Internship at Barcelona Supercomputing Center: Generating Virtual Water Distribution Networks for flow control by a reinforcement learning agent

BELFADIL Anas

May 6, 2020

Part I

Internship data:

Barcelona supercomputing center is a prestigious research center, it hosts one of the most powerful supercomputers in the world, ranked at number 30. The center focuses on high performance computing in four main fields: Computer Sciences, Life Sciences, Earth Sciences and Computer Applications in Science and Engineering. My internship was kindly hosted by the Earth Sciences department.

Student's data	
NAME:	Belfadil Anas
UAB ACADEMIC TUTOR:	Anna Cima
Company's data	
COMPANY'S NAME :	Barcelona Supercomputing Center
TUTOR:	David Modesto Galende
Internship's data	
DATES:	March 1st - May 1st
OBJECTIVE:	Create virtual models of water distribution systems that will be used to train an artificial agent.

Part II

Internship project:

1 Motivation:

The scarcity of realistic networks is a known problem in water distribution networks research. It is due to safety reasons, and because of the time and costs associated with data collection 1. To overcame this difficulty, researchers proposed the generation of synthetic (virtual) WDSs. In 3 R. Sitzenfrei presented a review of different generators and their specific advantages.

Inspired by these generators especially 1, we created an algorithm for generating virtual WDNs that could be used to train an artificial agent to control the flow on the network.

2 Algorithm steps:

Our objective is to generate a very large number of random WDN's that are hydraulically 'realistic' i.e having mean pressure and velocity within usual intervals. This is obtained by penalizing pressures and velocities outside the usual intervals (see section 1.6. pipe sizing).
Here we present an overview of the algorithm. These steps will be discussed in more detail in the following sections.

Algorithm 1 Generate a random model of water distributing system

[1]

```
BEGIN
  Generate a random layout G(E,V) from street network data
  Add elevation values to the nodes of G
  Clean the graph from clustered nodes, self loops and parallel edges
  Add water demands to the nodes
  Create the main distribution network
  Add reservoirs as nodes and connecting them to the main distribution network
  Add pipe roughness to the edges
  Pipe sizing
  Add valves
  Discard networks with negative pressures
  Save network as Inp file and network statistics as CSV
END
```

2.1 Layout:

We use randomly selected areas from freely available street network data (OpenStreetMap) to generate the layout of the network. Mair et al. (2012)(2) have investigated network similarities between street, urban drainage and water supply networks. Their results showed that there are strong coherences between the three network-types. Moreover, when comparing the layouts and characteristics of the different WDN generation algorithms, R. Sitzenfrei (2016)(3) concluded that this aforementioned approach provides the most realistic layout due to utilizing the correlation with the street network data.
The layout is stored as a Networkx graph G(E,V).

2.2 Elevations:

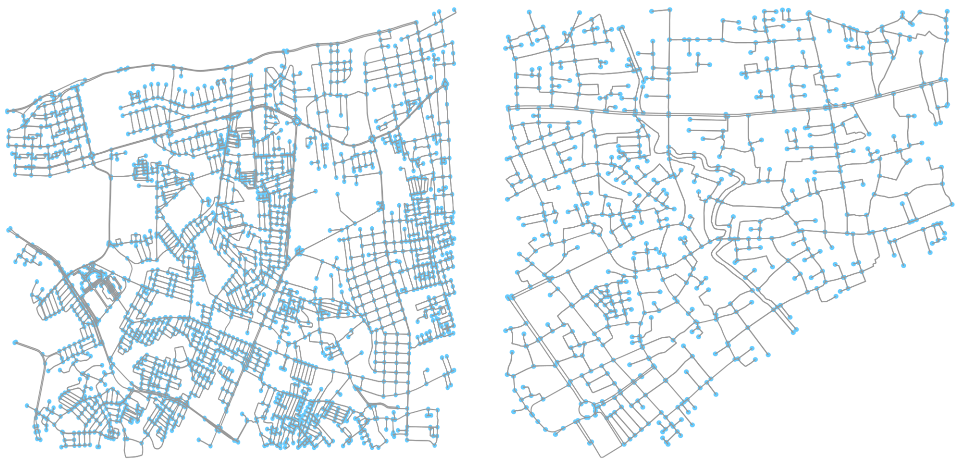
From the elevation map of the same region (Gmap API), we generate elevation values of the nodes of G.

2.3 Cleaning the graph:

To have a more clean layout we merge the clustered nodes that are within a tolerance value from each other (tol = 15m was taken). We also delete self loops if there are any and parallel edges, keeping just one edge between any two nodes u and v.

Figure 1: Layout examples generated from street network data

(a) Area = 36 km^2 (b) Area = 16 km^2



2.4 Demands:

The network is divided to communities of 4 nodes each, and each community is randomly assigned a value from the average demand values consumed by different land use types from the table:

Table 1: Average demand values consumed by different land use type in gal/day/acre from McGraw Hill (2000)(4)

Land use type	**Average demand in gal/day/acre**
Low_density_residential	1670
Med_density_residential	2610
High_density_residential	4160
Single_family_residential	2300
Multifamily residential	4160
Office_commercial	2030
Retail_commercial	2040
Light_industrial	1620
Heavy_industrial	2270
Parks	2020
Schools	1700

We consider that each node serves an area of 0.2 acre, the demands are then stored in m^3/s.

For the pipe sizing we use the peak hour demand, which is taken as the average daily demand multiplied by a peaking coefficient, the usual values for this coefficient are presented in table 2. For our algorithm we take a peaking coefficient value of four.

Table 2: Typical peaking coefficient from McGraw Hill (2000)4

Ratio of rates	**U.S. Range**	**Common Range**
Maximum day : average day	1.5 - 3.5 : 1	1.8 - 2.8 : 1
Peak hour : average day	2.0 - 7.0 : 1	2.5 - 4.0 : 1

2.5 Create the main network:

To create the main network we subdivide it to sectors, and connect these sectors between each other through 600mm pipes.

2.6 Roughness:

For simplicity a single value of roughness chosen randomly from roughness_values list is assigned to all the pipes.

2.7 Reservoirs:

The number of reservoirs to be added to the network is estimated form the number of pipes and the total area of the network. The network is then divided to 'NUMBER OF RESERVOIRS' random communities either by using FLUIDC ALGORITHM (5) or by using the ZONAGE() function that divides the network to communities each with at most n nodes. Then, for each community a reservoir node is created adjacent to the node in the community with the highest elevation, and is connected to it with a pipe of length 50m. Finally, the reservoir is assigned a random head value from the head_values list.

2.8 Pipe sizing:

First, we create a WNTR WATERNETWORKMODEL object from the Networkx graph we have constructed until now. Then, we start with the same diameter for all the pipes in the network 200 mm except the ones on the main distribution network which start with a diameter of 600 mm. Afterwards, the diameters are adjusted for each pipe to insure reasonable velocity and pressure values for most edges and nodes, we aim at velocity values $<=$ 2.5 m/s and pressure values between 20 and 100m.
These values are taken from Paez & Filion 7 who surveyed different recommendations for pipe sizing found on different references.
Table 3 presents values and constraints recommended from different expert sources. It includes values for the minimum diameter d_{min} the maximum flow velocity v_{max}, the maximum unit head loss Sf_{max}, the minimum and maximum allowable pressures p_{min} and p_{max}; and sometimes they are defined for different demand conditions where D_{max} is the maximum demand condition (commonly defined as the peak hour demand plus some fire scenario), and D_{ave} is the average demand condition (commonly defined as the mean daily demand).

Table 3: Recommendations and constraints for pipe sizing7

Reference	d_{min}	v_{max}	Sf_{max}**(1/100)**	p_{min}	p_{max}
Cesario (1995)	200 mm	$\bar{x} = 2$m/s	$\bar{x} = 6.2$	$\bar{x} = 23$m	$\bar{x} = 77$m
Survey		$x \in [1-6]$m/s	$x \in [1-15]$	$x \in [14-42]$m	$x \in [42-151]$m
Recommended		1.5m/s @$d < 600$ mm	1 − 2 @$d \geq 600$mm 2 − 5 @$d < 600$mm	28 - 35 m	63 - 77 m
Trifunovic (2006)		1m/s @large d 1.5m/s @small d	1-2 @large d 2-5 @midrange d 5-10 @small d	20 - 30 m	60 - 70 m @flat zones 100 - 120 m @hilly zones
Kujundzic (1996)				$\bar{x} = 28.5$m $x \in [20-40]$m	$\bar{x} = 93.6$m $x \in [70-160]$m
GLUMR (2012)	75 mm @no fire protection 150 mm @with fire protection			14 m @D_{max} 24.5 m @D_{ave}	70 m
Mississippi Dept. of Health (2001)	100 mm	1.5 m/s		14 m	56 m

Algorithm 2 Simple algorithm for pipe sizing

[1]

```
BEGIN
  Start with the 600mm diameter for the main pipes and 200mm for the rest
  While the mean pressure in the network is negative:  assign the next bigger
diameter to all pipes
  Repeat n times (n:  number of diameter types)
    For each pipe
      If the velocity < 0.5 and pressure > 70 make the diameter smaller
      If the velocity > 1.5 and pressure < 40 make the diameter bigger
      If the velocity < 0.01 make the diameter smaller
      If the velocity > 2.5 make the diameter bigger
END
```

2.9 Valves:

We add valves to the main distribution network at the intersections of three or more pipes and at the pipes leading to the reservoirs.

2.10 Saving as INP file:

Finally, we discard networks with negative pressures (are usually around 10% of the networks) and we save the network as an INP file that could be used by the reinforcement learning module for training the agent.

2.11 Conclusion of part1:

The algorithm presented can generate very big numbers of virtual water distribution systems, this is due to the randomization in the generation pipeline, starting with the random selection of the street location from which the layout will be generated and the total area of the network ($1km^2$ to $36km^2$) . Also, the random assignment of demands, the random selection of roughness, the semi-random placement of reservoirs and finally the placement of valves after sectorisation all add to the variability and randomness of the generated networks.

Figure 2: Examples of INP files created by the algorithm with pressure values and main distribution networks in red

Part III

Comparison with networks from literature using graph theory indexes

In this part we are going to present the results obtained from 100 random networks generated with the algorithm of part 1. First we compare the hydraulic performance of our networks to reference networks from the literature then we compare the two sets using graph theory indexes.

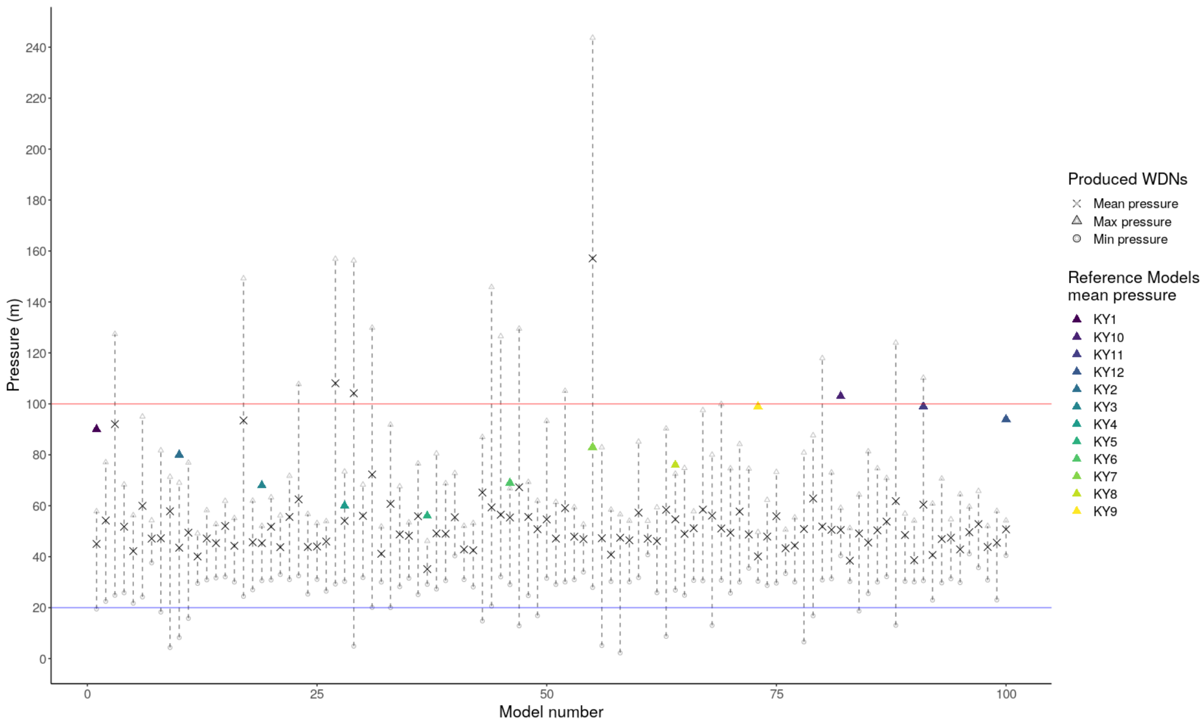


Figure 3: Pressure variations in 100 networks generated randomly, compared with 12 reference networks

3 About the reference networks:

The set of reference networks are from the Research Database of Water Distribution System Models (8), it is a set of 12 different networks developed from several small and medium actual systems in the state of Kentucky.

4 Hydraulic characteristics:

4.1 Pressure:

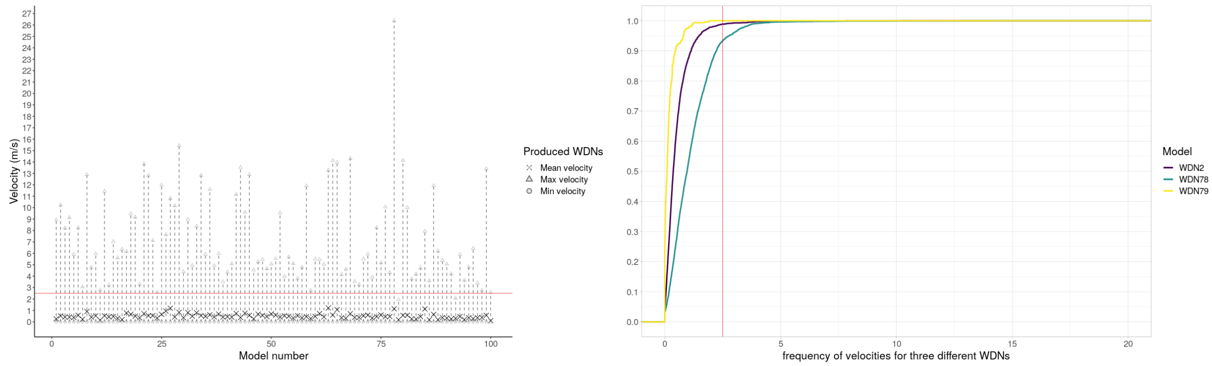
The majority of the generated networks have the pressure values within the interval [20-100]m, with just 3% of mean pressure beyond 100m, and 1% with mean pressure beyond 130m.

4.2 Velocity:

All the maximum velocities are well beyond the 2.5 m/s, although as we can see the mean velocities remain below 1 m/s, this means that those big velocities are exceptions in the network.

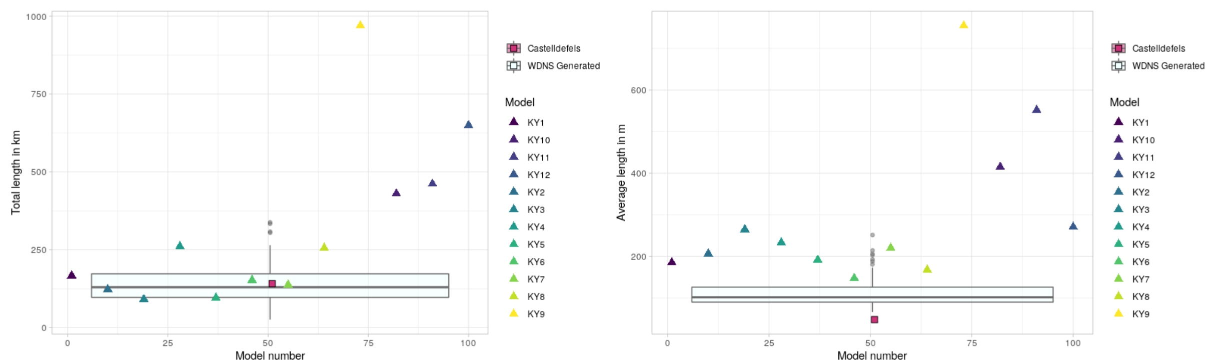
In fact when we draw the cumulative frequency plot for some of the networks, for example number 78 which seem to have a very big max velocity of 27m/s, and compare it with 79 that have a max velocity below 2.5m/s and also number 2 which is more representative of the general result. We can see that in all cases more than 90% of the velocities are less than 2.5m/s and 99% are less than 5m/s. The same kind of pattern is noticed when we run the simulations on the reference networks where 90% of values are within the 2.5m/s but the max velocity can also go to around 30m/s (network KY10 for example).

We can conclude that hydraulic characteristic of our networks are acceptable especially considering that they don't contain any regulatory elements such as pressure reducing valves, pressure sustaining



(a) Velocity variations in 100 networks generated randomly (b) Cumulative frequency of velocities in WDNs number 2, 78 and 79

Figure 4: Velocity analysis in the generated networks



(a) Total length of the network (b) Average length of the pipes

Figure 5: Length box-plot for 100 generated networks vs plots of the 12 reference networks vs Castelldefels network

valves or pressure relief valves, nor any pumps.

5 Layout characteristics:

5.1 Pipe lengths:

The average length of pipes in our networks is generally smaller than the reference networks, this is probably because most of our networks are generated from streets of high density areas(cities with more than 100k populations). This is moreover confirmed when plotting Castelldefels network characteristics where we find the average length to be even lower than our averages.

5.2 Graph theory indexes:

A WDN can be represented as a graph G = (N, E), containing a set N of n nodes (i.e., water demand and supply nodes) and a set E of m edges (i.e., water distribution pipes). Since the direction of the water flow in the pipes is subject to occasional changes (e.g., pressure changes, pump activation, change of status of shut-off valves, etc.), WDNs are represented as undirected graphs. Moreover, under

$$k_{avg} = \frac{1}{n}\sum_{i=1}^{n} k_i = \frac{2m}{n} \qquad k_{max} = max(k_i)$$

$$R_m = \frac{m-n+1}{2n-5} \qquad q = \frac{2m}{n(n-1)}$$

100 Generated WDNs	**Min.**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**
Nodes	191	587	909	958.9	1276.2	2912
Edges	271	851	1256	1411	1808.8	4826

Model	**KY1**	**KY2**	**KY3**	**KY4**	**KY5**	**KY6**	**KY7**	**KY8**	**KY9**	**KY10**	**KY11**	**KY12**
Nodes	777	513	264	939	409	763	500	1283	1118	886	731	2262
Edges	903	595	344	1118	498	1033	624	1523	1284	1036	836	2396

Table 4: Summary of the number of nodes and edges of 100 generated WDNs and of the 12 reference networks

normal circumstances, there exists at least one path between every pair of nodes and therefore, WDNs are connected graphs. Furthermore, a WDN can be represented as a graph where no edges cut one another (except at their nodes). This planarity could be violated in a WDN due to pipes crossing without intersecting, but this situation is uncommon in reality, therefore, planarity can be considered as a good approximation for WDN.

Given the graph G, the following connectivity indexes can be defined(7):

The degree k_i of a node i represents the number of edges incident to this node. The average degree of a graph k_{avg} is defined as the average degree of all the nodes of the graph. It is a measure of connectivity and gives information on the sparseness of the network: treelike network structures have an average degree of approximately 2, whereas an average degree of about 4 indicates a more complete network with grid pattern. As can be derived from the average degree of the observed real networks ($2 < k << 4$), WDNs tend to be rather sparse.

R_m is the meshedness coefficient that represents the fraction between the current number of loops and the maximum number of loops in a planar graph.

q is the link density that represents the fraction between the current number of links and the maximum number of links in a non-multigraph graph (i.e., a graph without edges that share the same adjacent nodes which in WDSs context could be understood as a network without parallel pipes).

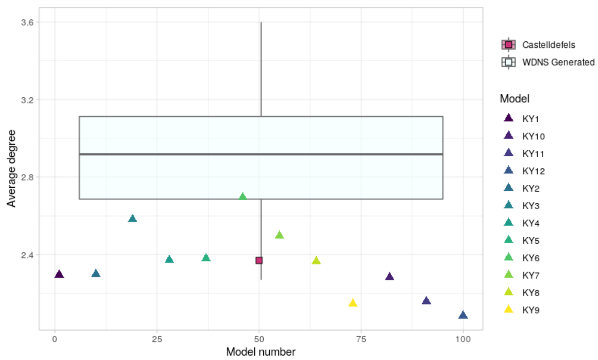
Using our algorithm we generated 100 WDNs covering surfaces from $4km^2$ to $36km^2$, and the following characteristics:

Here we present other graph indexes of our networks against the reference ones:

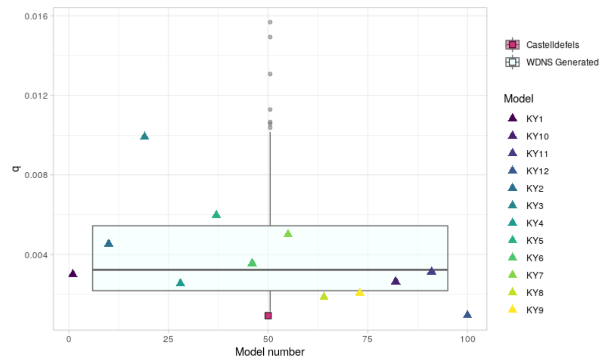
We notice that the link density q of the generated networks is within the interval of variation of the reference networks. However, the average degree and meshedness are generally bigger in the generated sets. When we investigate more we can see that generally our networks have less nodes with one degree, about the same number of nodes of two and three degrees, and more nodes with four and 5 degrees, with existing but negligible percentages of nodes with degrees of six and beyond.

Therefore, if it's necessary we can adjust the average degree by deleting some edges to bring the degrees of some nodes from four to one.

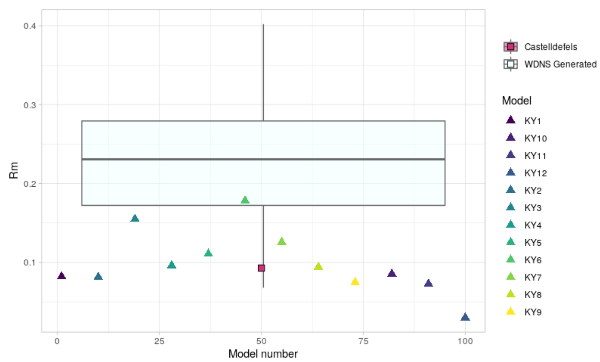
Also more real networks have to be investigated to get better data on the true intervals of variations of their characteristics.



(a) Average degree k_{avg}



(b) link density q



(c) Meshedness R_m

Figure 6: Box-plot of 100 generated WDNs vs plot of 12 reference models

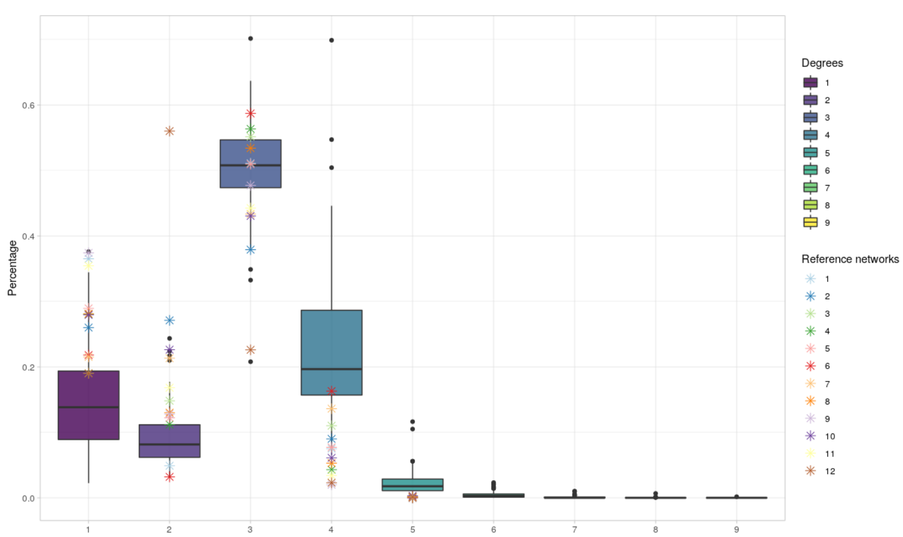


Figure 7: Degree frequency for 100 generated WDNs vs 12 reference networks

Part IV
Conclusions:

This internship was very interesting, it allowed me to put in practice a lot of the knowledge acquired in the master program. In fact, it relied heavily on programming with python for acquiring the data, modeling the water distribution network and running hydraulic simulations for sizing the network. The R programming language was also used for data visualization and to compare the performance of the created networks to reference networks.

The main difficulty faced in the time of this internship was the shutdown due to the Covid-19 pandemic. However, thanks to the efforts of the team at Barcelona Supercomputing center especially my tutor Mr. David Modesto, we were able to continue working from homes and to successfully attain the objectives of the internship.

We created and implemented on python an algorithm for generating virtually infinite numbers of WDNs. It was shown that these WDNs give good hydraulics performance considering they are lacking control elements such as pressure reducing valves, pressure sustaining valves, pressure relief valves, and pumps.

For taking this project further, we can suggest the following points:

Introducing a parameter for varying the average degree of the networks. In fact, figure 6 shows that the average degree of the produced networks is generally 20-25% bigger than the reference set of networks. It could be controlled, if necessary, by deleting some edges from nodes that have a degree of four or more.

More investigation could be carried to check the influence of including less dense areas (now cities with populations more than 100k were used) on the average length of the pipes.

The reservoir placement method could be improved, in fact one promising method that avoids negative pressures was implemented. Although as expected it has driven the mean pressure up, and an important percentage of the networks presented pressures outside the targeted 20-100m interval. Such a method should be coupled with pressure reducing valves to be give all the pressures in the desired range.

The process could ideally be be taken off-line by using local Open Street Maps fro producing the layout and Digital Elevation Models for elevations.

References:

1. M. Mair, W. Rauch and R. Sitzenfrei. Spanning tree based algorithm for generating water distribution network sets by using street network data sets. World Environmental and Water Resources Congress 2014

2. Mair M., Sitzenfrei R., Möderl M. and Rauch W. (2012). Identifying multi utility network similarities. In: World Environmental and Water Resources Congress 2012@ Crossing Boundaries, ASCE, Albuquerque, New Mexico, pp. 3147-53.

3. Sitzenfrei R. A Review on Network Generator Algorithms for Water Supply Modeling and Application Studies. In: World Environmental and Water Resources Congress 2016 May 22–26, 2016 | West Palm Beach, Florida

4. Water Distribution Systems Handbook - McGraw Hill (2000)

5. Parés, Ferran & Garcia-Gasulla, Dario & Vilalta, Armand & Moreno, Jonatan & Ayguad, Eduard & Labarta, Jesús & Cortés, Ulises & Suzumura, Toyotaro. (2017). Fluid Communities: A Competitive, Scalable and Diverse Community Detection Algorithm.

6. Fast unfolding of communities in large networks (2008) Vincent D. Blondel , Jean-Loup Guillaume , Renaud Lambiotte and Etienne Lefebvre

7. Paez, D., & Filion, Y. (2016). Use of network theory and reliability indexes for the validation of synthetic water distribution systems case studies. Sustainable Energy Technologies and Assessment

8. Research Database of Water Distribution System Models Matthew D. Jolly; Amanda D. Lothes; L. Sebastian Bryson, P.E., M.ASCE; and Lindell Ormsbee, P.E., D.WRE, F.ASCE