

## A Python Tool for Generating Synthetic Demand Scenarios

Simone Santopietro<sup>1</sup>, Francesco Granata<sup>2</sup>, Carla Tricarico<sup>3</sup>,  
Giovanni de Marinis<sup>4</sup>, Rudy Gargano<sup>5</sup>

<sup>1,2,3,4,5</sup> Università di Cassino e del Lazio Meridionale

Via G. Di Biasio 43, 03043 Cassino - Italy

<sup>1</sup>s.santopietro@unicas.it

### ABSTRACT

*Demand modelling has a great impact on the outcome of hydraulic simulations of water distribution systems. Its stochastic nature has to be considered in order to obtain more reliable results. A stochastic approach, based on a mixed probability distribution, able to model residential water demand has been implemented in Python. This distribution considers both the event of null and not null water demand and allows to estimate the statistical parameters, solely on the number of supplied users and on the related average demand pattern. This approach is well suited to be used in hydraulic simulation workflows (e.g. using the EPANET toolkit) and a Python implementation could encourage a more refined demand characterization also in non-scientific environments.*

**Keywords:** Demand modelling, Stochastic approach, Python

## 1 BACKGROUND

Recently Water Distribution Systems (WDS) simulation models are becoming more and more refined, incorporating for example pressure driven simulation, or leakage models. The main and widely used hydraulic simulation software nowadays is EPANET 2 [1]. This software has recently been transformed into an open source project and since then many efforts have been made to simplify the code structure and to make it more scalable. This would allow to easily extend the software features as scientific research moves forward.

When engineers perform hydraulic simulations, they need to consider many variables: geometric, hydraulic and geographic. Among these, water demand influences deeply the performances of WDSs and a more robust representation of it, taking into account its stochastic nature, would lead to more robust results.

In scientific literature, many contributions have been made toward a more accurate characterization of the water demand. In the last two decades, end user approaches have gained popularity. In these approaches water demand is represented as a superimposition of single rectangular pulses following stochastic processes, like the Poisson Rectangular Pulse (PRP - [2]), or the Neyman-Scott Rectangular Pulse (NSRP - [3], [4]). These approaches need to model the phenomena at least through three stochastic variables: frequency, duration and intensity of the residential demand, which contribute to define the water demand for each user. In these approaches, a detailed knowledge of the users is required thus the parameters estimation process can be difficult. A simplification attempt in this direction has been made by [5], where the demand pulses are considered at the dwelling level.

Recently in [6] a stochastic approach for modelling residential water demand by means of a mixed probability distribution has been proposed. A mixed distribution is well suited to describe both the events of null and not-null water demand, that could occur for a small to medium-sized users' cluster. This approach has proven to have good performances for a number of users ranging from 200 to 1250.

Aim of this work is to implement the water demand modelling approach introduced by [6] into a Python tool capable to be easily included in hydraulic simulation workflows as an example in EPANET. This would allow to consider a more accurate stochastic characterization of the residential water demand, when using custom code implementation for WDS performance analysis.

## 2 METHODOLOGY

As it is well known, residential water demand follows daily trends according to the habits of the users connected to the WDS. These demand patterns present typical forms with spikes at the times of maximum consumption and minimums, that can assume null values, during the night hours.

Here we consider the dimensionless water demand by means of the demand coefficient:

$$C_D(t) = \frac{Q(t)}{\mu_Q} \quad (1)$$

where  $\mu_Q$  is the daily mean water demand, and  $Q(t)$  the value of water demand at the time  $t$ . If we consider the Cumulated Distribution Function (CDF) of  $C_D$  applying the total probability theorem we can explicit two components:  $F_0$  the probability of null consumption, and  $F^*$  that represents the CDF of the demand coefficient when it is not null.

$$F[C_D(t)] = F_0(t) + (1 - F_0(t))F^*(t) \quad (2)$$

Understandably the peculiarity of the users' habits affects significantly the probability of  $C_D(t) = 0$ . In addition [7] demonstrated that the  $F_0$  depends significantly on the discretization time interval  $\Delta t$  and on the number of users supplied. The approach proposed by [6] and here referenced was developed and tested for  $\Delta t = 1$  minute and for a number of users ranging in the interval 200÷1250. To estimate the probability of null consumption for each time of the day, an empirical equation has been proposed.

$$F_0(t) = \exp \left[ -5 \frac{N_{us}}{1000} \mu_{C_D}(t) \right] \quad (3)$$

Eq.(3) depends on the number of users as above mentioned and on the entity of the mean demand coefficient for every particular time of the day  $\mu_{C_D}(t)$ .

The event of  $F_o(t) = 1$ , in other words the probability of null demand is a certain event, it is a condition theoretically never reachable. Therefore, a relation contemplating a maximum value for  $F_o$  has been retrieved from observation on experimental data:

$$F_o^{max} = 1 - 0.25 \left( \frac{N_{us}}{1000} \right)^{2,5} \quad (4)$$

For the  $F^*$  CDF a logistic distribution has been considered, introduced by [8]

$$F_{C_D}[C_D(t)] = \left[ \exp \left( - \frac{\pi}{\sqrt{3}} \frac{C_D(t) - \mu_{C_D}(t)}{\sigma_{C_D}(t)} \right) + 1 \right]^{-1} \quad (5)$$

The logistic CDF has the peculiarity of having an integrable probability density function and for this particular application has to be truncated, being that  $C_D > 0$ . The considered distribution is bi-parametric and thus only the mean  $\mu_{C_D}$  and standard deviation  $\sigma_{C_D}$  have to be estimated.

The mean demand coefficient  $\mu_{C_D}(t)$  is an input of the approach, being very specific to the considered users' habits. As for the second parameter, the variation coefficient has been considered  $CV = \mu_{C_D}/\sigma_{C_D}$  and Eq.(6) allows to estimate it in relation to the mean and number of users.

$$CV(t) = 0,1 + \frac{6}{\left( \frac{1}{4} \mu_{C_D}(t) \cdot N_{us} \right)^{3/4}} \quad (6)$$

Combining Eq.(2) with Eq.(5) and the knowledge of the parameters  $\mu_{C_D}(t)$ ,  $CV(t)$  and  $F_o(t)$  (Eq.(3) and (6)) it is possible to generate synthetic demand pattern with the following relation:

$$Q(t) = \mu_Q C_D(t) = \mu_Q \left[ 1 - \frac{\sqrt{3}}{\pi} \ln \frac{1-F(t)}{F(t)-F_o(t)} CV(t) \right] \mu_{C_D}(t) \quad \text{for } F(t) > F_o(t) \quad (7)$$

### 3 THE ALGORITHM

The implementation of the above presented approach has been performed in Python. This choice was motivated by the increasing popularity of this programming language, in general and in water research (e.g. [9]). Its simplicity and velocity in the development process make it a great tool for quickly prototyping algorithms. In order to make the source code easily available to the widest possible spectrum of users, the PEP 8 [10] writing style standard has been adopted. The algorithm was implemented as a module to be easily available in the Python environment with the common *import* statement.

The main inputs of the algorithm are the mean demand coefficient pattern  $\mu_{C_D}(t)$  and the number of users  $N_{Us}$ .  $\mu_{C_D}(t)$  is a vector with length equal to the number of discretization intervals of one day (e.g. if  $\Delta t = 1\text{min}$   $\text{length}(\mu_{C_D}(t)) = 1440$ ). For each considered cluster of users, the following steps are performed:

- The model parameters are calculated:  $F_o(N_{Us}, \mu_{C_D}(t))$  and  $CV(N_{Us}, \mu_{C_D}(t))$ ;  
Considering  $N_{Us}$  and the vector  $\mu_{C_D}(t)$ , by means of Eq.s(3) and (6), it is possible to calculate  $F_o$  and CV, that have the same length of  $\mu_{C_D}(t)$
- For every time step a value of the probability F is generated in the interval (0,1);
- The generated probability is then confronted with the corresponding  $F_o(t)$ ;
- If  $F(t) < F_o(t)$  the resulting  $C_D(t) = 0$ , otherwise it will be calculated according to Eq.(7).

Then a vector containing a synthetic daily demand coefficient pattern is returned for the specific inputted pattern and the considered  $N_{Us}$ .

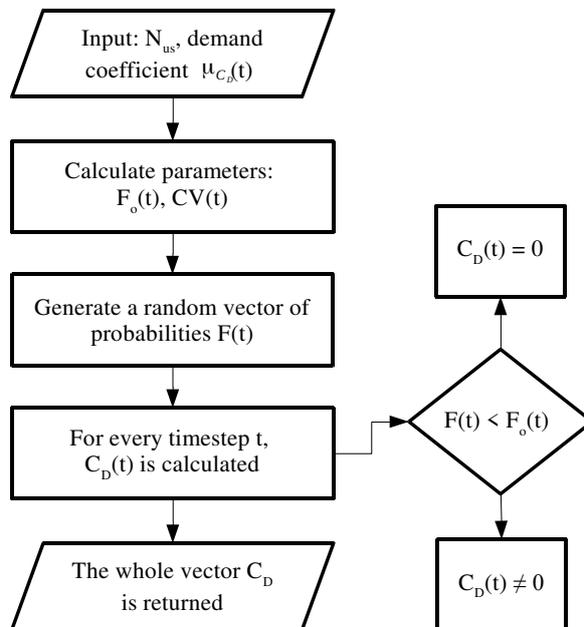


Figure 1. Flow chart of the implemented algorithm

The Python module is called `mixedmodel` and the function used to generate the pattern is called `generate_pattern` (Figure 2).

```
def generate_pattern(pat,n_users,n_days,add_stat = False):
```

Figure 2. `generate_pattern` function declaration.

The arguments needed are respectively  $\mu_{C_D}(t)$ ,  $N_{Us}$ , the number of days for which the generation is needed and an additional parameter (`add_stat`) that if it is set to `TRUE` returns the calculated parameters ( $F_o(t)$  and CV).

As already underlined above, the software EPANET is one of the most used to perform water simulations, in practical applications. Specifically, when the aim is to perform multiple elaborations, for example in heuristic or stochastic analysis, it is necessary to implement ad hoc codes, that invoke the EPANET toolkit recursively. A general example of the EPANET toolkit implementation, using a Python wrapper [11] is given in Figure 3.

```

import epamodule as em

# Open the EPANET toolkit
em.ENopen(infile,rptfile)

# Generate and set pattern

# Open the hydraulic solver system
em.ENopenH()

# Initialize the network solver
em.ENinitH()

# Run an hydraulic simulation step by step
step = 1

while step > 0:
    err = em.ENrunH()
    step = em.ENnextH()

    # Save output of the simulation

# Close the hydraulic solver
em.ENcloseH()

# Close the toolkit
em.ENclose()

```

*Figure 3. Pseudocode for the EPANET toolkit implementation*

The above implemented function can be used in the common adopted workflow using the toolkit, when setting the parameters for the next simulation (by means of the `ENsetpattern()` statement).

About the parameters needed for the pattern generation, a clarification has to be made. The number of users can be known from municipal archives where available, otherwise can be estimated by GIS methodologies such as proximity analysis or graph theory algorithms performed on national geographical databases (e.g. real estate registries or national statistic institutions). The demand pattern on the other hand can be either deduced by means of in-situ measurements or estimated considering the following characteristics:

- $\frac{1}{T} \int_0^T \mu_{CD}(t) dt = 1$
- several equations in technical literature allow to estimate the peak phenomenon in relation to the user number or the minimum demand (e.g. [12], [13]);

## 4 A PRACTICAL APPLICATION

By means of this approach and its implementation, it is possible to quickly generate synthetic demand scenarios, for example within Monte Carlo simulations. The mean demand coefficient pattern and its shape are very specific features of the considered users and thus the  $\mu_{Cd}(t)$  is required as an input. The same goes for the number of users we are willing to model, because the considered statistics are strongly dependent on it.

As an example, here are presented the results obtained from the application of this implementation on the case study of Castelfranco Emilia (CE) [3], an Italian town with a number of users equal to 596.

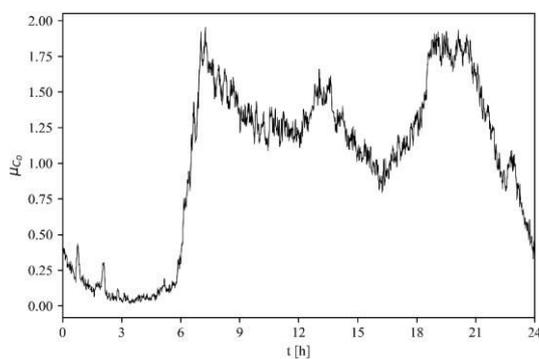


Figure 4. The mean demand coefficient for CE

In Figure 4 is shown the demand coefficient pattern averaged over an observation period of 50 days. This has been used as an input for the function along with the  $N_{us}$ . In the following Figure 5 the  $F_0$  and CV parameters are shown.

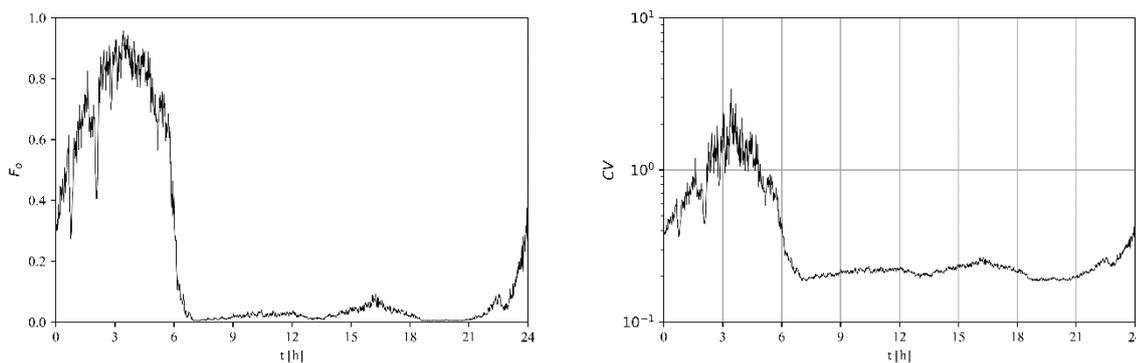


Figure 5. The probability of null demand and the coefficient of variation returned by the python function, setting the variable `add_stat = True`

As it is expected both the  $F_0$  and CV reach the highest values during the night hours. These parameters are then used to generate the demand patterns by mean of Eq.(7). In Figure 6 the output of 50-day generation is displayed with the mean generated demand coefficient. The generated patterns and the related mean have a time resolution of one minute and they are in good agreement with the mean demand coefficient of Figure 4.

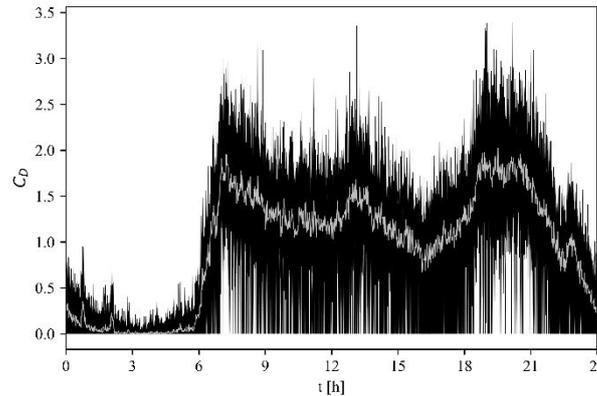


Figure 6. Demand pattern generation for 50 days with a time interval of 1 minute and related mean

## 5 CONCLUSIONS

A stochastic approach to model residential water demand has been implemented into a Python module. This approach allows to easily generate synthetic demand patterns for residential users ranging from 200 to 1250 with a time resolution of 1 minute. The required inputs are only the number of supplied users and a demand pattern congruent to their local habits. The water demand is modelled by means of a mixed distribution, able to contemplate at the same time a discrete random variable, that takes into account the null demand, and a continuous positive random variable, for the not null demand. This structure makes it very suitable to be implemented in hydraulic simulations of water distribution systems when a more accurate stochastic characterization of the water demand is required. This Python implementation allows to easily include this kind of approach inside water simulation workflows.

### List of symbols

$C_D(t)$  = [-] demand coefficient at the time  $t$

CDF = Cumulative Distribution Function

$CV(t)$  = [-] variation coefficient of the demand coefficient at the time  $t$

$F_o$  = [-] probability of occurrence of null request

$F^*$  = [-] CDF of the flow demand when it is different to zero

$F_o^{\max}$  = maximum value of the probability of null water demand

$N_{us}$  = [-] Number of users

$Q(t)$  = [ $L^3/T$ ] demanded water flow at time  $t$

$\mu_{C_D}(t)$  = [-] mean demand coefficient at the time  $t$

$\mu_Q$  = [ $L^3/T$ ] mean daily water demand

$\mu_Q(t)$  = [ $L^3/T$ ] mean water demand at the time  $t$

$\sigma_{C_D}(t)$  = [-] deviation standard of demand coefficient at the time  $t$

## References

- [1] L. A. Rossman, *EPANET 2 USERS MANUAL*. 2000.
- [2] S. Buchberger and L. Wu, “Model for Instantaneous Residential Water Demands,” *J. Hydraul. Eng.*, vol. 121, no. 3, pp. 232–246, Mar. 1995.
- [3] S. Alvisi, M. Franchini, and A. Marinelli, “A stochastic model for representing drinking water demand at residential level,” *Water Resour. Manag.*, vol. 17, no. 3, pp. 197–222, 2003.
- [4] V. H. Alcocer-Yamanaka, V. G. Tzatchkov, and F. I. Arreguin-Cortes, “Modeling of Drinking Water Distribution Networks Using Stochastic Demand,” *Water Resour. Manag.*, vol. 26, no. 7, pp. 1779–1792, May 2012.
- [5] R. Gargano, F. Di Palma, G. de Marinis, F. Granata, and R. Greco, “A stochastic approach for the water demand of residential end users,” *Urban Water J.*, vol. 13, no. 6, pp. 569–582, Aug. 2016.
- [6] R. Gargano, C. Tricarico, G. del Giudice, and F. Granata, “A stochastic model for daily residential water demand,” *Water Sci. Technol. Water Supply*, vol. 16, no. 6, pp. 1753–1767, Dec. 2016.
- [7] S. G. Buchberger and G. Nadimpalli, “Leak Estimation in Water Distribution Systems by Statistical Analysis of Flow Readings,” *J. Water Resour. Plan. Manag.*, vol. 130, no. 4, pp. 321–329, 2004.
- [8] P. K. Swamee, “Near lognormal distribution,” *J. Hydrol. Eng.*, vol. 7, no. 6, pp. 441–444, 2002.
- [9] D. Steffelbauer and D. Fuchs-Hanusch, “OOPNET: An object-oriented EPANET in Python,” *Procedia Eng.*, vol. 119, pp. 710–718, 2015.
- [10] “PEP 8 -- Style Guide for Python Code,” *Python.org*. [Online]. Available: <https://www.python.org/dev/peps/pep-0008/>. [Accessed: 03-Jul-2017].
- [11] *epanet-python: python wrapper for epanet library*. Open Water Analytics, 2017.
- [12] R. Gargano, C. Tricarico, F. Granata, S. Santopietro, and G. de Marinis, “Probabilistic Models for the Peak Residential Water Demand,” *Water*, vol. 9, no. 6, p. 417, Jun. 2017.
- [13] Javier Martínez-Solano, Pedro L. Iglesias-Rey, Rafael Pérez-García, and P. Amparo López-Jiménez, “Hydraulic Analysis of Peak Demand in Looped Water Distribution Networks,” *J. Water Resour. Plan. Manag.*, vol. 134, no. 6, pp. 504–510, Nov. 2008.