A LSTM-based Generative Adversarial Network for End-use Water Modelling

Xie Yukun

Delft University of Technology

A Master of Science Thesis

Submitted to the Department of Environmental Engineering

Supervised by

Mirjam Blokker

Riccardo Taormina

June 2023

Abstract

Research pertaining to end-use water analysis plays a pivotal role in enabling local communities to enhance their management of pipelines, water resources, and associated policies. Nowadays, various end-use models have been developed based on diverse databases and measurements. Nonetheless, a predominant drawback prevalent in most of these models is their limited spatial scope and sluggish computational speed. This thesis endeavors to address these challenges through the proposition of a generative adversarial network (GAN) based stochastic end-use demand model. The SIMDEUM model, a stochastic end-use model, was first published in 2010. Since its inception, it has garnered substantial recognition and validation from numerous researchers. Within this thesis, the GAN model utilizes SIMDEUM as the training set and undergoes validation utilizing a comprehensive measurement dataset, encompassing over 1000 households from the Netherlands and the United States. Remarkably, the GAN model attains an error rate of 12% for end uses, coupled with an R^2 value exceeding 0.8 for the overall model. In contrast to SIMDEUM, the GAN model significantly enhances computational speed by more than 500%. Furthermore, the GAN model can be tailored to specific requirements and seamlessly processes raw data.It is concluded that the GAN-based stochastic water use model presented in this thesis adeptly simulates end-use water demand.

Keywords: End-use water demand, Generative adversarial network (GAN), Long short-term memory (LSTM)

Table of Contents

1. Introduction

The increase in the number of cities and global population, combined with climate change, poses a huge challenge to water availability and water distribution systems (Yang et al., 2018). Although residential water accounts for a relatively small proportion of total water use compared to agricultural and industrial water, the growth rate of residential water use is fast (Flörke et al., 2013). According to the World Resources Institute, residential water use increased by 600% from 1960 to 2014 (Madias et al., 2022). Rapid urbanization makes water scarcity more severe in developing countries (Sivakumaran et al., 2010).

In response to water scarcity, governments and organizations have proposed numerous water conservation programs. These measures include water conservation initiatives, water restrictions, source substitution and education, etc (Makki et al., 2015). However, the effectiveness of water conservation programs cannot be simply measured by numerical changes in water use (Makki et al., 2015). A study shows that higher-income people tend to buy premium equipment, resulting in more water-efficient showers and laundry, but on the other hand, 24% of higher-income people contribute to 80% of the water used for irrigation (Willis et al., 2009). To validate the effects of the policy and find the methods to further reduce water use, an in-depth understanding of end-use water is necessary (Gato-Trinidad et al., 2011). Besides political significance, the study of end-use water has important implications for the exploitation and reuse of water resources (Mazzoni et al., 2023). Drinking water is high-quality water, but drinking water is not necessary for every end use in daily use. Being able to use rainwater or recycled water for irrigation or toilets would help significantly in saving drinking water (Willis et al., 2013). The volume and availability of local water resources (storm water, roof water & recycled water) are highly variable. Learning and building end-use models can help governments to systematically plan and utilize these possible local water resources and achieve integrated urban water management (Kumudu et al., 2011). Given that domestic water consumption is discontinuous and each end use has its specific water quality, bottom-up end-use water models help to understand how pollutants/nutrients accumulate (or dilute) in sewers and how future changes in water use may affect them (Bailey et al., 2020; Blokker et al., 2008). Water leaks are widespread in households, accounting for an average of 5% of household's total water consumption (Gato-Trinidad et al., 2011). In certain instances, leaks have been

observed to reach an alarming range of 25% to 53% (Gato-Trinidad et al., 2011). With an End-use water demand model, detection of leakage or unknown water use pattern becomes possible (Mazzoni et al., 2023). Moreover, the analysis of end-use serves as a valuable tool for water companies in formulating effective pricing strategies and incentivizing users to cultivate more responsible water consumption habits (Barberán et al., 2000).

End-use water modeling is a method for summarizing and modeling end uses, such as toilet, shower and so on. Numerous researchers have endeavored to develop distinct models for end-use water demand, with various spatial and temporal scales. Early and widely used models include the regression models (Sivakumaran et al., 2010) and the PRP models (Steven et al., 2007). The SIMDEUM model proposed by (Blokker, 2010) gives actual physical meaning to the model after analyzing the statistical data of users and end uses, realizing the transformation from simulating the phenomenon to analyzing the essence. Cominola (2016) proposed a stochastic simulation model based on the US data and Rathnayaka (2017) launched a new end-use water demand model similar to SIMDEUM based on the Australian statistical data. Compared with SIMDEUM, this model adds new end uses (e.g. evaporative cooler) and different user groups (Rathnayaka et al., 2017b).

Nevertheless, the existing end-use demand models have limitations in spatial and temporal scope. From a spatial perspective, people in different regions have different living habits and water using habits. In Australia, people prefer to shower in the morning and showering is the largest water consumption end use, but in the United States, people prefer to shower at night and the largest water consumption end use is the toilet (DeOreo et al., 2016; Gato-Trinidad et al., 2011). Although models such as SIMDEUM have the ability to be extended to different regions, people with a good understanding of the model and distribution analysis of the corresponding regions are required. Given this circumstance, multiple nations, including the United States, Australia, and the Netherlands, have independently formulated their respective end-use water models. (Blokker et al., 2011; Mayer et al., 1999; Willis et al., 2009). Measurements to obtain statistical values require significant human and financial resources, and the results are limited by its spatial representation and not generalizable (Mayer et al., 1999). This geographical limitation exists not only between countries, but also between industries. Encompassing every facet of life through categorical modeling is an arduous task, bordering on impossibility. From a temporal perspective,

the use of explicit functions to compose time series lacks flexibility. The explicit functions have domains (e.g. one day) and cannot track changes outside the domain. For many models, each day of the simulation is independent and the mutual information between days is zero. The combination of domain and resolution (e.g. 1 day and 1 s for SIMDEUM) makes the model accurate within the defined time scale, but blurs other time scales and additional patterns are required for better simulation (Rathnayaka et al., 2017b). For example, variables (or parameters) that need to be tracked and measured over long periods of time are often ignored or defined as states, such as summer and winter, which makes it difficult for models to simulate the gradual transition of seasons.

The existing models also suffer from two types of algorithmic problems. Firstly, linear equations are widely used to describe the relationship between variables and drivers. For example, the frequency of shower in Stochastic Demand Generator suggested by (Duncan et al., 2008) is only related to household size. However, studies showed that the increase in water use is not simply proportional to the number of people in the household (Mazzoni et al., 2023; Rathnayaka et al., 2011). Water consumption is influenced by many factors which makes it difficult to be described as a combination of parameter-based linear functions (Wang et al., 2018). Secondly, statistical models have correlation problems between variates. Multiple variates may each come from a statistically meaningful distribution, but do not constitute a statistically meaningful multivariate distribution. Creaco et al. (2015) employed a bivariate-normal equation to establish the correlation between intensity and duration. However, as the number of variables expands, the demand for computational power and time escalates exponentially. Therefore, a model with higher computational efficiency is needed. Finally, models founded on probability distributions encounter challenges when attempting to establish sensible constraints on water usage. For example, users in SIMDEUM are likely to use the same end use (e.g. WC) continuously in a short period, due to its high probability of occurrence during that period. While this approach generates outcomes of macroscopic significance, it often compromises the rationality of individual behavior, thereby resulting in excessive water consumption within a short timeframe.

Compared to traditional statistical models, artificial intelligence models have the advantage of using fewer assumptions and obtaining good results on variables with non-linear relationships (Ghalehkhondabi et al., 2017; Wang et al., 2018). In the case of SIMDEUM, for instance, assumptions regarding the distribution type are necessary,

whereas such requirements are not obligatory within the deep learning model. The efficacy of a neural network predominantly hinges upon factors such as the composition of the training dataset, the neural network's architecture, and the underlying algorithm employed. Generative Adversarial Network (GAN) represents a deep learning model rooted in game theory, enabling both semi-supervised and unsupervised learning approaches (Creswell et al., 2018). GANs are implicit generative models, which means that GANs are devoid of modeling-induced errors (Ian et al., 2020). Another reason to use deep learning models is that urban water use is characterized by high complexity and high variability, so demand models are required to have the ability to quickly change on a small scale in response to the possible changes (Mitchell et al., 2007). The changes could include technical upgrade of end uses, adjustment of water policy in some regions and the emergence of new water use strategies (or habits). With deep learning models, the raw dataset can be directly used to update the model without the need for further analysis of data, such as penetration ratios and efficiency.

SIMDEUM effectively emulates residential end-use water demand through the distinct simulation of end uses and individual users, achieving a commendable level of accuracy. Nevertheless, SIMDEUM is not exempt from encountering the challenges highlighted in the above pertaining to traditional models. This thesis proposes a novel generative adversarial network (GAN) based end-use water model for short-term residential water modelling. The main research question for this model is: To what extent can the model achieve a prediction accuracy comparable to SIMDEUM in the final application by processing the raw data directly? On this basis, the model endeavors to address the limitations encountered by traditional models in terms of space, time and computational speed.

2. Methods and Materials

2.1. SIMulation of water Demand, and End-Use Model(SIMDEUM)

According to the findings, the predominant indoor water demand can be accurately represented by singular rectangular pulses. In light of this principle, SIMDEUM effectively characterizes these pulses using the following equation (Blokker, 2011):

$$\mathrm{Q} = \sum B(I, D, \tau) \quad (1)$$

$$B(I, D, \tau) = \begin{cases} I & \tau < T < \tau + D \\ 0 & elsewhere \end{cases} \quad (2)$$

Where I represents the pulse intensity (L/s); D represents the pulse duration (s); τ represents the pulse starting time(s).

SIMDEUM is characterized by its adaptability and versatility. Typical inputs for SIMDEUM include the duration (e.g. 2days), household type, weekday/weekend, pattern, etc. Furthermore, users can customize most variates, including the composition of user groups, the behavior of users and the frequency of end uses, by changing or introducing statistical data in SIMDEUM. For a one-day simulation, the output of SIMDEUM is an array with 4 dimensions – time, users, end uses and patterns. Users can aggregate different dimensions in the output to enable analysis of different aspects.

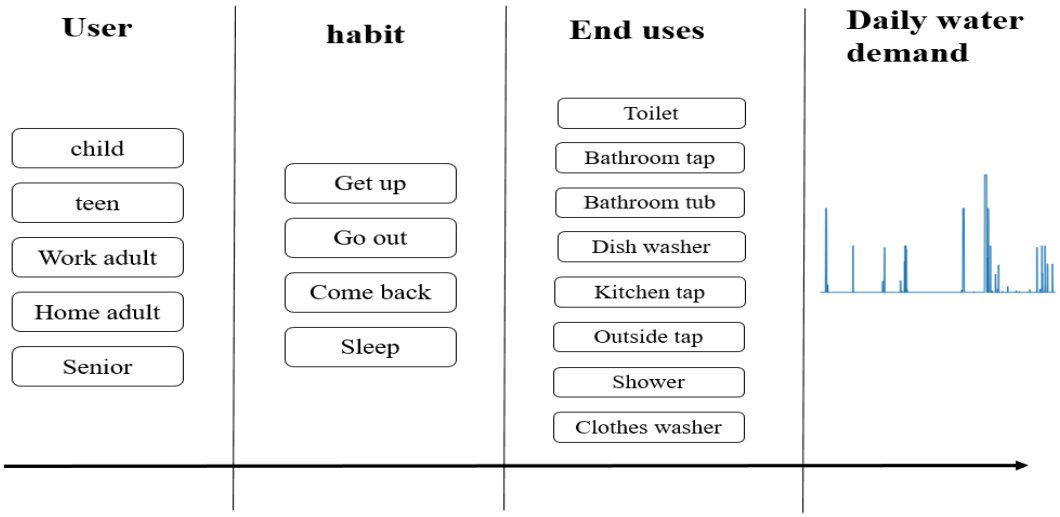


Figure 1. Structure of SIMDEUM

Different from traditional models, SIMDEUM does not totally rely on measurements, but is based on surveys of specific household members (or groups) and household appliances. SIMDEUM consists of 8 different end uses (see Figure 1). Except the intensity, other parameters of end uses, including the duration, frequency and start time, are related to the user. Among 8 different end uses, dishwasher, washing machine, kitchen tap and outside tap are shared by households. The use of the toilet, bathroom tap, bathtub and shower are controlled by the performance of each individual user. The influencing factors for the end uses include the age of the user, the user's diurnal habit, and the time end uses are used. After determining the parameters of users and habits, SIMDEUM employs a randomized process to generate end-use parameters, drawing

from the corresponding distribution for each parameter. In Figure 1, SIMDEUM calculates from left to right and finally obtains the daily water demand. Although SIMDEUM has a resolution of seconds, it does not support sub-daily simulations (e.g. from 3pm to 4pm).

As a stochastic model, SIMDEUM has been proven to be effective and efficient in simulating residential water demand (Blokker et al., 2017). Nowadays, SIMDEUM has been extended to the non-residential fields. With the support of prior knowledge and specific pattern, SIMDEUM has a root mean square error (RMSE) less than 30% and a coefficient of determination (R2) greater than 0.7 on time scales of 5 minutes, 1 hour, and 1 day (Blokker et al., 2011). Due to the lack of data, the simulation of SIMDEUM for hotel and nursing house is not ideal (Blokker et al., 2011). Besides normal water demand, SIMDEUM are extended to other water types, such as hot water and waste water (Blokker et al., 2017). SIMDEUM has been used by researchers and students in their studies (Blokker et al., 2017). Considering that SIMDEUM has achieved good results in the field of end use modelling, this study will use SIMDEUM data as input to simulate and compare the results with SIMDEUM.

2.2. Dataset

The dataset used in this thesis has two sources – SIMDEUM (python version, "pySIMDEUM") and measured data. Measured data includes data from Zandvoort (Netherlands) and REU2016 (USA). SIMDEUM data are mainly used for training models due to its reliability and large amount and the measured data are mainly used for validation. The basic information of the datasets is summarized in Table 5.

It is found that over 99% of water events in SIMDEUM are less than 1200s (20min) in duration. This thesis utilizes data with a resolution of 4 seconds instead of 1 second, resulting in a majority of water events having a duration of less than 300 seconds. In order to accurately represent water usage in line with the precision of 0.1 L, a variable is employed to track and store the accumulated amount of water consumed. This variable is consistently updated and propagated until it reaches the threshold of 0.1 L, at which point it is added to the water consumption at the corresponding moment. This approach imposes a limitation on the data resolution of SIMDEUM in this thesis, restricting it to 0.1 L intervals to replicate the behavior of real smart water meters during simulation.

In SIMDEUM, the different combinations of user types can be grouped into 3 household types – one person, two persons and family. This thesis combines these three types with weekend/weekday feature as input and creates a 4000-day dataset. In addition to the family type and the weekday information, other conditions are randomly specified inside SIMDEUM. The feature composition of this 4000-day SIMDEUM dataset is shown in Table 1 and detailed end use composition is shown in Table 2. Weekends and weekdays in this experiment are equally weighted variables. To prevent the model from focusing more on weekdays, the ratio between weekday and weekend is 1:1 in the dataset instead of 5:2. After inputting the features, SIMDEUM outputs the results in the form of a Dataframe (pandas package). The DataFrame results are indexed by time, with the end-use name as the column name. Since there is only one pattern in pySIMDEUM, there is also one pattern in the SIMDEUM dataset.

Table 1. Feature composition of the dataset

Household Size (unit: day)		
	Weekday	Weekend
one person	500	500
two person	500	500
family	1000	1000

Table 2. Composition of the end use in the dataset

End Use dataset (unit: count(ratio))			
Toilet	40622(35.4%)	**Bath tub**	39(<0.1%)
Bathroom Tap	28376(24.5%)	**Kitchen Tap**	30965(26.7%)
Outside Tap	2524(1.8%)	**Shower**	4839(3.9%)
Washing Machine	6195(5.4%)	**Dishwasher**	2639(2.3%)

Basic information for water users in Zandvoort area in shown in Table 3. In the Zandvoort area, there are not only residents, but also 3 hotels and 32 clubs. Clubs account for about 9% of the total water consumption, hotels for 24%, and residents for 68%. Among hotels and clubs, NH Hotel has the greatest impact (14%).The Zandvoort dataset contains not only water consumption in the Zandvoort area, but also information on water nodes and households in the Zandvoort area. Blokker (2010) studied the residence time of water at four different locations in the Zandvoort area by means of

tracers (Blokker, 2010). Location 1, 2 and 4 are based or close to the residential apartment; location 3 is located at NH Hotel. Table 9 shows the basic information in the Zandvoort area.

Table 3. Basic information about the Zandvoort area (Blokker, 2010)

	Amount	**Q(m^3/h)**
Small beach club	21	1.02
Large beach club	11	1.10
Residence	310	5.70
Apartment building	26	10.50
NH Hotel	1	3.247
Beach Hotel	1	1.783
Palace Hotel	1	0.50

REU2016 dataset used in this thesis refers to the Residential End Uses of water report, version 2 (DeOreo et al., 2016). The report covers 1,000 households in 23 regions of the United States. Compared with the SIMDEUM data, bathroom taps and kitchen taps are collectively referred to faucets in REU2016. The proportion of faucet use increases sharply, with a corresponding decrease in other end uses (except for bathtubs). Compared with Zandvoort dataset, the REU2016 dataset does not include demographical information on households.

Table 4. REU2016 End Use Dataset (DeOreo et al., 2016)

End Use	**Count (ratio)**	**End Use**	**Count (ratio)**
Toilet	124685(18.9%)	**Bathtub**	1742(0.3%)
Faucet	495931(75.2%)	**Outside Tap**	4754(0.7%)
Shower	17079(2.6%)	**Cloth washer**	11184(1.7%)
Dishwasher	3982(0.6%)		

Table 5. Summary of the basic information on the three datasets

	SIMDEUM	**Zandvoort**	**REU2016**
Year	/	2009	2013
Real data	x	√	√
Type of data	Residential	residential & non-residential	Residential
Household Information	√	√	x

Number of households	4000	371	1000
Time span	/	207 days	403 days
Resolution	4 second	15 minutes	10 second
Precision	0.1 L	1 L	0.01 L
End use number	8(6)*	/	13(5)

*8 represents the total end use number included in SIMDEUM and 6 represents end use number used in this thesis

2.3. Generative Adversarial Network (GAN)

In machine learning, there are two distinct types of models – discriminative models and generative models. Discriminative models focus on learning the boundary while generative models aim to learn and model the joint probability distribution of the input features and the labels. The generative adversarial network (GAN) is an implicit generative network widely used in image recognition/generation, text mining and speech synthesis (Aggarwal et al., 2021). GAN was first proposed by (Goodfellow et al., 2014) and can perform supervised, unsupervised, and reinforcement learning tasks.

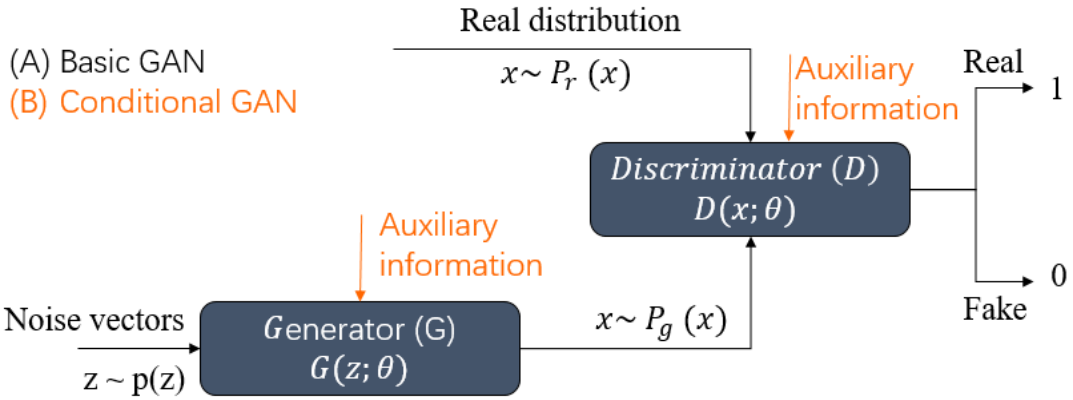


Figure 2. Structure of (A) Basic generative adversarial network (GAN); (B) Conditional GAN (CGAN)

The basic GAN consists of a generator and a discriminator (Figure 2). Generator receives inputs (noise) from a prior distribution p(z) and converts it into samples as real as possible by means of generator functions $G(z\ ;\ \theta^G)$. The generated samples are then mixed with the real samples and passed to discriminator to distinguish the authenticity. Traditional generative network used Kullback-Leibler (KL) divergence (also called relative entropy) to measure the distance between the fake distribution and the real distribution, but the KL divergence is asymmetric, making it difficult to balance

generation accuracy and generation diversity (Arjovsky et al., 2017). To optimize, GAN uses Jensen-Shannon (JS) divergence to measure the distance of distributions.

The loss of the discriminator and generator based on the JS divergence are calculated as follows:

$$\text{Loss}_{discriminator} = -\mathbb{E}_{x\sim P_r}\left[\log\left(D(x)\right)\right] - \mathbb{E}_{\boldsymbol{x\sim P_g}}\left[\mathbf{log\left(1 - D(x)\right)}\right] \quad (3)$$

$$\text{Loss}_{generator} = -\mathbb{E}_{\boldsymbol{x\sim P_g}}\left[\mathbf{log}\left(\boldsymbol{D}(\boldsymbol{x})\right)\right] \quad (4)$$

where D(x) is the discriminator's judgment on inputs, calculated by JS divergence, $D(x) \in (0,1)$; P_r and P_g are the distributions of the real and generated samples, respectively. It can be seen that $-\mathbb{E}_{x\sim P_g}\left[\log\left(1 - D(x)\right)\right]$ in Equation (3) is the opposite of $-\mathbb{E}_{x\sim P_g}\left[\log\left(D(x)\right)\right]$ in Equation (4). Attempting to minimize the opposite loss function results the generator and discriminator against each other.

2.4. Conditional Wasserstein GAN

Although GAN has achieved good results in different fields, instability is nowadays still a difficult part in the training processes (Jabbar et al., 2021). Theoretically, GAN converges when generator and discriminator reach Nash equilibrium (Goodfellow et al., 2020). However, this adversarial learning method is found to be instable in practical for two reasons (Appendix 7.2):

1. The support set often comes from low-dimensional manifolds, which cannot fill the high-dimensional space in the model, resulting in the gradient vanishing.

2. Model collapse caused by the asymmetry of KL divergence in JS divergence

To optimize the convergence problem of GAN, scholars have proposed different solutions to specific problems (Yan-Lin et al., 2022). Artificial noise addition on the generator and discriminator and batch normalization can mitigate the gradient vanishing slightly by increasing the alignment of real and fake distributions. To fundamentally solve the convergence problem, Arjovsky et al. (2017) proposed Wasserstein distance (equation 5) to substitute JS divergence and subsequently

employed it to construct a Wasserstein Generative Adversarial Network (WGAN).

$$\mathrm{W}(P_r, P_g) = inf_{\gamma \sim \prod(P_r, P_g)} \mathbb{E}_{(x,y)\sim\gamma}[\| x - y \|] \tag{5}$$

In which, $\prod(P_r, P_g)$ represents the joint distribution of all possible combinations of P_r and P_g; $\| x - y \|$ is the norm of x – y and *inf* is infimum. Compared to JS/KL divergence, the Wasserstein distance is smooth and provide meaningful gradients in high-dimensional space for two non-overlapping distributions.

Conditional GAN (CGAN) is a variant of GAN proposed by (Mirza et al., 2014) and its structure is shown in Figure 2. Conditional GAN increases the ability of GAN to output specific results by introducing conditional variables in both generators and discriminators. In this thesis, WGAN-GP is introduced into the CGAN to get more stable gradients while outputting specific results. It is worth to mention that the labels need to be excluded in the process of gradient calculation. The conditional improved loss function is converted to:

$$\mathbb{E}_{\tilde{x}\sim P_g}[\mathrm{D}(\tilde{x}|y)] - \mathbb{E}_{x\sim P_r}[\mathrm{D}(x|y)] + \lambda \cdot \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}[(\| \nabla_{\hat{x}} \mathrm{D}(\hat{x}|y) \|_2 - 1)^2] \tag{6}$$

WGAN-GP has relatively stable and smooth gradients in the case of strong discriminators. A common training method is to preferentially train discriminators several times, and then have the perfect discriminator drive the generator. This thesis uses WGAN-GP as the base logic and the training algorithm used in this thesis is shown in Table 6.

Table 6. Algorithm of WGAN-GP used in this thesis

Algorithm – WGAN-GP
Require: The gradient penalty coefficient λ, the number of discriminator iterations per generator iteration $n_{discriminator}$, the batch size m, Adam parameters α, β_1, β_2
Require: Initial discriminator parameters ω_0, initial generator parameters θ_0
While θ has not converged **do**
For t = 1,…, $n_{discriminator}$ **do**
For i = 1,…,m **do**
Sample real data $\mathbf{x} \sim P_r$, label **y**, latent variable $\mathbf{z} \sim p(z)$, $\varepsilon \in [0,1]$
$\tilde{\mathbf{x}} \leftarrow G_\theta(\boldsymbol{z}, \boldsymbol{y})$
$\hat{\boldsymbol{x}} \leftarrow \varepsilon \mathbf{x} + (1-\varepsilon)\tilde{\mathbf{x}}$
$L^{(i)} \leftarrow D_\omega(\tilde{\mathbf{x}}|\boldsymbol{y}) + D_\omega(\boldsymbol{x}|\boldsymbol{y}) + \lambda(\| \nabla_{\hat{\boldsymbol{x}}} D_\omega(\hat{\boldsymbol{x}}|\boldsymbol{y}) \|_2 - 1)^2$
end for

$\omega \leftarrow Adam(\nabla_\omega \frac{1}{m}\sum_{i=1}^{m} \mathrm{L}^{(i)}, \omega, \alpha, \beta_1, \beta_2)$

end for

Sample a batch of latent variables $\{\mathbf{z}^{(i)}\}_{i=1}^{m} \sim p(\boldsymbol{z})$ and label **y**

$\theta \leftarrow Adam(\nabla_\theta \frac{1}{m}\sum_{i=1}^{m} -\mathrm{D}_\omega(G_\theta(\boldsymbol{z}, \boldsymbol{y})), \theta, \alpha, \beta_1, \beta_2)$

end while

2.5. Conditional LSTM-WGAN Model

Yu et al. (2021) employed Long Short-Term Memory Generative Adversarial Network (LSTM-GAN) to generate time series data, yielding favorable outcomes. In order to validate the applicability of this methodology for water demand modeling, this thesis proposed an LSTM-WGAN end-use water demand model. The proposed model is built upon the algorithm presented in Table 6.

2.5.1. Model overview

In this thesis, the term "LSTM-GAN End Use Model" (LGEUM) is used to express the overall model and term "GAN" is used to describe the GAN part in the overall model. The LGEUM includes two parts, the "end use model" and the "user behavior model". These two names will be used to refer specifically to two sub-models. Besides the GAN, the LGEUM includes a data pre-processing part and a result processing part.

Referring to SIMDEUM, a total of 8 end uses are used in the LGEUM – Wc, bathtub, bathroom tap, kitchen tap, outside tap, shower, washing machine and dishwasher. The LGEUM employs a single tap as the smallest spatial scale and a temporal scale of 1 second. The LGEUM provides three family types (single person, two persons, and families with more than 2 people) and distinguishes between weekdays and weekends. Designed with flexibility in mind, the LGEUM can incorporate additional end uses and demographic characteristics and can be trained on various temporal or spatial resolutions.

2.5.2. Dataset pre-processing

The GAN has fixed input patterns in both end use model and user behavior model. To

have the ability to process raw data, the LGEUM wraps data pre-processing functions to convert the raw data into the desired format.

The preprocessing functions for the end use model provide two raw data input formats: 1. Input intensity and duration. This input mode is suitable for data that can be described by intensity and duration alone (e.g. SIMDEUM); 2. Input complete water events. This input mode is suitable for real data (e.g. REU2016).

The user behavior model simulates the start time and frequency of water use events at the same time. For the input required by the user behavior model, the length of the input represents the frequency of use and the value represents the start time of the event. The maximum length (frequency) of the input is set to 50 because the probability of a household using the same end use more than 50 times in a day is less than 5% in SIMDEUM. Value from 0-1 represent the start time of the event during the day – 0 for no occurrence, 0.5 for 12:00 and 1 for 24:00. To facilitate subsequent result processing (section 2.5.4), the start times are sorted in descending order.

For both end use model and user behavior model, one-hot encoding is used to label categorical features. A typical label includes 12 features – one person, two persons, family, weekday/weekend, WC, bathtub, kitchen tap, bathroom tap, outside tap, shower, washing machine and dishwasher. For most features in the label, 1 means the feature exists and 0 means not; for weekday information, 1 means weekday and 0 means weekend. Only one 1 is allowed for the same type of feature (e.g., 8 end uses are of the same type). To prevent labeling errors and improve efficiency, labels can be generated by inputting names or index numbers. Besides generating labels, remove unwanted features from labels by index is also feasible. The GAN model can automatically adapt to the input label shape, making it easy to extend/reduce the features as required.

The LGEUM provides functions to repeat samples for a specific end use or feature. In this thesis, the samples for bathtubs are repeated 20 times and the samples for outside taps are repeated 2 times.

2.5.3. Selection of kernels and Hyperparameters

In the LGEUM, Long Short-Term Memory (LSTM) and 1D-Convolutional Neural Network (1D-CNN) are employed as kernels in the generator and discriminator

components. LSTM is a variant of RNN and has commendable performance in the processing of long sequences and time series (Zhou et al., 2023). By controlling the transmission state, LSTM can remember important information for a long time and forget the unimportant information. The 1D-Convolutional Neural Network (1D-CNN) represents a specialized variant of CNN suitable for effectively handling time series and textual data. In this thesis, 1D-CNN is used in the discriminators. The rapid computational speed of 1D-CNN supports WGAN-GP in enabling efficient and repeated training of the discriminator without significantly consuming extensive time and computational power.

The model is highly sensitive to the optimizer parameters, gradient penalty weight and learning rate. In the experiment, the parameters for performing the grid search are shown in Table 7. To conserve computational power, if the generator loss function shows an increasing trend for 20 consecutive epochs, the model will be truncated and subjected to review. In addition to grid search, the rest hyperparameters are adjusted according to the results of the model.

Table 7. Parameter set used in grid search

	Range
Optimizer	Default Adam, PMSProp, Adam (beta_1=0.5, beta_2=0.9)
Gradient penalty	0.01, 0.05, 0.1, 0.3, 1
learning rate	0.0001, 0.0002, 0.0005, 0.001
Activation	Tanh, ReLu, LeakyReLu(alpha = 0.1), LeakyReLu(alpha = 0.2)
Dropout rate	0.1, 0.2, 0.5

The hyperparameters used in the model can be divided into three categories: the overall hyperparameters used for training (Table 8), the generator hyperparameters (Table 9) and the discriminator hyperparameters (Table 10). In Table 8, the maximum time steps of LSTM is set to 300 because long input sequences may result in vanishing gradients. In Table 9 and Table 10, the structure of generator and discriminator is inspired by (Yu et al., 2021).

Table 8. Training parameters

Batch size(end use model)	64
Batch size(user behavior model)	96

Random latent vector	32
LSTM length(end use model)	300
LSTM length(user behavior model)	50
Optimizer	Adam (beta_1=0.5, beta_2=0.9)
Learning rate	0.0002
Training epoch(end use model)	100
Training epoch(user behavior model)	500
Gradient penalty weight	0.1
Extra discriminator training	5
Loss function generator	Reduced mean
Loss function discriminator	Reduced mean

Table 9. Generator structure

	End Use Model	**User behavior Model**
Input layer	32(random vector) + 6(end use)	32 + 6(end use) + 4(family type)
Layer 1	Dense, 1800, LeakyReLU	Dense, 200, LeakyReLU
Layer 2	LSTM, 4, Dropout = 0.2	LSTM, 8, Dropout = 0.5
Layer 3	LeakyReLU(alpha = 0.2)	LeakyReLU(alpha = 0.2)
Layer 4	LSTM, 8, Dropout = 0.2	LSTM, 16, Dropout = 0.5
Layer 5	LeakyReLU(alpha = 0.2)	LeakyReLU(alpha = 0.2)
Layer 6	Dense, 1	Dense, 1

Table 10. Discriminator structure

	End Use Model	**User behavior Model**
Input layer	6(end use) + 1	6(end use) + 4(family type) + 1
Layer 1	Conv1D, 64, kernel size = 3	Conv1D, 64, kernel size = 8
Layer 2	LeakyReLU(alpha = 0.2)	LeakyReLU(alpha = 0.2)
Layer 3	Dropout = 0.3	Dropout = 0.3
Layer 4	Conv1D, 128, kernel size =3	Conv1D, 128, kernel size = 8
Layer 5	LeakyReLU(alpha = 0.2)	LeakyReLU(alpha = 0.2)
Layer 6	Dropout = 0.3	Dropout = 0.3
Layer 7	Flatten layer	Flatten layer
Layer 8	Dense, 1	Dense, 1

2.5.4. Result processing

The main function of result processing is to merge the results from two sub-models. In result processing, days are used as the basic time unit. Water consumption beyond 24:00 will be added to the morning of the same day (instead of the next day). The water use event will regain a 1s resolution (from 4s resolution) by interpolation.

Masks are utilized to delineate the boundary and, consequently, ascertain the valid data points. The length of a water event (or frequency) is deemed concluded when a value surpassing the defined boundary is encountered. For the end use model, the presence of an intensity value below 0.001 indicates the end of a water use event. For the user behavior model, the start time does not necessarily converge to 0 or 1, making it difficult to determine the noise. The judgment boundaries are shown in Table 11.

Table 11. Boundaries for valid start time for different end uses

End Use	**Boundary**
Toilet	0.025 < Start time < 1
Bathtub	0.05 < Start time < 1
Kitchen Tap	0.002 < Start time < 1
Bathroom Tap	0.002 < Start time < 1
Outside Tap	0.15 < Start time < 1
Shower	0.05 < Start time < 1

2.6. Model Environment

The model in this thesis is trained based on TensorFlow 2.8-GPU and Python 3.8. In addition to TensorFlow, some TensorFlow components (e.g. TensorFlow_docs) and sklearn are necessary.

2.7. Validation Methods

This thesis uses Root Mean Square Error (RMSE, equation 7), coefficient of determination R^2, Relative Error (RE, equation 8) and Graphical Comparison as performance metrics.

For the end use model and the user behavior model, the results generated by the GAN are compared with SIMDEUM. The comparisons encompass various aspects of end uses, including intensity, duration, volume, frequency, and patterns. In the calculation, the SIMDEUM value is treated as the true value and the value generated by GAN is considered as the predicted value.

For the overall end use model, the results generated by the LGEUM are compared with the Zandvoort dataset. The number of households assigned to each water node in the Zandvoort area is derived by computing the ratio between the base water demand and the total water demand. Subsequently, the LGEUM is employed to simulate the corresponding number of households. The generated water demand data for each water node is integrated with EPANET 2.2 (15-minute resolution), which is utilized to analyze and determine both the water demand and water age within the Zandvoort area. In the validation process, the generated water demand and water age is compared with SIMDEUM, EPANET model (Top-down water demand model) and Zandvoort measurement data. The non-residential patterns, such as those from clubs and hotels, which are not encompassed within the GAN model, are maintained in the results derived from the EPANET model.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}} \times 100\% \tag{7}$$

$$\mathrm{RE} = \frac{\hat{y} - y}{y} \times 100\% \tag{8}$$

Where $\hat{y}$ is predicted value and y is true value (or measurement value).

In order to assess and verify the extensibility, the GAN model is fine-tuned using the REU2016 dataset. Seventy percent (70%) of the REU2016 dataset is allocated as the training set, while the remaining thirty percent (30%) is designated as the test set. Given the absence of resident information, the extensibility validation process exclusively focuses on the end use model. The intensity, duration, and volume of the end use events generated by the GAN are subjected to comparison with the test set of the REU2016 dataset.

3. Results & Discussion

3.1. End Use Model

3.1.1. Training overview

The end use model is designed to determine the intensity and duration of water events for each end use. There is no household type distinction between end uses, so no additional label entry is required for this section. The end use model is trained for 100 epochs. The training time for each epoch ranged from 234s to 323s, with a total of around 8.3 hours. In the test, using CNN in the discriminator is more than 4 times faster than using LSTM.

3.1.2. Feasibility of GAN simulation end use

In contrast to SIMDEUM, end use events in GAN do not exhibit identical intensities. Shower events are taken as examples in Figure 4 because of their long duration and variable pattern. In Figure 4, it can be seen that the shower event generated by GAN has an upward trend while SIMDEUM has a stepwise change. The intensity of the event in Figure 4(b) is not stable because of three reasons: 1. the results of GAN are calculated from random points of the prior distribution (normal distribution in this thesis); 2. The variability of intensity in the dataset; 3. The structure of GAN and number of epochs.

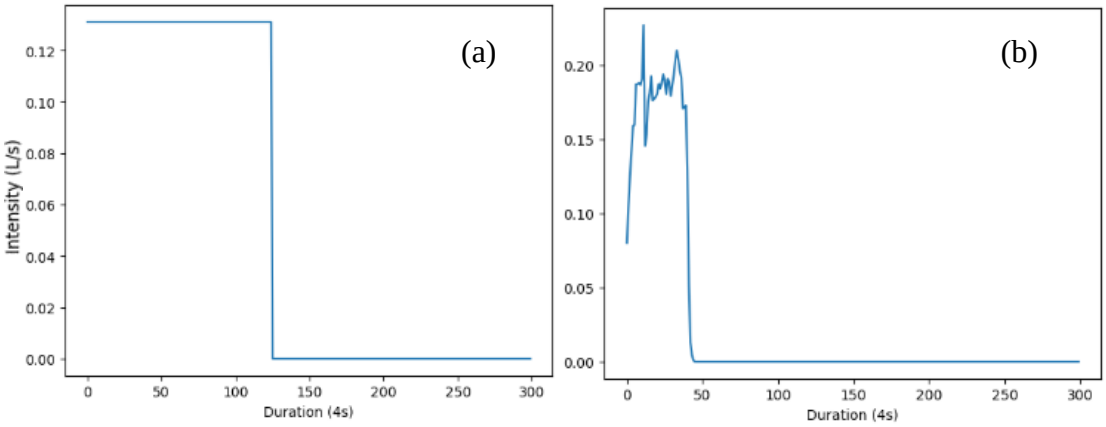
Ultimately, the GAN model incorporates five end uses. The inclusion of washing machines and dishwashers is omitted due to their fixed and non-continuous patterns in pySIMDEUM. Simulating washing machines and dishwashers (abbreviated as W&D) in GAN has the following challenges:

1. The duration of most W&Ds exceeds 1200s, which means that the GAN needs a lower resolution to cover them.
2. The simulation of GAN is variable. Currently, the pattern of W&Ds in the dataset is fixed and GAN has no advantage over SIMDEUM on the simulation of W&Ds.
3. Water demand of W&Ds is not continuous. In early experiments, separate simulations of the water use peaks and intervals of W&Ds were tried. Due to the fixed pattern in dataset, GAN simulates the water use peak well (like bathtub in section 4.1.3). However, GAN needs another model to simulate the interval between peaks. This extra

model requires to determine the number and duration of intervals. In practice, the combination of events can easily lead to distortion and an extra model leads to a decrease in computing speed and an increase in training difficulty.

There are two feasible methodologies to simulate W&Ds in GAN: 1. reducing the resolution; 2. treating W&Ds as continuous events (as in REU2016). Both of them require a more diverse W&Ds dataset and is subject to future validation. In addition, GAN is going to allow access to exogenous results. In addition to introducing new end uses into the model, end uses with relatively fixed patterns can be independently calculated and subsequently integrated into the final results.

Figure 4. The shower event generated by (a) GAN; (b) SIMDEUM



3.1.3. Result comparison

The comparison of event intensity and duration between SIMDEUM and GAN are shown in Table 12. This comparison is based on simulations of SIMDEUM and GAN for 1000 days each.

Table 12 illustrates that simulations for longer duration end uses (e.g., shower) tend to exhibit shorter durations, while shorter duration end uses demonstrate longer durations. This discrepancy may arise from the model's limited capacity to accurately differentiate between labels. In general, the intensity of the end use is generally underestimated in the whole Table 12. Notably, the extent to which an end use is overestimated in terms of duration (e.g., Kitchen Tap) correlates with the extent to which it is underestimated in terms of intensity.

Table 12. Comparison of event intensity and duration between SIMDEUM and GAN

	Intensity(L/s)			Duration(s)		
	SIMDEUM	**GAN**	**RE(%)**	**SIMDEUM**	**GAN**	**RE(%)**
Toilet	0.084	0.075	-10.7%	60.177	60.248	0.1%
Bathtub	0.198	0.187	-5.6%	603.223	581.20	-3.6%
Kitchen Tap	0.042	0.032	-23.8%	26.221	36.716	40.0%
Bath Tap	0.054	0.045	-16.7%	27.741	40.292	45.2%
Outside Tap	0.090	0.093	3.3%	233.762	167.05	-28.5%
Shower	0.131	0.124	-5.3%	525.101	421.13	-19.8%
Overall RMSE			13.1%			28.5%

	Volume(L)		
	SIMDEUM	**GAN**	**RE(%)**
Toilet	5.055	4.519	-10.6%
Bathtub	119.438	108.68	-9.0%
Kitchen Tap	1.101	1.175	6.7%
Bath Tap	1.498	1.813	21.0%
Outside Tap	21.039	15.535	-26.1%
Shower	68.788	52.220	-24.0%
Overall RMSE			18.0%

Among the end uses, the toilet and bathtub have small errors in all three metrics. the toilet benefits from its largest sample size and the bathtub benefits from its fixed pattern and 20 times repeating. However, mere repetition ignores the variability of the water event itself. This may lead to high accuracy in the comparison with SIMDEUM but distortion in the comparison in the measurement data. Improving the diversity of samples is an important direction for further optimization of the LGEUM model. Although the kitchen faucet has a small volume error, this is due to the neutralization of intensity and duration. Outside Tap has the largest volume error (-26.1%), which is mainly caused by its short duration (-28.5%). Results for short-duration end uses (e.g. taps) suffer from lower intensity and longer duration compared to SIMDEUM. One possible reason is the decrease in resolution (1s to 4s) in the dataset. Overall, most relative errors in Table 10 are negative, indicating that the water consumption of events in GAN model is less than that of SIMDEUM.

Figure 5 shows the probability distribution of the total water consumption for four typical end uses. The bathtub is excluded because of its fixed water pattern in the SIMDEUM. In Figure 5, the probability distribution of SIMDEUM is a combination of multiple explicit probability distributions, represented by the solid blue line. The probability distribution of GAN is generated from a histogram of 1000 days of data, represented by the orange scatter.

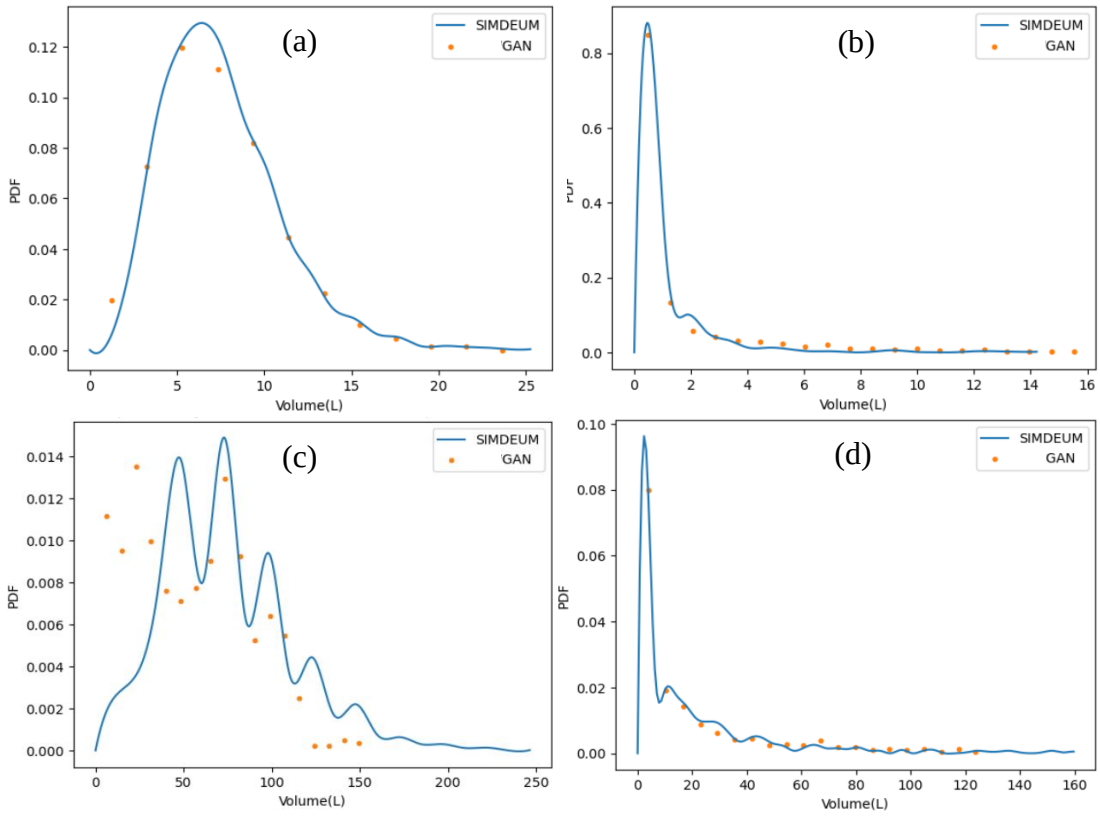


Figure 5. Water consumption probability distribution of (a) Toilet; (b) Bathroom Tap; (c) Shower; (d) Outside Tap

In Figure 5, the water consumption distribution of GAN's toilets (a), bathroom taps (b), and outside taps (d) exhibits a resemblance to that of SIMDEUM. There are clear peaks in the distribution of showers in Figure 5(c) because the duration of shower depends on the user's age and the intensity depends on the shower type in pySIMDEUM. Given the relatively narrow disparity in intensities, the shower water consumption is primarily regulated by 5 user types in pySIMDEUM, thereby leading to the presence of 5 peaks in Figure 5(c). The orange scatter(GAN) is overall to the left compared to the solid blue line(SIMDEUM), which corresponds to the fact that the shower event in GAN consumes less water in Table 12.

3.2. User Behavior Model

3.2.1. Training overview

The user behavior model is designed to determine the start time and frequency of water consumption events for households consisting of different user groups. To train the user behavior model, the full label described in the method section is required. The model was trained for 300 epochs, for a total of 2.1 hours.

3.2.2. Feasibility & Result comparison

Table 13 shows the comparison of average frequency of end uses in GAN and SIMDEUM based on 4000 days. Considering the significant differences in frequency among various end uses throughout a day, in addition to relative error(RE), Table 13 introduces weighted relative error to assess the actual impact of a specific end use on the overall user behavior model. Weighted RE in Table 13 is calculated as:

$$\text{Weighted RE} = \left(\frac{\text{frequency(SIMDEUM)}}{\text{sum of frequency(SIMDEUM)}}\right) \times \text{RE} \qquad (22)$$

In Table 13, there are 6 end uses with relative errors of less than 8%, including less than 5% for toilet and bathroom taps. The smallest relative error comes from the bathroom tap (-3.59%) while the largest relative error comes from the bathtub (-87.5%). The large error in the bathtubs is mainly due to the small sample size. The bathtub is hardly used in GAN's simulation. Although the average frequency of bathtub in Table 13 is 0.01, for one- and two-person households, the frequency of bathtub is 0, which means the user behavior model fails to simulate the bathtub.

In contrast to the end use model, the user behavior model adopts days as the fundamental unit instead of end uses. This change aims to explore potential correlations between different end uses within a single day. However, this method leads to a fact that the start time pattern of bathtub can't be repeated for 20 times like events. The method of adding weights to the bathtubs without changing the small sample size was tried, but it did not bring improvement to the overall model. Due to the low frequency (weight), the imprecision of the bathtub has limited impact on the overall simulation, resulting in a weighted RE of -0.15%. To improve the simulation of end use with small

sample sizes (or low frequency, e.g. bathtub), future optimizations to the structure of model and dataset are required. Other ways to address bathtub uncertainty include adding human intervention (e.g. repeats) or building new models.

Overall, Simulation of end use frequencies using GAN is feasible and has yielded good results. GAN and SIMDEUM are similar in terms of the frequency of end uses, with a RMSE of weighted RE equal to 4.2%.

Table 13. Average daily frequency of end uses in GAN and SIMDEUM

	SIMDEUM	**GAN**	**RE(%)**	**Weighted RE(%)**
Toilet	15.01	15.74	4.86%	1.58%
Bathtub	0.08	0.01	-87.5%	-0.15%
Kitchen Tap	15.92	14.77	-7.22%	-2.49%
Bathroom Tap	12.24	11.80	-3.59%	-0.95%
Outside Tap	1.00	1.07	7.00%	0.15%
Shower	1.84	1.71	-7.06%	-0.28%
Overall RMSE			36.2%	4.2%

Not all end uses in SIMDEUM have unique patterns (or distribution). Most end uses follow the user's presence pattern, which means the higher the probability of user presence, the higher the probability of the end use being used. This algorithm ignores the uniqueness of each end-use itself, inevitably leading to conflicting end uses at the same time and heavy use of end uses in a short period of time. The GAN model provides channels for each end use to obtain its unique “pattern”. These “patterns” interact with each other to create a holistic “pattern” for a household. The holistic “pattern” in GAN varies with household composition and external conditions. Under the influence of the holistic pattern, the pattern of all end uses in the GAN is determined simultaneously, rather than one by one.

In the Figure 6, the start time pattern of the 3 end uses with the highest weighted RE (in Table 13) are compared. The pattern curves in Figure 6 are derived from the same household combination as in the SIMDEUM dataset. The R^2 scores for (a), (b), (c) in Figure 6 are 0.73, 0.79 and 0.14, respectively. The R^2 scores indicates that the start time pattern of toilet and bathroom taps of GAN has a good agreement with SIMDEUM's.

Figure 6(a) shows a peak in the use of toilets in GAN at around 11:00, which is not observed in SIMDEUM. In Figure 5(c), SIMDEUM's bathroom tap has a start time pattern with distinct peaks. The reason for this phenomenon is that unlike the bathroom taps, the kitchen taps have its own start time pattern and GAN does not simulate it effectively. In Figure 5(c), the GAN maintains a similar pattern to that in Figure 5(a) and (b). This may be caused by the impact of other end uses on the kitchen tap. In addition, the patterns in the model are closely related to the masks described in the Table 11. Therefore, changing the hyperparameters in the mask can make the model more closely match the data to be simulated.

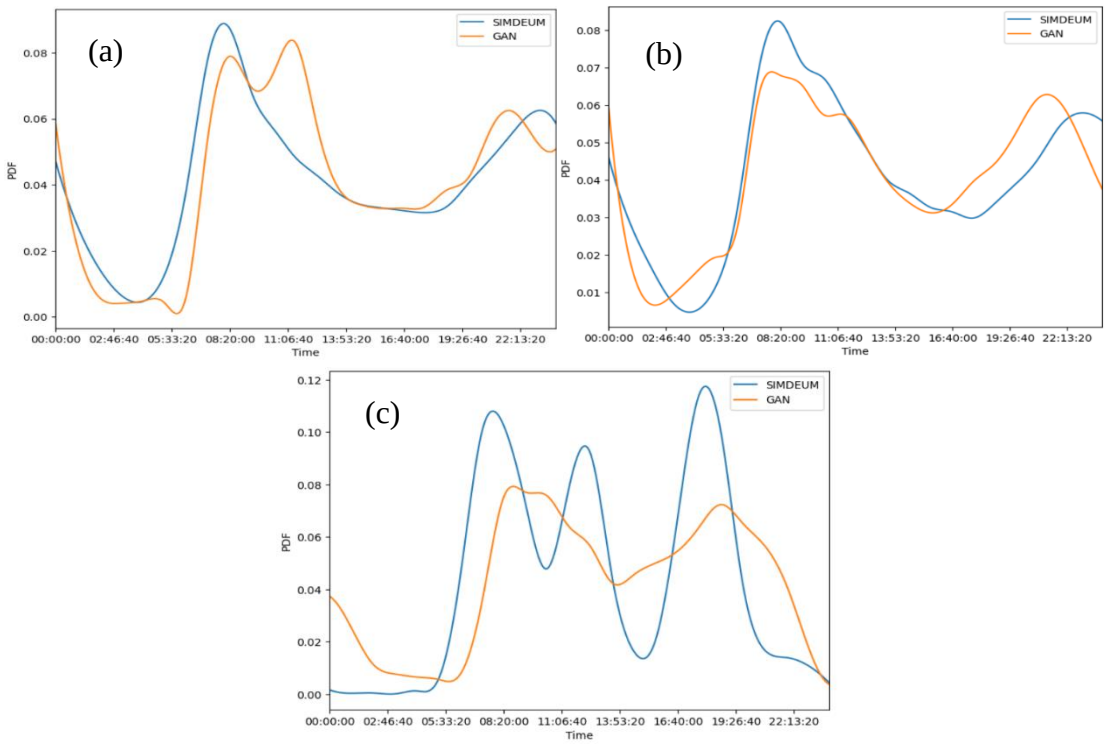


Figure 6. SIMDEUM and GAN's start time pattern for (a) Toilet; (b) Bathroom Tap; (c) Kitchen Tap

3.3. LSTM-GAN End Use Demand Model (LGEUM)

3.3.1. Model Overview

The LSTM-GAN Water demand model discussed in this section consists of two separate sub-models, the end use model (section 3.1) and user behavior model (section 3.2). The model adds events provided by the end use model to the corresponding

timeline based on the start time and frequency provided by the user behavior model. The output of the model contains 3 dimensions – time, end use and intensity. During the process of sub-model merging, within the same household, the toilet will consistently employ the same event whenever it is used, while other end uses will invoke the end use model for a new simulation each time they are utilized. In practice the end use is influenced by many other factors, such as season, wealth and climate (Mayer et al., 1999), which is not included in the model. The model uses uniform global variables for each part, so that features can be easily introduced or replaced.

The model is able to process on raw data, without the need to understand the data. Under different conditions, the average computing time of GAN simulation for one day is 0.0527s. Compared with SIMDEUM, GAN has 546% faster computing speed (average 0.288s). When running the simulation for an equivalent number of days, GAN consumes 2181MB of memory, whereas SIMDEUM utilizes 1697MB. In general, GAN is more suitable for large scale and large number of simulations.

3.3.2. Flow Validation

Figure 7 and Table 14 show the cumulative frequency comparison of Q_{max} between SIMDEUM and GAN for different time scales, based on 300 households. To compare with the GAN, washing machines and dishwashers are excluded from the SIMDEUM in Figure 7. RMSE in Table 14 is composed of two parts: the absolute values and as percentages of the SIMDEUM means.

It can be seen from the Figure 7, the orange line representing GAN is predominantly positioned to the left of the blue line representing SIMDEUM. This indicates that GAN exhibits a more restrained behavior in terms of Q_{max}. The negative RMSE in Table 14 validates this conclusion. There are two explanations for this phenomenon: 1. Water events are more dispersed in GAN, with fewer short-term bursts of water use. There are few cases where multiple end uses are used consecutively at the same time; 2. In section 3.1.3, it is found that the events simulated by GAN have a lower intensity compared to SIMDEUM. In Figure 7(a), GAN and SIMDEUM have clear maximum difference based on 5 min Q_{max}. GAN reaches a maximum water consumption of 50L, while SIMDEUM reaches 70L. In Figure 6(b), the hourly Q_{max} of GAN and SIMDEUM match well with an R^2 score of 0.96. Overall, The Qmax of GAN is basically consistent with SIMDEUM, and the average R^2 is greater than 0.8.

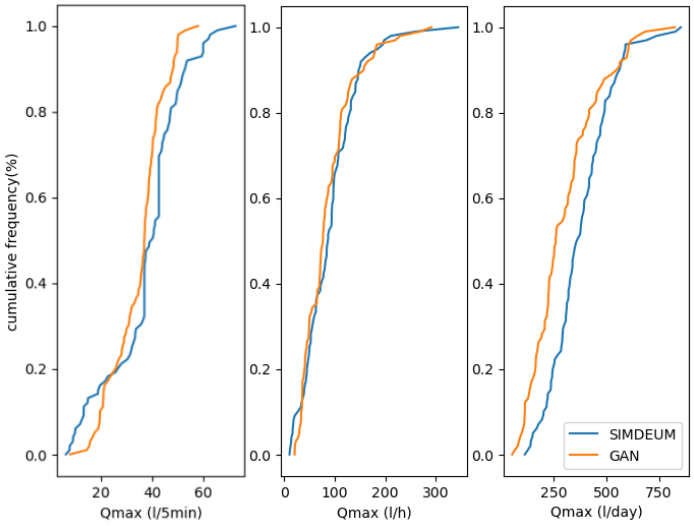


Figure 7. Cumulative frequency comparison of Q_{max} between SIMDEUM and GAN: (a) per 5 min; (b) per 1 hour; (c) per day

Table 14. Statistics of Q_{max} comparison between SIMDEUM and GAN

	RMSE	**R^2**
Q_{max} (l / 5 min)	5.41 (-13.5%)	0.71
Q_{max} (l / hour)	11.23 (-12.4%)	0.96
Q_{max} (l / day)	84.18 (-22.3%)	0.77

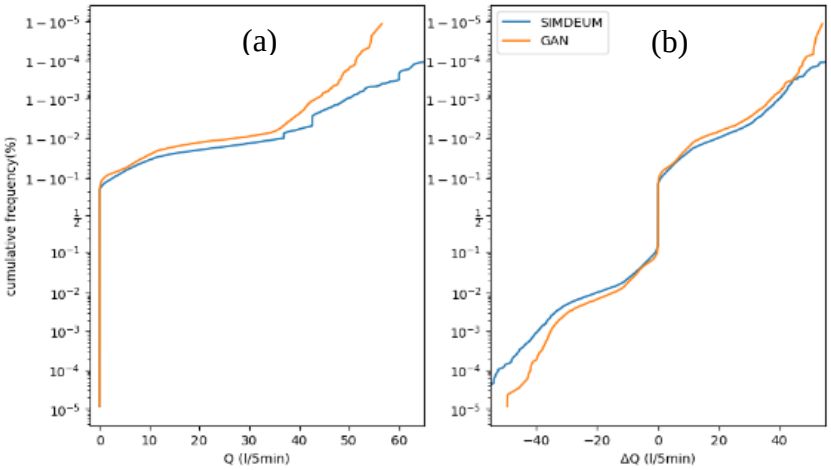


Figure 8. Cumulative frequency comparison of (a) Q and (b) Q_{diff} between GAN and SIMDEUM

A comparison (500 households) of the flow rate Q and the flow difference Q_{diff} for a time scale of 5 minutes is shown in Figure 8. The R^2 of Figure 8(a) and (b) are 0.91 and 0.96, respectively. In Figure 8(a) GAN uses less water, consistent with Figure 7 and in (b), GAN has smaller flow changes. Overall, GAN not only has less short-time water

consumption, but also the change of water consumption is smoother.

3.3.3. Water Demand Validation

Figure 9 shows the water consumption pattern in the Zandvoort area on a time scale of 15 minutes. The green dashed line plots the average pattern of GAN (2000 households), which has no geographical difference. Due to the absence of washing machine and dishwasher in the result of GAN, the probability distributions instead of the demand are compared in the Figure 9. Due to the presence of hotels, Zandvoort has a higher midnight water demand than Haarlem, while the rest of the day the water demand is roughly the same. GAN currently only provides residential water simulation and the water consumption of GAN in Figure 9 is significantly lower at midnight than the actual consumption. In addition, there is a rapid decrease in water GAN's consumption between 20:30 and 22:00, which is much stronger than the actual change.

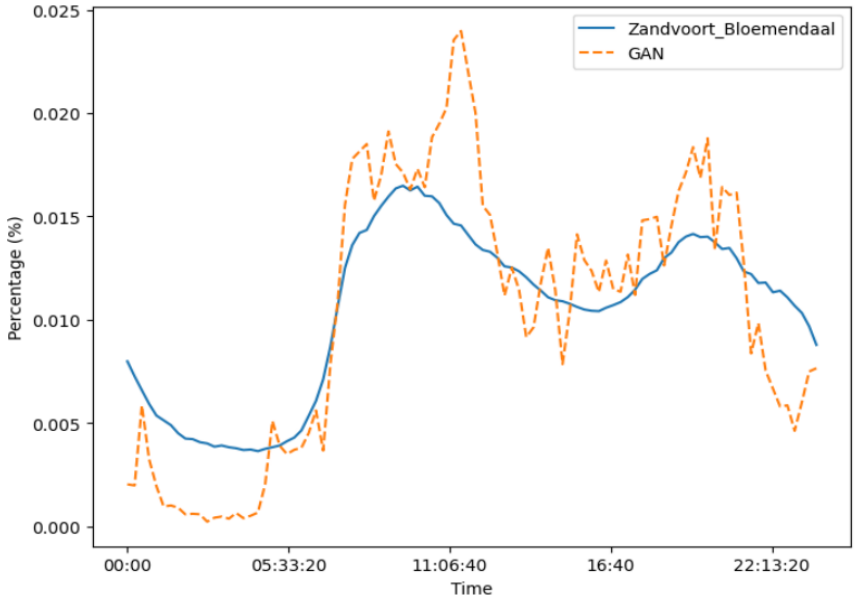


Figure 9. Water demand patterns in the Zandvoort, Haarlem area and GAN

Compared to the residents, clubs and hotels provide significant amount of water consumption at midnight in the Zandvoort area, especially the NH Hotel (location 3), which counts for 14% hotel water consumption. Without correction, SIMDEUM and therefore LGEUM are not suitable for simulating water consumption for the entire Zandvoort region. In Figure 10, the GAN retains the midnight water demand of NH Hotel in location 3. The Figure 10 refers to (Blokker, 2010), in which the time scale of SIMDEUM and the measurement curve is 5 minutes while the time scale of GAN is 15 minutes.

In both Figure 9 and Figure 10, the GAN has a noon water consumption peak from 10:30 to 12:30, which is higher than measurements. The main components of this peak are 49.4% Toilet, 34.3% shower and 9% kitchen tap. This result is consistent with Figure 6(a) and shows that toilets in GAN may be overused during this time period. Despite the retention of midnight water demand in NH hotel, GAN has a lower midnight (1:00-5:30) water consumption in Figure 10 compared with measurements. This indicates that retaining part of the water demand cannot completely solve the problem of LGEUM's lack of pattern. Compared to SIMDEUM, GAN maintains the same downward trend in the evening as the measurements, and the water demand in the early morning is more like the measurements. In addition, the different cut-off values for different end uses in the mask may play an important role because mask will lead to a decrease in the number of value around 0 and 1(around 0:00). However, mask has a great influence on the boundary, while the influence on the other values is negligible.

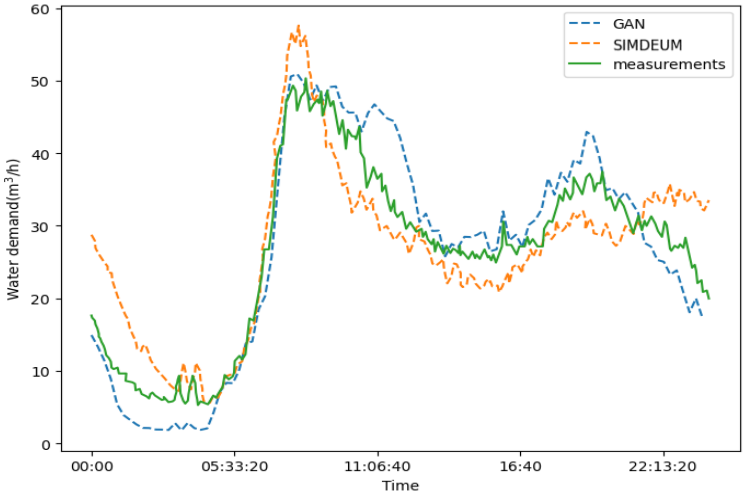


Figure 10. Comparison between measured and simulated water demand in the Zandvoort area (Blokker, 2010)

The possible changes to Figure 10 caused by the absence of washing machines and dishwashers in the GAN model must be considered. In Netherlands, washing machines account for about 14% of total water use, and dishwashers account for about 3% (Mazzoni et al., 2023). In SIMDEUM, the washing machine are mainly used in the morning (7:30 - 11:00) and followed by the afternoon (19:00 - 21:00). Therefore, the main effect of the washing machine on the water demand pattern in Figure 10 is a possible increase of up to 14% in the morning and afternoon peaks, with a corresponding small decrease in the other hours. Considering that washing machine are more often used on weekends, during which they are used more evenly than weekdays,

the influence of the washing machine on the water consumption pattern may be less than 10%.

Same as Figure 10, all information (except LGEUM) in the Figure 11 are referenced to (Blokker, 2010) rather than regenerated. In Figure 11, the blue, orange and green lines represent the original EPANET model, LGEUM and SIMDEUM, respectively. The water age of SIMDEUM and LGEUM are calculated using the methodology described in section 2.7. The light blue area represents the 90% confidence interval of LGEUM. The confidence intervals are obtained from multiple simulations of the LGEUM, employing diverse initial latent vectors, representing the uncertainty of the model (not to variation). Since pySIMDEUM contains only one pattern, the LGEUM trained on it lacks pattern variation as well. There are additional patterns are evident in alternative versions of SIMDEUM, offering further opportunities for the enhancement of LGEUM.

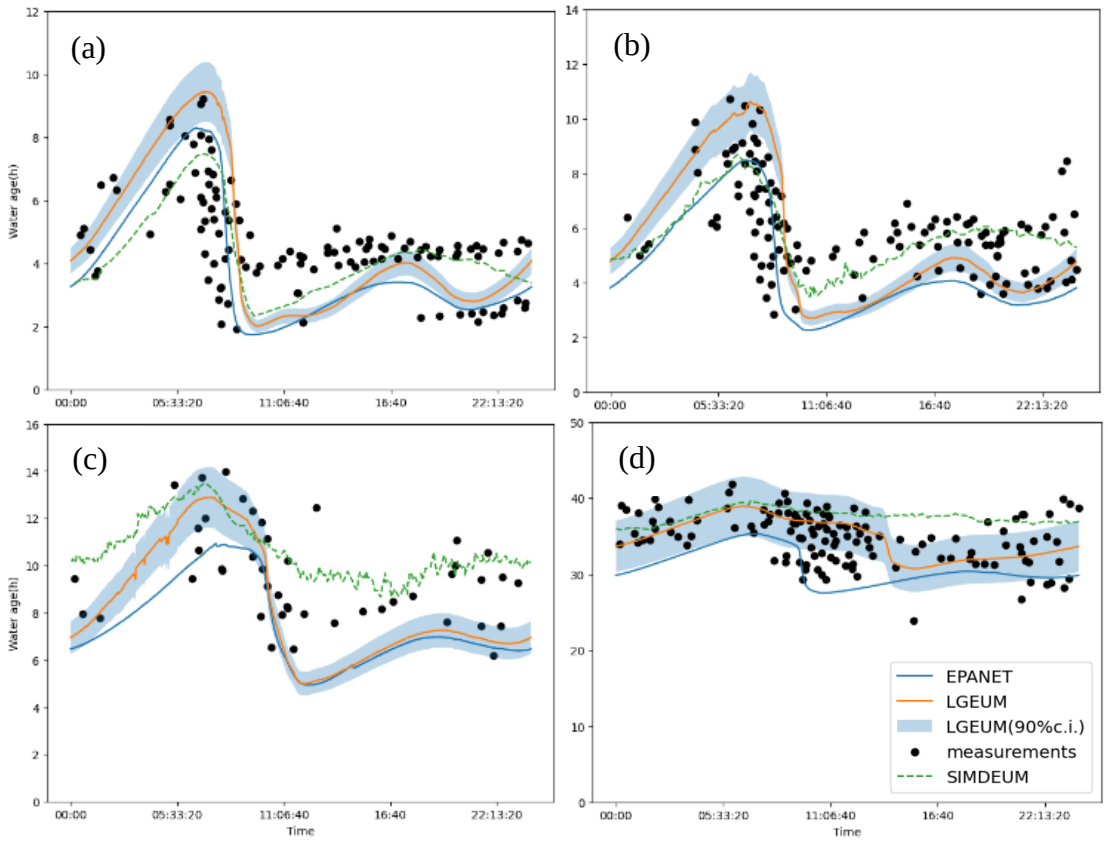


Figure 11. Measured residence time and modelled water age at 4 different locations in the Zandvoort area: (a) location 1; (b) location 2; (c) location 3; (d) location 4 (Blokker, 2010)

In Figure 11 (a), (b)&(c), it can be seen that GAN maintains a similar trend. In Figure 11 (a)&(b), GAN has a longer water age in the morning and similar water age in the

afternoon comparing with EPANET and SIMDEUM. In Figure 11(c), GAN is very similar to EPANET because location 3 is the NH hotel and GAN keeps the pattern of NH hotel from EPANET model. Figure 4(d) shows the water age of an apartment building of 15 residences and GAN covers almost all measurement points at this position. Overall, the GAN matches the measurement points better than the original EPANET model for the Zandvoort area but weaker than SIMDEUM. This indicates that the bottom-up water demand model can better simulate water demand, while the pattern diversity of GAN still needs to be improved. Since the actual composition of the measurements is not known, it remains to be verified whether the GAN is correct for the composition of the total water consumption.

3.4. Extensibility validation

In this section, extensibility of the GAN is validated with the REU2016 dataset from the United States. The GAN is trained with the same neural network structure and data preprocessing for 15 more epochs on top of the original model. Each epoch takes about 400s, with a total of 6000s. Due to the incompleteness of the REU2016 data on user, there is no extension validation of the user behavior model in this section. In the experiment, it was found that the outside tap has an excessive duration in the REU2016 dataset (Table 15), with an average of 1972s. Therefore, the resolution of outside tap in the REU2016 dataset is corrected to 20s instead of 4s while other end uses maintains the same resolution. It is recommended to introduce an automatic resolution adjustment mode in the data pre-procession part, which allows the resolution to be adjusted according to the event length. By introducing this mode, the GAN model can better process raw data.

Table 15 presents a comprehensive comparison of intensity and duration between GAN and REU2016 measurements. The end use exhibiting the most substantial error in Table 15 is the bathtub, with an error rate of 21.22%. A significant contributing factor to this disparity is the fixed pattern attributed to the bathtub in SIMDEUM, whereas it exhibits variability in REU2016. This results in a fact that the bathtub lacks training (too much modular training) although bathtub has a low error in Table 12. This highlights the necessity for enhancing GAN specifically for end uses that exhibit limited diversity. In REU2016, kitchen taps and bathroom taps are collectively referred to as faucets. This change reduces conflicts between similar end uses and significantly increases the

sample size of faucet. As a result, faucet has the lowest RE (-4.05%) in Table 15. In REU2016, the outside tap is not only used for irrigation, but also for other purposes such as car washing, which has a huge demand of water, with longer duration and higher intensity. The distinctive features result in a lower error for the simulation of outside taps compared with Table 12. Despite the small percentage of toilets and showers in REU2016, small errors were achieved in Table 15. One possible reason is the higher similarity of the toilet and shower events to SIMDEUM.

Table 15. Comparison of event intensity and duration between GAN and measurements (meas)

	Intensity(L/s)			**Duration(s)**		
	meas	GAN	RE(%)	meas	GAN	RE(%)
Toilet	0.158	0.133	-15.8%	65.34	72.58	11.1%
Bathtub	0.284	0.278	-2.11%	276.77	342.75	23.6%
Faucet	0.045	0.046	2.22%	38.59	36.11	-6.43%
Outside Tap	0.328	0.297	-9.45%	1972.16	1882.64	-4.54%
Shower	0.128	0.151	18.0%	477.16	379.18	-20.5%
Overall RMSE			11.6%			15.2%

	Volume(L)		
	meas	GAN	RE(%)
Toilet	10.32	9.65	-6.49%
Bathtub	78.6	95.28	21.22%
Faucet	1.73	1.66	-4.05%
Outside Tap	646.86	559.14	-13.56%
Shower	61.07	57.26	-6.23%
Overall RMSE			12.09%

The water demand probability distribution for two end uses with the lowest volume error (faucet and shower) are showed in Figure 12. The JS divergence for Figure 12(a) and (b) is 0.095 and 0.290, respectively. It can be seen from the values in Table 15 and the shape in Figure 12, GAN has substantial changes to adapt to REU2016. GAN simulations for REU2016 have smaller errors in intensity, duration and volume compared to simulations for SIMDEUM. Most end uses have a RE lower than 10%. Both the bathtub and outside tap have low weight in the model, and thus better results can be expected from GAN if weighted RMSE is introduced. In general, compared to

SIMDEUM, GAN has better results on REU2016 in this validation. Two possible reasons are: 1. The training of REU2016 is based on SIMDEUM and more training epochs may bring better results; 2. GAN is more suitable for changeful datasets. These conclusions need to be further verified.

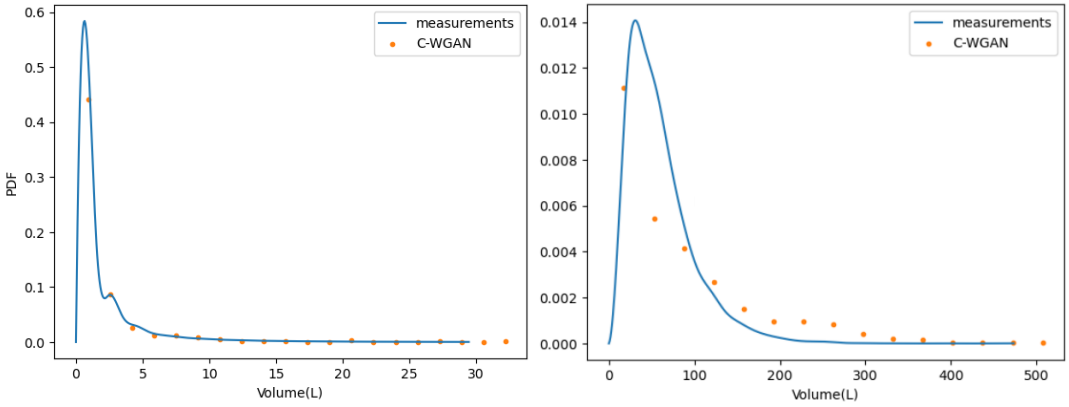


Figure 12. Water consumption probability distribution of (a) Faucet; (b) Shower

Facing water scarcity, developing countries often do not have the capacity to conduct large-scale research (Sivakumaran et al., 2010). The fine-tuning and raw data processing ability of GAN may help developing countries achieve better water management at a lower cost. On the basis of a generic GAN model in the further, the developing country can obtain a model adapted to the local dataset through small-scale data collection and fine-tuning training.

3.5 Uncertainty analysis

The LGEUM exhibits two primary sources of uncertainty. The first pertains to the absence of washing machines and dishwashers, which may have influenced the results by less than 10%. Although the impact of these appliances on total water consumption is limited, their influence on water usage patterns remains unclear, warranting further investigation. The second source of uncertainty arises from the lack of variation within the LGEUM. Certain end uses, such as bathtubs, maintain fixed patterns within the LGEUM. Additionally, the users in LGEUM lack diversity, as only one pattern is accessible in the current version of pySIMDEUM. It is advisable to incorporate more diverse patterns into LGEUM for training purposes, thereby mitigating the uncertainty attributed to variability.

4. Conclusion

A LSTM-GAN end use model (LGEUM) is developed in this thesis. For the research question, the LGEUM simulates the end use demand well based on the SIMDEUM dataset. LGEUM attains an overall RMSE of 18% and 4.2% for the end-use model and user behavior model, respectively, when evaluated against SIMDEUM. Additionally, LGEUM exhibits substantial similarity in water consumption distributions, further corroborating its effectiveness. In validation with measured data, LGEUM maintains an R^2 score above 0.7 on all time scales. The LGEUM shows good agreements with SIMDEUM and Zandvoort (Netherlands) measurements both numerically and in terms of distribution. In addition, LGEUM has good extensibility and achieves a 12% overall RMSE on the REU2016 dataset (US), which is even lower than that of SIMDEUM. In terms of computational speed, the GAN model improves 546% compared to SIMDEUM. However, LGEUM is still lacking in terms of variation and simulation of W&Ds. It is recommended to introduce more patterns in LGEUM to make the model more diverse and realistic. As new end uses and features are introduced in the future, a better way to balance the sample size is worth investigating.

LGUEM presents viable solutions to several constraints found in conventional models. The GAN model establishes its fundamental logic through extensive training on the SIMDEUM dataset, while enabling users to further train the GAN model with local databases, thereby surpassing spatial limitations. Moreover, the GAN model allows the utilization of training sets with varying resolutions, thereby alleviating temporal constraints. Additionally, by simulating all events in a day simultaneously, the GAN model effectively addresses the computational limitations encountered in traditional models. Above all, the strong adaptability, fast computational speed and simple update methods make GAN models promising in the field of end-use water modeling.

5. Acknowledgements

I would like to use this section to thank **Mirjam Blokker** & **Riccardo Taormina** for their patience and suggestions.

In addition, I would like to thank my colleagues at KWR and Filippo Mazzoni for their help.

6. Reference

Aggarwal, A., Mittal, M., & Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. International Journal of Information Management Data Insights, 1(1), 100004. doi:https://doi.org/10.1016/j.jjimei.2020.100004

Arjovsky, M., & Bottou, L. J. a. p. a. (2017). Towards principled methods for training generative adversarial networks.

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. Paper presented at the International conference on machine learning.

Bailey, O., Zlatanovic, L., van der Hoek, J. P., Kapelan, Z., Blokker, E., Arnot, T., & Hofman, J. (2020). A Stochastic Model to Predict Flow, Nutrient and Temperature Changes in a Sewer under Water Conservation Scenarios. 12(4), 1187.

Barberán, R., Inmaculada, V., & Gracia, F. (2000). Water Price Impact On Residential Water Demand In Zaragoza City. A Dynamic Panel Data Approach.

Blokker, M. (2010). Stochastic water demand modelling for a better understanding of hydraulics in water distribution networks.

Blokker, M. (2011). Stochastic water demand modelling: IWA publishing.

Blokker, M., Agudelo-Vera, C., Moerman, A., Thienen, P., & Pieterse-Quirijns, I. (2017). Review of applications for SIMDEUM, a stochastic drinking water demand model with a small temporal and spatial scale. Drinking Water Engineering and Science, 10, 1-12. doi:10.5194/dwes-10-1-2017

Blokker, M., Pieterse-Quirijns, E. J., Vreeburg, J., & Dijk, J. (2011). Simulating Residential Water Demand with a Stochastic End-Use Model. Journal of Water Resources Planning and Management, 137, 511-520. doi:10.1061/(ASCE)WR.1943-5452.0000002

Blokker, M., Vreeburg, J., Buchberger, S., Van Dijk, J. J. D. W. E., & Science. (2008). Importance of demand modelling in network water quality models: a review. 1(1), 27-38.

Cominola, A., Giuliani, M., Castelletti, A., Abdallah, A., & Rosenberg, D. E. (2016). Developing a stochastic simulation model for the generation of residential water end-use demand time series.

Creaco, E., Farmani, R., Kapelan, Z., Vamvakeridou-Lyroudia, L., & Savic, D. (2015). Considering the Mutual Dependence of Pulse Duration and Intensity in Models for Generating Residential Water Demand. Journal of Water Resources Planning and Management, 141, 04015031. doi:10.1061/(ASCE)WR.1943-5452.0000557

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative Adversarial Networks: An Overview. IEEE Signal Processing Magazine, 35(1), 53-65. doi:10.1109/MSP.2017.2765202

DeOreo, W. B., Mayer, P. W., Dziegielewski, B., & Kiefer, J. (2016). Residential end uses of water, version 2: Water Research Foundation.

Duncan, H. P., & Mitchell, V. G. (2008). A Stochastic Demand Generator for Domestic Water Use. Modbury, SA. https://search.informit.org/doi/10.3316/informit.568150280628065

Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F., & Alcamo, J. (2013). Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study. Global Environmental Change, 23(1), 144-156. doi:https://doi.org/10.1016/j.gloenvcha.2012.10.018

Gato-Trinidad, S., Jayasuriya, N., & Roberts, P. (2011). Understanding urban residential end uses of water. Water science and technology: a journal of the International Association on Water Pollution Research, 64, 36-42. doi:10.2166/wst.2011.436

Ghalehkhondabi, I., Ardjmand, E., Young, W. A., & Weckman, G. R. (2017). Water demand forecasting: review of soft computing methods. Environmental Monitoring and Assessment, 189(7), 313. doi:10.1007/s10661-017-6030-3

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. J. A. i. n. i. p. s. (2017). Improved training of wasserstein gans. 30.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. J. A. i. n. i. p. s. (2014). Generative adversarial nets. *27*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2020). Generative adversarial networks. *63*(11 %J Commun. ACM), 139–144. doi:10.1145/3422622

Jabbar, A., Li, X., & Omar, B. (2021). A survey on generative adversarial networks: Variants, applications, and training. Computing Surveys, 54(8), 1-49.

Jiuxiang, G., Zhenhua, W., Jason, K., Lianyang, M., Amir, S., Bing, S., . . . Tsuhan, C. (2018). Recent advances in convolutional neural networks. Pattern Recognition, 77, 354-377. doi:https://doi.org/10.1016/j.patcog.2017.10.013

Madias, K., & Szymkowiak, A. (2022). Residential Sustainable Water Usage and Water Management: Systematic Review and Future Research. Water, 14(7). doi:10.3390/w14071027

Makki, A. A., Stewart, R. A., Beal, C. D., & Panuwatwanich, K. (2015). Novel bottom-up urban water demand forecasting model: Revealing the determinants, drivers and predictors of residential indoor end-use consumption. Resources, Conservation and Recycling, 95, 15-37. doi:https://doi.org/10.1016/j.resconrec.2014.11.009

Mayer, P. W., DeOreo, W. B., Opitz, E. M., Kiefer, J. C., Davis, W. Y., Dziegielewski, B., & Nelson, J. O. (1999). Residential end uses of water.

Mazzoni, F., Alvisi, S., Blokker, M., Buchberger, S. G., Castelletti, A., Cominola, A., . . . Franchini, M. (2023). Investigating the characteristics of residential end uses of water: A worldwide review. Water Research, 230, 119500. doi:https://doi.org/10.1016/j.watres.2022.119500

Mazzoni, F., Alvisi, S., Franchini, M., & Blokker, E. (2023). Exploiting high-resolution data to investigate the characteristics of water consumption at the end-use level: A Dutch case study. Water Resources and Industry, 29, 100198. doi:https://doi.org/10.1016/j.wri.2022.100198

Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets.

Mitchell, V. G., Duncan, H., Inman, M., Rahilly, M., Stewart, J., Vieritz, A., . . . Coleman, J. (2007). State of the art review of integrated urban water models. Paper presented at the NOVATECH 2007-6ème Conférence sur les techniques et stratégies durables pour la gestion des eaux urbaines par temps de pluie/6th International Conference on sustainable techniques and strategies for urban water management.

Rathnayaka, K., Malano, H., Arora, M., George, B., Maheepala, S., & Nawarathna, B. (2017a). Prediction of urban residential end-use water demands by integrating known and unknown water demand drivers at multiple scales I: Model development. Resources, Conservation and Recycling, 117, 85-92. doi:https://doi.org/10.1016/j.resconrec.2016.11.014

Rathnayaka, K., Malano, H., Arora, M., George, B., Maheepala, S., & Nawarathna, B. (2017b). Prediction of urban residential end-use water demands by integrating known and unknown water demand drivers at multiple scales II: Model application and validation. Resources, Conservation and Recycling, 118, 1-12. doi:https://doi.org/10.1016/j.resconrec.2016.11.015

Rathnayaka, K., Malano, H., Maheepala, S., Nawarathna, B., George, B., & Arora, M. (2011). Review of residential urban water end-use modelling. Paper presented at the 19th international congress on modelling and simulation.

Sivakumaran, S., & Aramaki, T. (2010). Estimation of household water end use in Trincomalee, Sri Lanka. Water International, 35(1), 94-99. doi:10.1080/02508060903533476

Steven, G. B., & Zhiwei, L. (2007). PRPsym: A Modeling System for Simulation of Stochastic Water Demands. World Environmental and Water Resources Congress 2007, 1-13. doi:doi:10.1061/40927(243)511

Wang, K., & Davies, E. G. R. (2018). Municipal water planning and management with an end-use based simulation model. Environmental Modelling & Software, 101, 204-217. doi:https://doi.org/10.1016/j.envsoft.2017.12.024

Willis, R., Stewart, R., Panuwatwanich, K., Capati, B., & Giurco, D. (2009). Gold Coast

Domestic Water End Use Study. Water, 36.

Willis, R. M., Stewart, R. A., Giurco, D. P., Talebpour, M. R., & Mousavinejad, A. (2013). End use water consumption in households: impact of socio-demographic factors and efficient devices. Journal of Cleaner Production, 60, 107-115. doi:https://doi.org/10.1016/j.jclepro.2011.08.006

Yan-Lin, H., Xing-Yuan, L., Jia-Hui, M., Shan, L., & Qun-Xiong, Z. (2022). A novel virtual sample generation method based on a modified conditional Wasserstein GAN to address the small sample size problem in soft sensing. Journal of Process Control, 113, 18-28. doi:https://doi.org/10.1016/j.jprocont.2022.03.008

Yang, A., Zhang, H., Stewart, R. A., & Nguyen, K. (2018). Enhancing Residential Water End Use Pattern Recognition Accuracy Using Self-Organizing Maps and K-Means Clustering Techniques: Autoflow v3.1. Water, 10(9). doi:10.3390/w10091221

Yu, Y., Srivastava, A., Canales, S. J. A. T. o. M. C., Communications,, & Applications. (2021). Conditional lstm-gan for melody generation from lyrics. 17(1), 1-20.

Zhou, L., Zhao, C., Liu, N., Yao, X., & Cheng, Z. (2023). Improved LSTM-based deep learning model for COVID-19 prediction using optimized approach. Engineering Applications of Artificial Intelligence, 122, 106157. doi:https://doi.org/10.1016/j.engappai.2023.106157

7. Appendix

7.1 Extra Equations

KL divergence: $$\mathrm{D}_{KL}(\mathrm{P} \parallel \mathrm{Q}) = -\sum_i P(i) ln(\frac{Q(i)}{P(i)}) \quad (1)$$

JS divergence: $$\mathrm{D}_{JS}(\mathrm{P} \parallel \mathrm{Q}) = \frac{1}{2}\mathrm{D}_{KL}\left(\mathrm{P(x)} \parallel \frac{\mathrm{P(x)+Q(x)}}{2}\right) + \frac{1}{2}\mathrm{D}_{KL}\left(\mathrm{Q(x)} \parallel \frac{\mathrm{P(x)+Q(x)}}{2}\right) \quad (2)$$

Loss function of GAN:

$$\mathrm{Loss}_{discriminator} = -\mathbb{E}_{x\sim P_r}\left[\log\left(D(x)\right)\right] - \mathbb{E}_{\boldsymbol{x\sim P_g}}\left[\mathbf{log}\left(\mathbf{1}-\mathbf{D}(\mathbf{x})\right)\right] \quad (3)$$

$$\mathrm{Loss}_{generator} = -\mathbb{E}_{\boldsymbol{x\sim P_g}}\left[\mathbf{log}\left(\boldsymbol{D}(\boldsymbol{x})\right)\right] \quad (4)$$

Wasserstein Distance:

$$\mathrm{W}\left(P_r, P_g\right) = inf_{\gamma\sim\prod(P_r,P_g)}\mathbb{E}_{(x,y)\sim\gamma}[\parallel x-y \parallel] \quad (5)$$

7.2 Two Type of Collapse in GAN

When we make the derivative of Equation (3) with respect to D(x) equal to 0, the optimal discriminator $D^*(x)$ is obtained as：

$$-\frac{P_r(x)}{D(x)} + \frac{P_g(x)}{1-D(x)} = 0$$

$$D^*(x) = \frac{P_r(x)}{P_r(x)+P_g(x)} \quad (6)$$

Substituting Equation (6) into Equation (3) and simplifying Equation (3) to:

$$\mathrm{Loss}_d = -\mathbb{E}_{x\sim P_r}\left[\frac{P_r(x)}{\frac{1}{2}\cdot\left(P_r(x)+P_g(x)\right)}\right] - \mathbb{E}_{x\sim P_g}\left[\frac{P_g(x)}{\frac{1}{2}\cdot\left(P_r(x)+P_g(x)\right)}\right] + 2\cdot\log 2$$

$$= -\mathbf{2} \cdot \boldsymbol{JS}\left(\mathbf{P}_{\boldsymbol{r}} \parallel \mathbf{P}_{\boldsymbol{g}}\right) + \mathbf{2} \cdot \mathbf{log}\, \mathbf{2} \tag{7}$$

From Equation (7), it can be seen that minimizing the loss function defined by the original GAN is equivalent to minimizing the JS divergence between the real distribution and the fake distribution under the condition of the optimal discriminator. The more the discriminator is trained, the closer the loss function is to Equation (7). One would deduce that optimize the discriminator first, and then optimize the generator based on the optimized discriminator can achieve the overall optimization. But in reality, the better the discriminator is trained, the worse the training of the generator would be. This is because the support sets of P_r and P_g are often low-dimensional manifolds in high-dimensional space. For example, a common method to obtain 128-dimensional generated data is that we first obtain a 32-dimensional support set from the prior distribution and extend it to 128 dimensions by means of a neural network. Obviously, a 32-dimensional support set (manifold) is not enough to fill a 128-dimensional space. The manifolds lead to a discontinuous distribution in the high-dimensional space and the probability of P_r and P_g perfect align is almost 0 (Arjovsky et al., 2017). This leads to $\mathbf{JS}\left(\mathbf{P}_{\boldsymbol{r}} \parallel \mathbf{P}_{\boldsymbol{g}}\right) = \mathbf{log2}$ **or** $\rightarrow \infty$ and the optimization for the generator based on the gradient descent does not work in this situation (Arjovsky et al., 2017):

$$\lim_{||D-D^*||\rightarrow 0} \nabla_\theta \, \mathbb{E}_{z\sim \mathrm{P(z)}}[\log(1 - \mathrm{D}(g_\theta(x)))] = 0 \tag{8}$$

Equation (8) is the direct cause of the gradient vanishing.

Another instability of GAN comes from model collapse, also called mode dropping. When D* (Equation 6) is substituted into $\mathrm{KL}\left(\mathrm{P}_g \parallel \mathrm{P}_r\right)$ (Equation 1):

$$\mathrm{KL}(\mathrm{P}_g \parallel \mathrm{P}_r) = \mathbb{E}_{x\sim P_g}\left[\log \frac{P_g}{P_r}\right] = \mathbb{E}_{x\sim P_g}\left[\log \frac{\frac{P_g}{P_r+P_g}}{\frac{P_r}{P_r+P_g}}\right]$$

$$= \mathbb{E}_{x\sim P_g}[\log(1 - D^*(x))] - \mathbb{E}_{x\sim P_g}[\log(D^*(x))] \tag{9}$$

Introduce Equation (7) into Equation (9) and organize:

$$-\mathbb{E}_{x\sim P_g}\left[\log\left(D^*(x)\right)\right] = \mathbf{KL}(\mathbf{P}_{\boldsymbol{g}} \parallel \mathbf{P}_{\boldsymbol{r}}) - \mathbf{2} \cdot \boldsymbol{JS}(\mathbf{P}_{\boldsymbol{g}} \parallel \mathbf{P}_{\boldsymbol{r}})$$

$$+2 \cdot log2 + \mathbb{E}_{x\sim P_r}\left[\log\left(D^*(x)\right)\right] \quad (10)$$

The left side of Equation (10) is the generator's loss and only the first two terms on the right side depend on P_g. In these two terms, JS divergence is symmetrical but KL divergence is not. $KL(P_g||P_r)$ penalizes untruthful samples greatly. Therefore, the model tends to generate safer samples and gradually loses diversity (Arjovsky et al., 2017).

7.3 Solution of WGAN

In equation 5, arbitrarily sample the joint distribution γ from $\prod(P_r - P_g)$ and calculate the mathematical expectation of the distance between the real samples and the fake samples based on γ. Among all joint distributions, the infimum of this mathematical expectation is the Wasserstein distance.

However, $inf_{\gamma\sim\prod(P_r,P_g)}$ in the equation (5) is not directly solvable. Therefore, the authors of WGAN transform equation (5) based on Kantorovich-Rubinstein Duality as (Arjovsky et al., 2017):

$$W(P_r, P_g) = \frac{1}{K} \cdot sup_{\|f\|_L\leq K}\left(\mathbb{E}_{x\sim P_r}[f(x)] - \mathbb{E}_{x\sim P_g}[f(x)]\right)$$

$$\approx \frac{1}{K} \cdot max_{\omega:\|f_\omega\|_L\leq K}\left(\mathbb{E}_{x\sim P_r}[f(x)] - \mathbb{E}_{x\sim P_g}[f(x)]\right) \quad (11)$$

Where $\| f \|_L$ is Lipschitz constant. Equation (13) can be solved by using a neural network with parameter ω to substitute f_ω. It is important to note that Equation (11) must satisfy the Lipschitz continuity, which is $\| f_\omega \|_L \leq K$. In the WGAN, the absolute value of the parameters will be clipped to no more than a fixed constant *c* in order to satisfy the restriction (K can be any value but can't be infinite). However, this method of weight clipping is prone to the following two problems (Gulrajani et al., 2017): 1. parameter bipolarity (convergence to c or -c); 2. the choice of constant *c* directly affects the model effect, and may lead to gradient vanishing and gradient explosion.

Gulrajani et al.(2017) proposed a novel structure of WGAN with a gradient penalty(GP) instead of weight clipping (Gulrajani et al., 2017). The gradient penalty forces the discriminator network to satisfy the 1- Lipschitz continuity. It was also found that

constraining the L2 norm of the gradient around 1 performed best (Gulrajani et al., 2017). The loss function in WGAN-GP is thus calculated as:

$$\text{Loss}_{total} = \mathbb{E}_{\tilde{x}\sim P_g}[\text{D}(\tilde{x})] - \mathbb{E}_{x\sim P_r}[\text{D}(x)] + \lambda \cdot \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}[(\| \nabla_{\hat{x}}\text{D}(\hat{x}) \|_2 - 1)^2] \quad (12)$$

In equation (14), the first two items are the Wasserstein distance and the third item is the gradient penalty; $\tilde{x}$ represents the fake samples; $\hat{x}$ represents samples from the interpolated space of the real and fake data, $\hat{x} = \varepsilon\tilde{x} + (1-\varepsilon)x$, where $\varepsilon \in [0,1]$.