Wroclaw, 15.03.2019 r.

Annex 3

The author’s review of her own research

Małgorzata Kutyłowska, D.Eng.
Wroclaw University of Science and Technology
Faculty of Environmental Engineering
Department of Water Supply and Sewage Systems

I. Diplomas and academic degrees held – with specified name, place and year of receiving, and title of PhD thesis

In 2002 I graduated from the Polish-Community-in-Belgium 14th Grammar School in Wroclaw and began studying at the Faculty of Environmental Engineering, Wroclaw University of Science and Technology (WUS&T). I graduated (magna cum laude) from WUST in 2007, receiving the Master of Engineering Degree Diploma in Water Supply and Sewage and Waste Disposal (specialization: Water Supply and Sewage Systems). In 2007 I began my doctoral studies under the supervision of prof. Małgorzata Kabsch-Korbutowicz. In 2010 I defended my PhD thesis entitled: *Forecasting the parameters of membrane separation processes using artificial neural networks*.

II. Employment in research units

In 2009, still as a PhD student, I took up a part-time job as research-teaching assistant lecturer at the WUS&T Faculty of Environmental Engineering, which up to this day has continuously been my principal workplace. Immediately after I had defended my PhD thesis, in 2010 I was employed full-time as assistant lecturer at the WUS&T Faculty of Environmental Engineering. In 2011 I was promoted to the position of research-teaching adiunkt (lecturer).

III. The author's scientific achievement according to art. 16 sect. 2 of the Academic Degrees and Titles and Arts Degrees and Titles Act of 14 March 2003

Title: *Regression and classification methods in the analysis and assessment of the failure level of water conduits*.
Author: Małgorzata Kutyłowska
Publishing House: Wroclaw University of Science and Technology Publishing House, Wroclaw 2019.
Publishing reviewers: prof. Andrzej Kotowski, hab. D.Eng., prof. Ziemowit Suligowski, hab. D.Eng.

Description of the thesis:

The current water supply systems are a very important, in fact essential, part of the municipal infrastructure, without which developed societies would not be able to function normally. The systems consist of many components performing different functions. These are: water sources and intakes, water pumping stations and treatment plants, water supply reservoirs and conduits – water transmission pipelines, mains, distribution pipes and house connections. At the current stage of knowledge development, when well proven methods of designing water supply systems are available and when hydraulics and, in general, fluid mechanics are commonly known and applied, it seems that as regards water distribution systems, the emphasis should be put on their improvement and proper operation as well as on research into their operational reliability to extend their service life.

The aim of this research work was to develop regression and classification models, using selected prediction (machine learning) methods. This monograph represents an innovative and vitally important approach to the assessment of the level of reliability of municipal water supply systems in Poland and it broadens, completes and sums up the author's previous research based on data acquired from different water supply systems. The range of the research presented in this work is considerably wider in comparison with the author's earlier analyses, and the modelling procedure has been modified to better demonstrate the adaptability and applicability of the three selected algorithms for quantitative and qualitative analysis and forecasts.

The regression problem consisted in predicting a dependent quantitative variable, i.e. a stationary failure rate of water conduits (λ). Whereas as part of the classification problem a dependent qualitative variable, i.e. a classification of types of damage to water conduits, was modelled. The quantitative and qualitative investigations and the modelling were based on the operational data for two water distribution systems, obtained from water companies (in city X and Y). The following three algorithms: the support vector machines (SVM) method, the K-nearest neighbour (KNN) method and the regression (RT) and classification trees (CT) method were selected for analysis and modelling. Each of the methods has different basic parameters taken into account when building models. This resulted in a great variety of model structures and in large differences between the modelling results. A major initial task was to create a few configurations of independent variables to check if a considerable expansion of the predictor vector to include more data (e.g. pipeline burial depth and pressure, and season) than in the author's previous investigations would affect modelling quality, the convergence of the predicted dependent variable with the actual values of the failure rate, and the accuracy of classifying types of failures. From the results of modelling based on testing sample data (representative for qualitative and quantitative assessments) one can draw several major conclusions summing up the investigations.

The main goal of this research was achieved since SVM, KNN, RT and CT were shown to be more or less suitable for the analysis and assessment of the failure rate of the investigated water conduits. Moreover, the commercial software Statistica was found to be quite sufficient for predicting the dependent variables on the basis of the available operational data. This fact is significant considering the possible applications of the machine learning methods.

In order to rationally plan repairs and upgrades of selected, most failure-prone sections of the water supply network and to acquire deeper knowledge about the level of reliability of the water conduits, water companies can use the proposed methodology to model selected dependent variables. However, this is practically possible only when one has a relatively large and representative set of real data since it was shown (for city Y) that information for only three years of operation can be insufficient to draw rational conclusions. Hence the indirect conclusion emerging from this research is that it is necessary to systematize the data collected in water companies.

The modelling results were not always satisfactory because of the large divergence from the actual values of failure rate λ or the poor accuracy of the classification of water conduit damage types. However, at each stage of the investigations the causes of this were identified as stemming from, e.g., the model structure or the peculiar character of the data. The main conclusions emerging from the analysis of the modelling results, and the advantages and disadvantages of the adopted research methodology are presented below.

At the present stage in the development of research on machine learning methods in reliability analysis one cannot arbitrarily exclude any prediction algorithm since as the above research has shown, the results can differ considerably depending on the accuracy and quality of the operational data, the adopted way of dividing the data and the creation of separate models for different types of water conduits. The results have shown the KNN algorithm to be the least effective prediction method. For mathematical modelling aimed at determining the level of failure and operational reliability of water distribution systems it is recommended to use at least two of the methods indicated in this monograph, whereby it will be possible to make a comparative analysis. If only one methodology were used, unreliable results could be obtained and the conclusions drawn on this basis would be erroneous (which would translate into, for example, the water company's irrational expenditures).

The number of support vectors and the number of localized vectors are major model parameters in the support vector machines method. An analysis of the regression results for both cities indicates that the number of vectors is several times larger for the larger set of training data and the more complex predictor vector (city X). But in this case, this is not a drawback since it resulted in a smaller (by an order of magnitude) cross-validation error and did not increase the capacity, in comparison with the models describing the failure rate (in city Y). This is a significant finding indicating that the inclusion of more data in the analysis and the larger range of the data do not translate directly into a greater complexity of the model. A major common denominator of the created regression models is that mainly models based on a linear or polynomial kernel function were selected to describe failure rate λ (in city X and Y), which means that the prediction of the considered dependent variable is relatively simple. Only in a few cases models with radial basis functions were selected, which did not result in a greater convergence with the real values at all.

In particular, the agreement between the predicted failure rates and the experimental data for the water mains in city X was middling. The correlation coefficients were in a range of 0.32–0.88 and the relative model errors ranged from 11.8% to 12.2%. The models without the type-of-damage predictor showed a greater convergence with the actual values of rate λ.

The SVM method's advantage, i.e. its resistance to outliers, noted in the literature on the subject, was not observed in the case of the house connections in city X in 2009. This could be due to the peculiarity of the real data, which for a few years had assumed constant values. But also for other types of conduits the outliers were found to be incorrectly modelled by the SVM models. Less detailed analyses of the research results were presented for city Y, since only three years of their operation were taken into account. Even though the relative errors of the models did not exceed 30% and the correlation was in the range of 0.76–0.99, the results of modelling using the SVM method cannot be regarded as entirely correct because of the poorly representative character of the operational data. The problem of the generation of constant values also occurs in the modelling of failure rate λ for house connections in city Y.

Summing up, one can say that the SVM method can be recommended for the analysis of the level of failure and operational reliability of water conduits, but one should bear in mind the limitations connected with outliers or constant values. Moreover, one can narrow the range of further investigations and limit oneself to models based on the linear or polynomial kernel function since extending the analyses to include, e.g., the sigmoidal kernel function not only does not result in improved quality, but also complicates the analyses. Especially that in the classification problem the models with the linear or polynomial kernel function were characterized by the highest classification accuracy – in a range of 44–72%. The relatively high classification accuracy was the result of mainly the correct classification of fracture or corrosion type damage, but it did not translate into satisfactory results for other, less predominant types of damage. The advantage of the classification of qualitative variables describing the failure frequency of water conduits is the relatively small number of model parameters to be predefined and selected.

Because of the high complexity of the regression problem in city X, in the case of the K-nearest neighbours (KNN) the models had a few times larger number of K-nearest neighbours in comparison with the regression problem in city Y. The optimal models were described mainly by the Chebyshev distance measure. Moreover, the modelling results were found to be identical as for the Euclidean distance measure and its square, as well as for the Manhattan distance measure, in the case of city Y. Therefore it seems sensible to limit the range of investigations to the use of only the Chebyshev measure and the Euclidean measure in order not to complicate the process of building and developing KNN models by taking into account other distance measures.

Similar observations were made when analysing the classification problem in which KNN-CZ and KNN-E predominated. An analysis of the results of modelling failure rate λ for the water mains in city X indicates that the KNN method is less useful than the SVM method because of greater discrepancies between the actual values and the predicted ones, even though the relative errors of about 22% and the correlation of about 0.42–0.90 were acceptable. In many cases the KNN models overestimated the failure rate and for some configurations (configurations B and D in city Y) practically constant values were obtained for the house connections and the distribution pipes, which indicates that the division of data according to conduit type and the creation of separate models for each conduit type does not result in better quality of modelling results. For the other configurations the modelling of failure rate λ for the house connections was correct, with correlation at the level of 0.91–0.98, model errors amounting to about 22% and the failure frequency correctly estimated for the year 2009 (unlike in the case of the SVM method). The results of predicting the dependent variable in city Y show the trend in the change of this variable at a relatively high correlation, but because of the lack of resistance of the models to outliers (e.g. λ = 0.66 fail./(km·a) for house connections), it is not recommended to use the KNN method for regression problems with a too small predictor vector and a small size of the group of testing data. Neither the classification task results are of particularly high quality, considering that the classification accuracy was in the range of 37–55% and only the most common types of damage were correctly classified. Judging by the results of the present research and the previous findings, the KNN method should not be widely used to assess the technical condition and the level of operational reliability of water distribution systems.

As opposed to the two algorithms discussed above, the regression and classification trees method has an important advantage: independent variables are ranked when creating a tree model, which is not possible in the case of the SVM method and the KNN method. An analysis of this ranking showed that the extension of the predictor vector has no significant effect on modelling quality and prediction results. As regards the models developed to predict the failure rate of water conduits in city X, it has been shown that the most important independent variables are conduit material, type and diameter. Similar observations were made for city Y, where the models were based only on the major predictors describing the pipelines. Thus one can say, with a large degree of certainty, that as regards the regression problem there is no need to extend the independent variable vector to include, e.g., conduit burial depth or pressure, since this does not translate into prediction quality, and the data are sometimes uncertain and impossible to acquire. Especially that also in the case of the classification problem it was found that three predictors (sometimes slightly different than the ones used in the regression problem) were enough. Not much higher classification accuracy was achieved when seven predictors were used.

The tree model's most important parameters is the number of end nodes and divided nodes and the type of the independent variable taking part in the division at each model building stage and in each leaf. The number of cases and the size of the learning vector has a bearing on the complexity of the tree. Therefore the optimal regression tree models were greatly extended (the number of end nodes was twice larger) in the case of city X in comparison with city Y. It is surprising that the optimal classification trees are much smaller than the regression trees. This could have influenced the quality of classification, which is comparable with that of the other methods, amounting to 37–72%. In the selected configurations in the regression problem for city X the results of modelling the failure rate of the water mains are unacceptable. It was only when models were built according to conduit type, a convergence of 0.84–0.89 was achieved, but at large model errors, reaching as much as 70%. The situation repeats itself since the other methods, also in many cases, did not perform well in evaluating the failure rate λ of the water mains, which means that it is necessary to improve the research technique for the prediction of the failure rate of water mains. A much better convergence was obtained for the distribution pipelines and the house connections in city X. No significant differences in prediction results between the different configurations were found. The correlation amounted to 0.86–0.97 and 0.84-0.96 and the model error to 9.7–11.8% and 11.4–13.5% for respectively the house connections and the distribution pipes. Equally satisfying results for the distribution pipes and the house connections were obtained in the case of city Y, for which the regression tree models were less complicated. The model error did not exceed respectively 27% and 20% and the correlation coefficients were as high as 1.0 and 0.99. But, as already mentioned, data for merely three years of operation can detract from the quality of modelling and statistically affect the results. Moreover, the problem of incorrect prediction in the case of constant or outlying real values of the failure rate was found to occur, especially in the analysis of the results for city X. Nevertheless, in comparison with the KNN method, the regression and classification tree models perform better in assessing the level of failure of water conduits. Also the simple structure of tree models is their advantage and so the RT and CT method can be recommended for analyses and reliability assessments of underground infrastructure.

The investigations have shown that it is necessary to systematize and group the operational data registered in water companies, whereby the possibilities of using the collected information for the purposes of developing reliability models will greatly increase.

IV. Review of other scientific and research achievements

I have published the total of 44 works: 29 articles (including 14 with IF), 11 chapters in monographs and 4 conference papers indexed in Web of Science. Moreover, I have co-edited 3 monographs and 2 conference proceedings publications, and co-authored 4 research reports (SPR series). In Annex no. 4 entitled: *List of published scientific works and information about educational achievements, research collaboration and science popularization*, I provided a bibliographic description of my major works – altogether 53, including 29 works written individually and 24 works written collectively.

In the period before receiving the PhD degree (2007–2010) I wrote (collectively) the total of 6 works, including 4 articles (3 of them with IF) and 2 chapters in monographs.

After receiving the PhD degree (in the years 2011–2019) I have written the total of 47 works (including 29 written individually and 18 written collectively). Thirty eight of the works have been published. These include:

- 11 articles in scientific journals with IF, including 8 written independently;
- 14 articles in scientific journals without IF, but with points awarded, including 10 written independently;
- 9 chapters in monographs, including 7 written independently;
- 4 conference papers indexed in the Web of Science database, all written independently.

The structure and numerical expression of my scientific and research achievements for the period respectively before and after receiving the PhD degree are presented in table 1.

Table 1. Overall numerical presentation of scientific and research achievements

Type of work	Before receiving PhD degree			After receiving PhD degree			Altogether		
	Ind.	Col.	Total	Ind.	Col.	Total	Ind.	Col.	Total
Articles with IF	0	3	3	8	3	11	8	6	14
Articles without IF	0	1	1	10	4	14	10	5	15
WoS papers	0	0	0	4	0	4	4	0	4
Chapters in monographs	0	2	2	7	2	9	7	4	11
Editing works	0	0	0	0	5	5	0	5	5
SPR reports	0	0	0	0	4	4	0	4	4
Total	0	**6**	**6**	**29**	**18**	**47**	**29**	**24**	**53**

After deducting the share of the works' co-authors, the total score amounts to 378.4 points, including 195.5 points for articles in journals with IF. The score for the period after receiving the PhD degree amounts to 352.3 points, including 177.5 points for articles in journals with IF.

The total IF amounts to 9.027, including 7.454 after receiving the PhD degree.

Depending on the database, the bibliometric factors (number of works, number of citations, Hirsch index) characterizing my scientific and research achievements amount to:

- Web of Science: number of works: 20, number of citations: 81, Hirsch index: 4.
- Scopus: number of works: 17, number of citations: 79, Hirsch index: 5.
- Google Scholar: number of works: 20, number of citations: 136, Hirsch index: 6.

Below I present (in chronological order) my interests and scientific and research achievements with reference to the list of scientific and research achievements, provided in Annex no. 4.

Period before receiving PhD degree (2007–2010)

In the period before receiving the PhD degree my scientific interests covered the following four research topics:

1. A feasibility study of the application of artificial neural networks (ANNs) to membrane process problems;
2. The modelling of the permeability of submerged membranes;
3. The modelling of qualitative parameters after coagulation and ultrafiltration processes, by means of ANNs;
4. The application of time series to permeate flux prediction.

Re 1. At the beginning of my doctoral studies I took up a then innovative subject of applying artificial intelligence (artificial neural networks (ANNs) to be more precise) to problems relating to membrane processes, which currently are increasingly commonly used in water and sewage treatment. A thorough survey of literature on the subject, presented in a scientific article (Annex no. 4, it. E26), showed the modelling results obtained by many researchers all over the world to be promising, which justified the undertaking of this subject. The next step was to begin my own research in the field of modelling. Initially the research focused on the prediction of the pore size of ultrafiltration membranes by means of ANNs. The research was published in a monograph chapter and presented at the scientific conferences: *Membranes and Membrane Processes in Environmental Protection* (Annex no. 4, it. E11) and *Nanostructured Materials and Membrane Modelling and Simulation* (Annex no. 4, it. L6). Using the results of laboratory tests, the average radius of the pores in selected membranes was calculated. In order to create a proper ANN architecture it was necessary to propose a sequence of training data (permeate flux J, pressure P and membrane thickness l_p). At the network's output there was one neuron describing the forecasted average membrane pore radius R_{pp}. The results yielded by the ANN simulation were found to be sufficiently convergent with the experimental data, which provided a motivation for undertaking further research on the application of artificial intelligence to membrane processes.

Re 2. In 2008, when serving my internship at Technische Universität Dresden, I began to model the permeability of submerged ZeeWeed 500c membranes, using an ANN.

I used real experimental data provided courtesy of the TU Dresden research team. The first modelling results were reported in an article published in a scientific journal with IF (Annex no. 4, A14). The permeability of the submerged membrane during the hybrid treatment of surface water in the coagulation/ultrafiltration system was conducted using different multilayer perceptron structures with one hidden layer. The turbidity of the feed solution, the pH, the temperature and the cross-membrane pressure were the input signals. Owing to the use of the ANN, the behaviour of the whole system could be predicted very precisely. The modelling of membrane permeability was continued over the next two years under an extended range of investigations. The final observations and conclusions were presented in one part of my PhD thesis.

Re 3. The experimental data provided by the TU Dresden research team were also used to run simulations of artificial intelligence application to predict the qualitative parameters (e.g. turbidity) after the integrated coagulation/ultrafiltration process. Moreover, the ANN structures obtained for the ultrafiltration water treatment and the integrated coagulation/ultrafiltration process were compared. It was shown that alum coagulant dosing had no significant bearing on the network architecture, e.g. on the number of hidden neurons. In the measurement run without coagulant dosing the convergence achieved during the verification of the network model amounted to 0.99, while in the second run the correlation coefficient amounted to 0.82. This was due to the different number of training epochs, which undoubtedly affects prediction quality. It was found that the use of more complicated activation functions did not at all translate into better simulation results. Therefore it is not recommended to use one universal ANN structure to predict permeate turbidity and the turbidity retention coefficient. Instead, each time one should analyse the variation of the input signals. The research results were published in a monograph chapter and presented at the conferences: *Membranes and Membrane Processes in Environmental Protection* and *Membrane Science and Technology Conference of the Visegrad Countries PERMEA 2010* (Annex no. 4, E10 and L4). In the course of the analyses a problem connected with the number of data used for model building was encountered (the qualitative variables were recorded at a relatively large time step), which was also noted in the work on the prediction of water turbidity after the treatment process (Annex 4, A13). It was found that the discrepancy between the measured values and the predicted ones could be due to the too small number of training data. Therefore it is suggested to use a larger measurement database to create a an optimal artificial neural network model. The modelling, using the ANN algorithm, of water quality after membrane processes was described at more length in another part of my PhD thesis.

Re 4. In parallel to the above investigations I was conducting feasibility studies of predicting, by means of ANN time series, the permeate flux during the ultrafiltration of detergent wastes. The input signals were: the recirculated flux, pressure of the feed flux, pressure of the concentrate and the temperature of the detergent wastes. Except for a few cases of zero permeate flux values, the convergence between the experimental data and the predicted ones was satisfactory, amounting to 91–94%. The research results were reported in an article published in a scientific journal with IF (A12), and an expanded version of this work was included in my PhD thesis.

Period after receiving PhD degree (2011–2019)

After receiving the PhD degree I continued the previously initiated research topic no. 3 (*Modelling of qualitative parameters after coagulation and ultrafiltration processes, by means of ANNs*) and reported the results of this research in a scientific journal with IF, listed in the JCR database (A11). A review of contemporary knowledge relating to the application of ANNs in broadly understood environmental engineering was presented in a publication (E25) which I co-authored with prof. Malgorzata Kabsch-Korbutowicz who was the supervisor of my PhD work.

After defending my PhD thesis in 2010 I was employed in my parent Faculty, as part of the Water Supply and Sewage Disposal Team at the Scientific Department of Water Supply. As a result, I considerably broadened my scientific interests, previously typically technological and now covering network problems relating to water supply and sewerage systems. I undertook to solve 5 new research problems (5–9), in particular such as:

1. Operational reliability, technical condition and operation of sewerage networks;
2. Benchmarking in the water supply and sewerage sector;
3. Operational reliability, technical condition and operation of water-pipe networks;
4. Modelling the level of failure of water-pipe networks by means of the ANN method.
5. Modelling the level of failure of water-pipe networks by means of regression and classification methods.

The above research topics, taken up after receiving the PhD degree, were initially investigated in the Scientific Department of Water Supply (headed by Halina Hotloś, hab. D.Eng.), and after Faculty reorganization in 2014, at the Department of Water Supply and Sewage Systems (headed by prof. Andrzej Kotowski).

It should be noted that the decided majority of my published works are single authored and my share in the co-authored works is dominant (amounting to 40–70%). Besides being published in scientific journals or presented at domestic and international conferences, the results of my research were also used in my lectures for students and PhD students at the parent Faculty (Annex 4, pt. III I), as well as at other universities. The lectures were part of my research and teaching internships (Annex 4, pt. III L), which I completed in order to improve my scientific qualifications. Moreover, I was awarded a scientific and research scholarship under *Human Capital Operational Programme – strengthening and development of the educational potential of universities. The development of the research-educational potential of the young academic staff at Wroclaw University of Science and Technology* to realize some of the research topics taken up after receiving the PhD degree (Annex 4, pt. III A).

Re 5. At the beginning I became interested in the operational reliability of sewerage networks and I carried out a review of the state of knowledge relating to this problem (Annex 4, E9), from which the following conclusions (presented at, i.e., the conference:

Interdisciplinary Problems in Environmental Protection and Engineering) emerged: using proper damage classification methods one should be able to explicitly describe the technical condition of sewage disposal systems and the operational reliability of sewage system should be assessed comprehensively, using not only typical statistical methods, but also mathematical calculations based on advanced methods of determining the probability of occurrence of a given system malfunction.

The aim of the next analyses concerning the technical condition and operation of sewage disposal systems in Poland (E7) was to find the missing links in the investigations, which was discussed in a paper delivered at the conference: *New Technologies in Water Supply and Sewerage Networks and Installations*. The paper presented and discussed the contemporary state of knowledge about the main causes and effects of failures of and the number and type of damage to municipal sewerage networks, including the ones located in mining areas. Most of the inspections were carried out using the video technique. Work E7 presented a methodology for calculating the rates of point and linear failures of sewers. On the basis of the calculation results the condition of combined sewers, sanitary sewers and storm water sewers, made of concrete, stoneware, PVC and brick, varying in their cross section and age, was assessed. Those were not only sewers which had been in service for, in many cases, over 100 years, but also newly built ones inspected before commissioning. It was indicated that systematic inspections were necessary to prevent serious failures, and also to acquire information needed to develop a strategy for the operation and upgrading of the sewerage networks. Moreover, in an industry journal (*Gas, Water and Sanitary Engineering* – Annex 4, E22) the level of failure of the sewerage networks was compared with benchmarking data. Conclusions emerging from the above research should be taken into account in the operation and modernization of the existing sewerage systems and in the design of new such systems. The literature data presented in the above work show that gravity sewerage network failure frequency studies and their range in Poland are very limited and often inadequate. This is due to poor quality of operational data and the lack of a uniform methodology for classifying and assessing the type and effects of the abnormalities in the functioning of the sewerage systems and their components. Benchmarking will appear again in my further research, but in connection with water distribution systems (which is discussed below).

Having thoroughly explored the state of relevant knowledge, I began my own research on the technical condition of the sewerage networks. First the sewerage networks in Głogów were investigated. The information on the way in which the sewage disposal system in the city operates was collected in (E8) and presented at the periodic conference: *Interdisciplinary Problems in Environmental Protection and Engineering*. The next step in this research was to be the modelling of sewerage network reliability indices, but because of the inadequate operational data I could not realize this at the time. This research topic was continued in an article published in a scientific journal with IF (Annex 4, A10), where I presented the history and characteristics of the sewerage system in Wrocław, which is a gravity sewerage system with pressure elements (pumping stations, rising mains). It was calculated that on average 7.3–8.3 sewer failures per month would occur, while the annual average intensity of failures amounted to 0.07 per kilometre of sewage canals and combined sewers (without storm sewers).

The technical condition and operational reliability of sewerage systems are inseparably bound up with video pipe inspections. The results of the video camera inspection of the sewerage network in Wroclaw were discussed in two articles published in an industry journal (*Instal*) – Annex 4, E20 and E21 and in a conference proceedings chapter (E4), as well as presented at the *IWA International Conference for Young Water Professionals* (L3). The main conclusions emerging from the above analyses are as follows. Firstly, it is necessary to keep all the sewerage systems in good condition. Particularly the storm sewer system belongs to the critical infrastructure in recent years. Secondly, considering the increasingly frequent sudden and excessive rainfalls, sometimes causing severe damage also on the earth’s surface, special care should be paid to the condition of storm drainage systems. The malfunctions and failures presented in (E4) are typical and they occur in both the combined system and sewers. Only selected types of failures were reported since the aim was to show the scale of the problem with which the operators of the sewerage networks have to deal on a daily basis.

Re 6. Benchmarking in the water supply and sewage sector completes the research on the operation, design, modernization and operational reliability of municipal system since it makes it possible to compare the basic benchmarks proposed by IWA for different settlement units, regions and countries. A comparison of the situation in this regard in Poland with that in selected regions in Europe was presented in (E24). Then a comparative analysis for two selected water distribution systems was carried out and presented in an article (E23) in an industry journal, in a conference proceedings chapter (E5) and at the conference: *Current Problems in Water Treatment and Distribution*.

Re 7. In parallel to the benchmarking studies I began research on determining the level of operational reliability and failure frequency of water supply networks on the basis of the available operational data. The main achievements in this regard were published in a scientific journal with IF (A9). The results of the investigations into the operation of the water supply networks in Głogów and in other cities in Poland showed that from the many factors involved, the material and diameter of the pipes, the period and conditions in which they were built, the season and the water pressure and its fluctuations had the decisive effect on the frequency of failures of the conduits. In general, the water supply networks in Poland are characterized by a much higher failure frequency than the networks in other countries. This is due to several reasons, mainly to the underinvestment in pipe replacement and to many years of neglect in the systematic maintenance and modernization of the networks. The current upgrading of the water supply networks, consisting in the renewal or replacement of the depreciated cast iron or steel pipes, contributes to a reduction in the frequency of network failures, lessens the possibility of secondary pollution of the water during failures, improves the quality of the water supplied to customers and reduces network pressure losses by eliminating the negative effect of incrustation and the deposits formed on the inner walls of pipes, reducing water losses due to leakage from untight and damaged pipes, reducing water network operating costs and improving reliability and operational safety of the water supply systems.

The research on determining the failure rate on the basis of the available operational data was then carried out on other water distribution systems. The research results were published in two works (E16 and E17). The main assumptions and conclusions (E17) can be summarized as follows. The failure rate of the distribution pipes and the house connections in the years 2008–2014 was determined on the basis of operational data made available by one of the water companies. The failure rate of the pipelines was determined in one of the designated zones of the analysed water distribution system. The average failure rate in the analysed period amounted to 0.18 fail./(km·a) for the distribution pipes and 0.44 fail./(km·a) for the house connections. In article (E16) a comparative analysis of the failure level of the water supply network in a large city and two medium-sized cities in Poland was carried out. The main conclusion emerging from this research is that it is necessary to systematically replace the old lateral service lines with new pipelines made of plastic.

The analyses of the operation and reliability of the water distribution systems also provided the basis for producing team reports (J3 and J4) containing the results of research conducted as part of the Faculty's statutory activities.

In the next step of my academic development I took up the subject of modelling as applied to water supply networks. My first attempts in this field were described in an article with IF (A8), in which I undertook to modify an existing model describing the failure rate. The results turned out to be promising, but only for the particular water distribution systems. In order to model other water supply networks it would be necessary to modify the model by including the parameters describing the actual conduits.

Re 8. The modelling of the failure rate and the availability factor of water conduits by means of artificial neural networks turned out to be one of the principal research topics. This subject was discussed in detail in a series single-authored publications in scientific journals with IF (A4–A7), in monograph chapters (E3 and E6), in an industry journal (E15) and in a conference paper indexed in Web of Science (E29). Moreover, the subject of applying ANNs to the analysis of the level of failure of water conduits was the basis for a grant for research under the young academic staff development project (J5). The published results of this research indicate that artificial neural networks are a valuable tool for analysing and assessing the level of the operational reliability of water distribution systems. In the above works different artificial neural network structures and various combinations of input signals were considered in order to determine the usefulness of this kind of modelling, even when only a small number of operational data are available. The modelling results were described in statutory research reports (J1 and J2). The conclusions emerging from the research on the application of artificial intelligence turned out to be promising and so I decided to extend the range of research to cover other prediction methods, as described below.

Re 9. Recently I became interested in the application of machine learning methods (in failure level analysis), i.e. the support vector machines (SVM) method, the K-nearest neighbour (KNN) method and the regression (RT) and classification trees (CT) method. In addition, I checked the potential of the MARSplines method for the modelling of the failure rate of water conduits. The application of the prediction methods to the modelling of the failure rate of water conduits was described in a series of single-authored articles published in scientific journals with IF (A1–A3), in conference monograph chapters (E1 and E2), in journals not included in the JCR database (E12–E14, E18 and E19) and in conference papers indexed in Web of Science (E27, E28 and E30), and it was presented at two international conferences (L1 and L2) – *IWA Eastern European Young Water Professionals Conference*. In some publications (E18 and E19) a comparative analysis of the SVM method and the ANN method was carried out. In most of the published works the advantages and disadvantages of the prediction methods and their application value were indicated. At the current stage of the research on the operational reliability of municipal systems, modelling and predicting the rate and type of failure of water conduits seems to be an innovative and particularly vital when decision must be made quickly in cases of serious failures. One should bear in mind that any modelling carries a forecast error. The choice of an optimal model should be contingent not only on achieving the best possible convergence, but also on the assessment of the effect of an incorrect estimate. The scale and effects of a failure of, e.g., distribution pipes are incomparably greater than those of house connections. In most of the above mentioned works on prediction methods the need for further research in this field was indicated. Therefore I decided to extend the modelling methodology (e.g. to include more detailed independent variables) and the applicability of the regression and classification methods to different water distribution systems, which led to broader considerations and to the presentation of the results of my latest research in my habilitation monograph.

Besides typically research (and teaching) tasks, my professional activity included other tasks.

In the years 2015–2016 I co-edited books – in the *Interdisciplinary Problems in Environmental Protection and Engineering* series (F4 and F5), and then in the years 2017-2018 in the *Interdisciplinary Problems in Environmental Protection and Engineering* series conference proceedings (F1, F2) indexed in Web of Science.

In 2017 I was a co-editor of the *Advances in Energy Systems and Environmental Engineering* series conference materials (F3) indexed in Web of Science.

In the years 2017–2019 I was a member of the scientific committee of domestic and international conferences (Annex 4, pt. III N). Since 2012 I have reviewed articles for publication in 10–20 scientific journals, most of which are listed in the JCR database (Annex 4, pt. III P).

signature of Mover