Stanisław Biedugnis (stanislaw.biedugnis@pw.edu.pl)
    The Main Fire Service School, Warsaw

Andrzej Czapczuk (andrzej.czapczuk@fbitasbud.pl)
    F.B.I. TASBUD S.A., Science and Research Division, Warsaw

The application of the 'K-nearest neighbour' method to evaluate pressure loss in water supply lines

Zastosowanie metody K-najbliższych sąsiadów do oceny strat ciśnienia w przewodach wodociągowych

Abstract

Water supply systems are complex engineering structures; certainly, the most important part is the water distribution network. The design of this element requires calculations and many analyses to arrive at the best solution. The main task of the calculation is to determine the flow rates through pipes, to determine pressure losses, height of tanks, pressure required in the supply pumping station, pressure levels in the individual nodes of the network. Correct execution of the calculations requires careful evaluation of the results obtained and accuracy in the solutions applied. The issue of controlling the results of calculations is difficult to present in algorithmic form as these are mainly based on the experience and knowledge of the designer. Classes of decisions describing the problems of pressure loss in the pipework were established in order to evaluate the results of calculations. Numerical experiments were carried out in this paper to show how the 'K-nearest neighbour' method can be used to evaluate pressure loss in water pipes.

Keywords: water distribution system, hydraulic calculations, evaluate of pressure loss, K-nearest neighbour method

Streszczenie

Systemy zaopatrzenia w wodę są skomplikowanymi układami obiektów inżynierskich, które pełnią różnorodną rolę w funkcjonowaniu całości. Najważniejszym elementem jest z pewnością system dystrybucji wody. W związku z powyższym projektowanie systemów dystrybucji wymaga przeprowadzenia obliczeń oraz wielu analiz mających na celu doprowadzenie do najlepszego rozwiązania. Podstawowym celem obliczeń jest wyznaczenie natężeń przepływów przez przewody, określenie strat ciśnienia, wysokości zbiorników, wymaganego ciśnienia na zasilaniu oraz ciśnienia w poszczególnych węzłach sieci. Poprawne zrealizowanie obliczeń wymaga wnikliwej oceny uzyskanych wyników oraz poprawności zastosowanych rozwiązań. Zagadnienia kontroli rezultatów obliczeń trudno jest ująć w postaci algorytmicznej, gdyż oparte są głównie na doświadczeniu i wiedzy projektanta. Do oceny wyników obliczeń zdefiniowano klasy decyzyjne opisujące problemy związane ze stratami ciśnienia w przewodach. W pracy przeprowadzono eksperymenty numeryczne pokazujące, w jaki sposób może być zastosowana metoda K-najbliższych sąsiadów do oceny strat ciśnienia w przewodach wodociągowych.

Słowa kluczowe: system dystrybucji wody, obliczenia hydrauliczne, ocena strat ciśnienia, metoda K-najbliższych sąsiadów

1. Introduction

Water supply systems are complex engineering structures; certainly, the most important part is the water distribution network. This is dictated by high construction costs as well as the complex processes of operation and renovation. The design of this element requires calculations and many analyses to arrive at the best solution. The main task of the calculation is to determine the flow rates through pipes, to determine pressure losses, height of tanks, pressure required in the supply pumping station, pressure levels in the individual nodes of the network.

For many years, computer technology has been used for the calculation of hydraulic water distribution systems [1, 2]. Such computer programs appeared in the second half of the 20th. century [3, 4]; however, they were relatively difficult to handle and required a lot of input work – they are now equipped with many functions [5, 6]. Despite significant advances in computer technology for calculating water distribution systems, the procedures that are used still have a predetermined course. Implementing such calculations requires the designer to take a number of decisions that affect both technical and economic parameters; however, other factors, such as reliability [7–10] and safety in relation to water supply systems, are often taken into account [11]. This requires careful assessment of the results obtained and accuracy in the solutions applied. Methods involving artificial intelligence or non-parametric regression can be a great support in the implementation of this kind of task as it can be difficult to handle with regard to devising the required algorithms.

In the literature, there are numerous examples of artificial intelligence methods being used to evaluate the results of analyses of water distribution systems [12–16]. This paper discusses the possibility of using the 'K-nearest neighbour' method to evaluate pressure loss in water supply lines.

2. Methodology for estimating pressure loss in water supply lines

The evaluation of pressure losses can be difficult because identical values of Δh_l, depending on conditions, may be considered to be high, within the optimum range, or relatively low. For example, Fig. 1 shows a case in which equal values of Δh_l for segments of different lengths, can be classified in one of the three categories above.

The basic decisive factor for pressure loss is the diameter of the pipeline, which is selected on the basis of speed. However, you can specify other parameters that affect pressure loss and are not related to velocity. The first of these is the length L of the pipe section, while the second is the coefficient of absolute roughness k.

When evaluating pressure losses, it is firstly necessary to determine whether Δh_l is correct for a given condition; if it is too low or too high, the reason for that situation should be determined. A review of results for individual sections, especially in the case of large water distribution systems, can be tedious, time consuming, and confusing. The causes of pressure losses that would require correction, or the introduction of specific operating instructions, such as periodic pipe rinsing, are quite varied. For the purposes of this paper, four classes
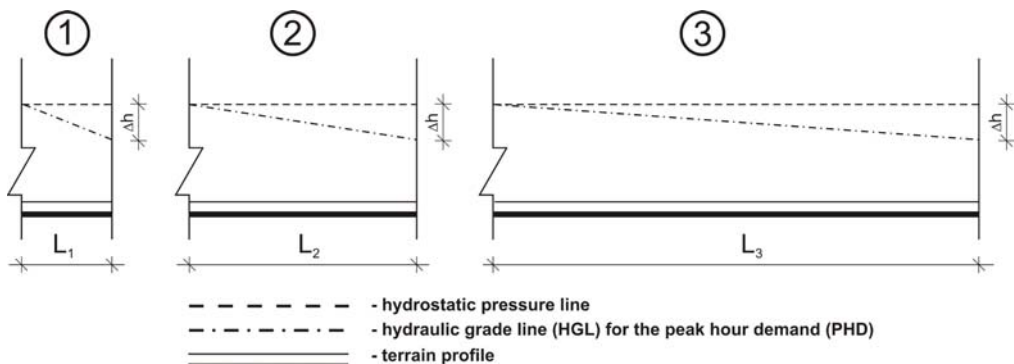
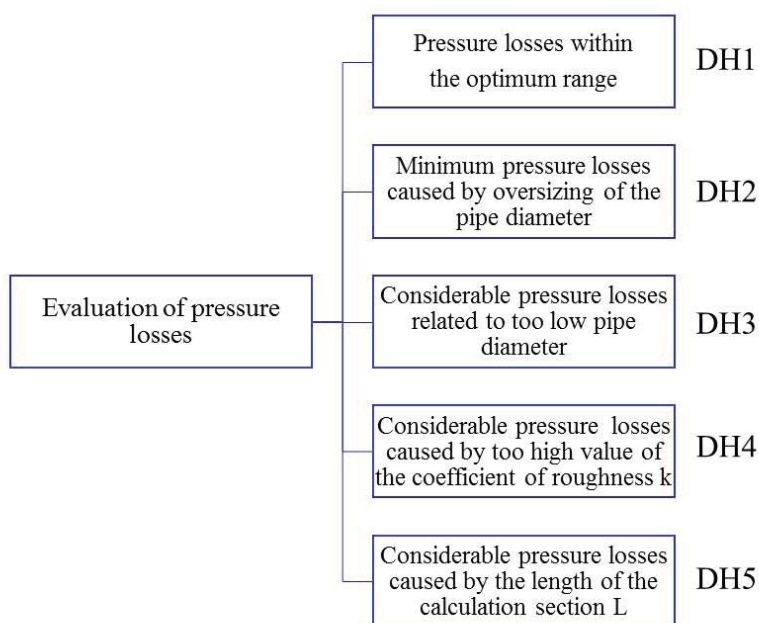Fig. 1. Evaluation of pressure losses on sections of various lengths L

Fig. 2. Classes for the assessment of pressure losses in pipelines of water distribution systems

have been identified; these describe the causes of incorrect assumptions *vis-à-vis* Δh_1 pressure losses. One further class corresponds to the range of appropriate values. Classes described with labels DH1–DH5 are shown in Fig. 2 [12].

Defining the DH1 class corresponding to the range of correct values of pressure losses on a computational section, takes into account not only flow velocity and diameter but also the length of the section and the coefficient of absolute irregularity k.

No proposal was found in the available literature, which would clearly define the optimal range of pressure losses Δh_1 in the section.

143

There are no proposals in the literature that would specifically define the optimal pressure loss range Δh_l in the section. However, proposals and instructions for maximally permissible, unit pipeline pressure drops are available; these are determined in relation to the diameter [16] or flow rate along the section [18]. Depending on the diameter, the following unit values for maximum pressure losses, conditioned by the need to maintain their respective flow rates, [16] are given:

- 15–20 ‰ for the pipework with diameters DN100–DN250 mm
- 15 ‰ for busbar with diameters up to DN500 mm
- 5 ‰ for pipework with diameters greater than DN500 mm

The limit values, as a function of the flow rate for a flat area, are given in Table 1.

Table 1. Permissible unit pressure loss for flat land, depending on the flow rate [18]

Stream (dm³/s)	Pressure-drop in the pipelines (‰)
0.1–72	10
72–480	5
480–4000	3.5

It seems, however, that the permissible unit pressure loss as a function of the diameter of the pipework or flow rate is insufficient when assessing pressure losses. The computer system should also detect situations in which the pressure losses are low or definitely too high. This paper assumes that the range of permissible pressure losses described in the DH1 class occurs when the following conditions are met:

- the flow rate is higher than 0.5 m/s;
- the flow velocity for individual diameters does not exceed the recommended values for the individual diameters;
- the roughness coefficient does not exceed the upper limit of resistance, assumed on the basis that $k = 1.5$ mm;
- pressure losses resulting from the length of the calculation of section L will not cause the pressure line to fall below the value required for a 4-storey building, i.e. the pressure line will not fall below 25.5 m, assuming that the losses resulting from the absolute roughness coefficient k are normal. The starting pressure is assumed to be 40 m; therefore, the pressure loss cannot exceed 14.5 m. This condition is required in order to ensure the required pressure, while keeping the pressure as low as possible.

The DH2 class describes the situation where low pressure losses are caused by oversized pipework or by restricted flow at the ends of the water supply network, where the flow rate is less than 0.5 m/s.

The DH3 class is characterised by a pressure loss which is due to too small a diameter in the pipeline. The flow rate exceeds the recommended values for individual diameters. The value of the roughness coefficient is below the upper limit of $k = 1.5$ mm.

The DH4 class describes conditions in which pressure losses are dictated by a high value of the absolute coefficient *k*, above the *k* limit of 1.5 mm.

The DH5 class refers to a situation in which the primary cause of pressure loss in the pipeline – below the required value –is the pressure loss associated with the calculated length of a particular section of the line. In the case where calculations of long water transits or mains are taken into account as a single, computational section, this variant signals the need for zoning of the system. This length is different each time, depending on the pipe diameter, flow rate and roughness.

The above classes will be assigned to individual computational sections of the water distribution system, using the '*K*-nearest neighbour' method.

Using the *K*-nearest neighbour approach requires a set of teaching examples containing arguments or variables and classes corresponding to them. Teaching examples, in the form of calculation results for sections of the water distribution system, were prepared using EPANET software and EXCEL spreadsheets. Pressure losses over the length of the section are calculated according to the Darcy-Weisbach formula.

PE100 polyethylene pipelines of the SDR17 series (PN-EN 12201-2:2011) were used for the calculations.

Hydraulic calculations were made using the following assumptions:

▸ internal diameters of pipelines D_w were used in the calculations;
▸ the minimum cable diameter was DN110;
▸ the maximum cable diameter was DN630;
▸ the following roughness coefficients were assumed: $k = 0.01; 0.1; 0.5; 1.0; 1.5; 2.0$ mm;
▸ the maximum calculated length of the sections was 3000 m.

The teaching examples relate to individual sections of the water supply network and were developed in such a way that they could be used to assess pressure losses in classes DH1–DH5. In order to induce the decision tree, the problem domain is defined with the following attributes:

▸ the length of the computational line *L*
▸ calculation flow on the Q_m section
▸ the absolute roughness coefficient of the pipework on a given section *k*
▸ linear pressure losses on the computational section Δh_1

All teaching examples were described by the DH1–DH5 label indicating that they belonged to a specific class characterising pressure losses. Tutorials are saved as an Excel spreadsheet file, where each row corresponds to learning examples and the columns correspond to attributes. A collection of 17019 teaching examples was formed – these represented all the DH1-DH5 classes described above.

3. The *K*-nearest neighbour method

A non-parametric regression algorithm was used for forecasting the value of random variable values and for classification. The task was to classify a new object into a positive or

a negative group based on what it is adjacent to. An important parameter in this method is the value of *K*, which is the number of teaching examples in the nearest neighbourhood. A new value for introduction into the space was sought on the basis of the value of those *K* neighbours closest. In the case of classification, the method takes into account the value of most of the examples in the neighbourhood, i.e. it sets the value of the new example by voting.

Choosing a neighbourhood, in other words choosing the value of the *K* parameter, is essential. With this method, this parameter is decisive with regards to the accuracy of the prediction. At a low *K* value there is a high variability of prediction, while at a high *K* value, there may be a significant, systematic shift of prediction values. On the basis of the analyses, the *K* value must be chosen to minimise the likelihood of misclassification. The *K*-nearest neighbour method proposes an optimal *K* value based on the cross-check method [20].

The Euclidean Mara is most often used to evaluate the distance between the points describing the training examples:

$$D(x,p) == \sqrt{(x-p)^2} \tag{1}$$

where:

x – is a new case to be classified,

p – is one of the training cases.

Application of the *K*-nearest neighbour method to evaluate pressure losses in water-supply lines

Calculations for water pipes were made using the **EPANET** programme. Pressure losses were then calculated along individually calculated sections using a spreadsheet; a suitable DH class was then assigned to each section. Calculations were made using different coefficients of roughness k and the lengths of the sectors calculated. This step was dictated by the fact that the variability of attributes across the whole possible range needed to be taken into account in the examples. Some of the hydraulic calculations were deliberately misplaced; however, these examples were described by the appropriate DH classes so that the expert system could identify the cause of the pressure losses. 12,754 teaching examples were obtained.

Based on the set of teaching examples, the *K*-nearest neighbour method was graded using different neighbourhood values.

Detailed classification results are included in the matrix of confusion [20]. This is a square matrix, in which information as to which classes individual examples actually belong, are in rows and information as to how they were classified by the classifier are in columns. The diagonal contains examples that are categorised as correct while those located beyond the diagonal are classified as incorrect. At the same time, examples beyond the diagonal indicate the classes to which they were classified, albeit incorrectly.

Tables 1 and 3 show the teaching outcomes of the *K*-nearest neighbour for neighbourhood *K* = 1 and *K*-5. The Euclidean measure was used to calculate the distance between teaching examples. Tables 2 and 4 summarise the classification of examples for *K* = 1 and *K*-5.

Table 2. Matrix of confusion for the nearest neighbourhood method for $K = 1$

	DH3	DH5	DH2	DH1	DH4
DH3	1296	141	43	0	7
DH5	96	855	0	20	0
DH2	42	0	971	0	7
DH1	0	31	0	556	0
DH1	6	2	13	5	164

Table 3. Summary of classification results for $K = 1$

	Total	Accurate	Invalid	Relevant (%)	Invalid (%)
DH3	1440	1296	144	90.00000	10.00000
DH5	1029	855	174	83.09038	16.90962
DH2	1027	971	56	94.54722	5.45278
DH1	581	556	25	95.69707	4.30293
DH1	178	164	14	92.13483	7.86517

Table 4. Matrix of confusion for the nearest neighbourhood method for $K = 1$

	DH3	DH5	DH2	DH1	DH4
DH3	1300	118	60	0	9
DH5	116	831	0	24	0
DH2	54	0	959	0	7
DH1	0	13	0	574	0
DH1	7	2	14	7	160

Table 5. Summary of classification results for $K = 5$

	Total	Accurate	Invalid	Relevant (%)	Invalid (%)
DH3	1477	1300	177	88.01625	11.98375
DH5	964	831	133	86.20332	13.79668
DH2	1033	959	74	92.83640	7.16360
DH1	605	574	31	94.87603	5.12397
DH1	176	160	16	90.90909	9.09091

4. Summary and Conclusions

The results obtained by the K-nearest neighbour method for neighbourhoods $K = 1$ and $K = 5$ are comparable. They differ in the accuracy of the classification for individual classes. In general, however, classification errors for both K values are quite high. The results obtained, along with previous experiments with other artificial intelligence methods, especially with the decision tree induction method, show that the K-neighbour approach is a much worse solution than induction or artificial neural networks.

References

[1] Biedugnis S., *Metody informatyczne w wodociągach i kanalizacji*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 1998.

[2] Rossman L.A., *EPANET 2 User's manual, EPA/600/R-00/057*, National Risk Management Research Laboratory, U.S. Environmental Protection Agency, Cincinnati, OH, USA 2000.

[3] Epp R., Fowler A.G., *Efficient Code for steady state Flows in Networks, Journal of the Hydraulics Division*, ASCE, Vol. 96, No. HY1, 1970, 43–56.

[4] Adams R.W., *Distribution Analysis by Electronic Computer*, Institute of Water Engineers, Vol. 15, 1961, 415–428.

[5] Knapik K., *Dynamiczne modele w badaniach sieci wodociągowych*, Wydawnictwo Politechniki Krakowskiej, Kraków 2000.

[6] *Pipe2010 Water Utility Modeling: A Comprehensive Guide to Hydraulic and Water Quality Modeling of Drinking Water Systems Using Pipe 2010*, KyPipe, http://kypipe.com.

[7] Królikowski A.J., *Niezawodność działania systemów zaopatrzenia w wodę*, Ekologia i Technika, Vol. 1, No. 1, 1993, 7–14.

[8] Wagner J.M., Shamir U., Marks D.H., *Water distribution reliability: simulation methods*, Journal of water resources planning and management, Vol. 114, Issue 3, 1988, 276–294.

[9] Wieczysty A., *Niezawodność miejskich systemów zaopatrzenia w wodę: praca zbiorowa*, Politechnika Krakowska, Kraków 1993.

[10] Kwietniewski M., Roman M., Kloss-Trębaczkiewicz H., *Niezawodność wodociągów i kanalizacji*, Arkady, Warszawa1993.

[11] Zimoch I., *Bezpieczeństwo działania systemu zaopatrzenia w wodę w warunkach zmian jakości wody w sieci wodociągowej*. Ochrona Środowiska, Vol. 31, Issue 3, 2009, 51–55.

[12] Czapczuk A., *System ekspertowy do oceny przepływów i strat ciśnienia w układzie dystrybucji wody*, Dysertacja, Wydział Inżynierii Środowiska, Politechnika Warszawska, Warszawa 2013.

[13] Czapczuk A., Dawidowicz J., Piekarski J., *Metody sztucznej inteligencji w projektowaniu i eksploatacji systemów zaopatrzenia w wodę*, Annual Set – The Environment Protection, Vol. 17, No. 2, 2015, 1527–1544.

[14] Dawidowicz J., *Evaluation of a pressure head and pressure zones in water distribution systems by artificial neural networks*, Neural Computing & Application, 2017, doi:10.1007/s00521-017-2844-8.

[15] Dawidowicz J., *Diagnostyka procesu obliczeń systemu dystrybucji wody z zastosowaniem modelowania neuronowego*, Rozprawy Naukowe, No. 268, Oficyna Wydawnicza Politechniki Białostockiej, Białystok 2015.

[16] Dawidowicz J., *System ekspertowy do oceny układu systemu dystrybucji wody sporządzony za pomocą wnioskowania indukcyjnego*, Annual Set the Environment Protection, Vol. 14, 2012, 650–659.

[17] Gupta R.K., *Analysis and Control of Flows in Pressurized Hydraulic Networks*, PhD, UNESCO-IHE Institute, Delft 2006.

[18] Łyp B., *Wybrane problemy wodociągów i kanalizacji w przestrzennym planowaniu zagospodarowania miast*, COIB, Warszawa 1992.

[19] Bishop C.M., *Neural Networks for Pattern Recognition*, University Press, Oxford 1996.

[20] Triantaphyllou E., Felici G. (Eds.), *Data mining and knowledge discovery approaches based on rule induction techniques*, Vol. 6, Springer Science & Business Media, 2006.