Measurement 149 (2020) 107026

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Reliability modelling and analysis of water distribution network based on backpropagation recursive processes with real field data

D. Vališ [a,b], K. Hasilová [c,*], M. Forbelská [d], Z. Vintr [a]

[a] Department of Combat and Special Vehicles, University of Defence, Brno, Czech Republic
[b] Faculty of Transport and Computer Science, University of Economics and Innovation, Lublin, Poland
[c] Department of Quantitative Methods, University of Defence, Kounicova 65, Brno CZ-66210, Czech Republic
[d] Department of Statistics and Operation Analysis, Mendel University in Brno, Brno, Czech Republic

ARTICLE INFO

Article history:
Received 2 May 2019
Received in revised form 29 July 2019
Accepted 1 September 2019
Available online 6 September 2019

Keywords:
Time series modelling
Dynamic linear models
Water distribution network
State-space models
Analysing field data
Kalman type recursor

ABSTRACT

Any water distribution network (WDN) is a key element in critical infrastructure. As time series data mining and modelling are developing apace, these provide promising tools for the research of water resources management. The aim of this article is to explain how incomplete information on the failures in water mains can be used and processed effectively. The data available and under investigation are the truncated failure field data of a regional water distribution network recorded during the last 15 years. We introduce the application of novel, non-trivial dynamic backpropagation recursive time-series models which are later successfully validated by the recorded data. Real field data of a WDN were used while these data were elaborated with a backpropagation Kalman recursor. The results can be used for predicting the reliability of a WDN, or as inputs into a water management system to optimise maintenance, crisis management and emergency planning.

1. Introduction

A water distribution network (WDN) is an important part of the critical infrastructure (CI) of a country. Needless to say, the distribution system is a very complex and vulnerable element of all CI networks. It usually consists of a pipeline placed in the ground and is of variable age, material, quality of installation, etc. The failure of a WDN commonly means that a large number of people will be without water supply, that there is a need for direct onsite action and that the system function restoration costs will be considerable.

The quality of water is monitored at the source and at the point where water enters the distribution network. Usually it is monitored at close and regular intervals [29]. Water quality records can be understood as a typical time series where the records are made in equidistant steps. Subsequent mathematical processing and prediction also fit this model, since there are numerous methods used for modelling time series related to WDNs.

Another very important property is distribution network dependability – mainly availability and reliability. When it comes to pipelines made of different materials, it is difficult to determine and monitor their real technical condition. The maintenance/diagnostics procedure usually consists of monitoring field parameters such as pipeline pressure (its deviations), or comparing the amount of water in the input and output [7]. Pipeline maintenance is carried out by removing sediments from the network (mechanically) or replacing pipe sections. In this case, the systems will most likely be connected to incongruent counterparts because of their different ages and materials.

In the available technical literature, some of will be discussed later, we can find numerous results of modelling time series of WDNs. Most of the time series, however, are focused on carefully observed and regularly recorded characteristics of water quality or daily demand. For this purpose, very sophisticated systems of measurement, observation and telemetry are applied. When it comes to observing and assessing distribution system condition or reliability, published results are rather scarce. Determining the degree of pipe network reliability is not easy and straightforward either, requiring, for example, directly measuring the water quality or the amount of water consumption.

1.1. Motivation

A failure is an event when the system – a WDN – is not able to meet requirements, i.e. the quantity and quality of water are not

* Corresponding author.
E-mail addresses: david.valis@unob.cz (D. Vališ), kamila.hasilova@unob.cz (K. Hasilová), marie.forbelska@mendelu.cz (M. Forbelská), zdenek.vintr@unob.cz (Z. Vintr).

https://doi.org/10.1016/j.measurement.2019.107026
0263-2241/

provided as required. A typical pipeline rupture, armature leakage, corrosion, etc. might serve as an example. As a system, this is a specific situation where a continuous structure is affected by failures in time (which is also continuous); therefore, it is necessary apply modified approaches for state-space models.

Typical state-space models are based on purely Gaussian assumptions, provided we take into account equidistance of time data and normality of residuals. In our case, however, these are more likely Poisson processes, since we observe the event occurrence depending on a continuous random variable which is not simply time, but also pipeline length. Another thing is that for common Gaussian state-space models, the process mean value is obviously independent of dispersion and vice versa. However, the thing which has to be taken into consideration is that the mean value of a Poisson process depends on the dispersion. Moreover, reliability measures such as failure rate are a mathematical pseudo-variable, not measured directly. Hence, if we are to model this as a time series using a state-space model, we need a very specific approach.

In spite of numerous modernisations and improvements, there is still difficulty in recording data, in particular failure data. The aim of this article is to show that incomplete and insufficient information on the failures of a WDN might be used and processed efficiently. We introduce a new approach to the use of specific field data of a complex CI structure. The main aim and contribution of the article then is to: (i) illustrate a way to model reliability measure development, (ii) specify how to estimate and predict the next possible behaviour of a system regarding its function and (iii) introduce a way to determine failure occurrences during the following period.

After the introduction, we focus on the state-of-the-art in the area of time series, state-space models and their relation to WDN. Next, we describe the form of field data on network failures. After this, we introduce a modelling methodology appropriate and suitable for this situation. Finally, there is space for results, discussion and conclusions.

2. State-of-the-art

While studying current and applicable publication sources, attention was paid to authors who deal with state-space models and time series. These approaches were searched in relation to WDN failures and their reliability modelling. Comparing published and achieved results with our intention was the main aim of the sources analysis.

2.1. General approaches to modelling WDN

Approaches to modelling of WDN include data mining, wavelet analysis, artificial neural networks, fuzzy inference systems, pattern recognition and specific methods for missing data reconstruction.

Deng and Wang [6] proposed a two-part framework based on a time series data mining. The approach is based on dividing the time series into ‘clouds’ and their analysis, which the authors tested on historical time series data.

Mounce et al. [28] proposed methods based on wavelet-based semblance analysis for monitoring water flow-pressure and turbidity for assessment of the risk of fouling and water deterioration. A case study had been performed on the UK WDN.

Another approach to water quality assessment is presented by Perelman et al. [36]. The authors introduced a general framework which composes data-driven estimation with sequential probability updating – an approach based on artificial neural networks is to indicate deviations in water quality time series, which was tested on real data.

Mounce et al. [30] presented an approach based on pipe system monitoring characteristics, i.e. flows, pipe pressure and water quality. They transformed time series data into vectors and applied pattern matching to automatically recognise specific deviations.

For large volume of water records data, Mounce et al. [29] proposed an automatic identification of new or abnormal patterns embedded in the data. They introduced support vector regression as a learning method for anomaly detection from water flow and pressure time series data.

Harding and Walski [13] described three modelling principles which were based on long time series analyses to describe contamination in the WDN of a large city. The EPANET and matrix approaches were applied for a wide variety of possible contaminant concentrations at the original source.

Firat et al. [7] applied two types of fuzzy inference systems (FISs) in order to predict municipal water consumption time series. The authors included both an adaptive neuro-fuzzy inference system (ANFIS) and a Mamdani fuzzy inference systems (MFIS) into their study. Water consumption models were constructed using a combination of the past values of the water consumption.

Barrela et al. [4] presented a simple and apparently highly effective method of reconstruction of missing data in flow time series. Their method uses a weighted function for forecast and backcast obtained from existing time series models. The authors performed a number of tests in order to demonstrate the methodologys applicability. A more general case was treated by Mora et al. [10]. The authors used Farrington’s model for censored data and applied it to the data containing information on water pipes.

It is obvious that time series are an important area of interest in this field. However, these works, inter alia, clearly show that most time series are focused on measurable and rather automatically achieved indicators, in particular when it comes to water quality. When dealing with modelling and further mathematical processing, the tools used are mostly mathematical soft methods such as fuzzy logic, neural networks and Bayesian networks.

2.2. Pipelines failures modelling

In our further research into the relevant literature, we concentrate on results analysing the failure of a WDN. Failures modelling comprises, most often, simple statistical assessment, simulations and methodology examining effects connected with the failures.

Pietrucha-Urbanik [38] dealt with financial efficiency in the case of water network maintenance. In a decision-making process, the author used a relatively simple statistical approach, assessed failure occurrence in the network and finally combined it with a proportional hazard method.

Asadi and Melchers [3] studied a sample of 10 exhumed pipelines from different locations after 34–129 years of service. They looked for confirmation on a Gumbel probability distribution of failure on such pipes. Ossai et al. [34] studied corroded pipelines and modelled their inspection and repair by Markov models.

Pietrucha-Urbanik and Studzinski [40] showed a simulation approach of WDN failures. When simulating and predicting the failure, they used a hydraulic model which took into account the effects of pumping stations with expansion tanks located in the network.

Piegdon et al. [37] introduced different possibilities of applying a geographical information system (GIS). Their major concern was to analyse the risk of failures in the water pipes of a WDN. Based on this information, they selected the pipelines with the highest failure risk, and therefore with the highest possibility of water supply interruption.

Pozos-Estrada et al. [41] presented results of three types of analysis created in relation to failures of WDNs: theoretical, experimental and in situ investigation. Pipeline failures occurred when the supplying system collapsed, which resulted in the failure of pre-stressed mains. In their work, they assessed how air pockets in a pipeline affect distribution system failure.

Wols and van Thienen [48] examined how climatic changes might affect the integrity of drinking WDN. The failure there was not only a delivery failure, but also a physical water network system failure. Studzinski [39] focused on the cost of water pipe failures. The analysis introduced in their work was based on real data from a regional system and included profit loss and corrective maintenance cost. Arsenio et al. [2] showed how topsoil deformation affects utility lines, i.e. a water network. For this purpose, they used ground movement data and set a replacement-prioritisation map in order to restore the network effectively.

Manuco et al. [23] presented a novel methodology for optimizing the inspections of large underground infrastructure networks based on failure likelihood, failure severity and incomplete information about the network features and parameters. They illustrated this approach on a real case of the large-scale maintenance optimisation problem of the underground sewerage network in Espoo, Finland. Williams and Kuczera [46] showed the results of a deeper understanding of the operational association with failure. Their research dealt with a regional WDN in Australia and was focused on deterioration and failures of iron pipes.

2.3. Failures modelling of WDN

In this section we searched through current and relevant results dealing with failures in a WDN. Most of the approaches are generic, but they are not always based on real data and easy to use in real conditions when it comes to a decision-making process, maintenance optimisation and/or crisis management.

Hwang et al. [16] dealt with an alternative and modified FMECA method used for assessing the risk of failure occurrence regarding the resilience of regional WDN in the USA. Kutylowska and Hotlos [19] presented fault diagnostics of a regional WDN in Poland; the presented method is based on a relatively simple statistical assessment. Francis et al. [8] used a Bayesian Belief Network to construct a knowledge model for pipe breaks in a water zone. The machine learning model was trained from pipe breaks and covariate data from a mid-Atlantic United States drinking WDN. Winkler et al. [47] presented a framework for processing and predicting WDN failure data. The authors applied boosted decision trees to predict failures and tested the approach on real data.

When searching through relevant literature, our major area of interest is processing the field data of a WDN. We have examined not only the approaches dealing with the prediction of a system condition, see Shamir and Howard [45], Herz [15], but also the approaches dealing with the prediction of maintenance data based failures, see Gat and Eisenbeis [11,44].

When assessing the condition and the behaviour of a WDN, we were highly interested in applying state-space models and back-propagation recursive approaches. There are not many works tackling this area; however, some applications of a generic approach to modelling of time series with a Kalman filter can be found in [22,12,17,33].

Our approach is in any case novel and based on a specific form of the Kalman filter, i.e. it takes into account the Poisson character of field data dealing with available system failures.

3. Data – form and origin

The analysed data are dichotomous and quantitative. They are real and were collected during the operation of a regional network of a WDN for about one million inhabitants during almost 15 years (2000–2014). The data are incomplete since we have only the numbers of failures in the system of a given length during a relevant month of each observed year. Unfortunately, no more information about the failure or the maintenance is available. An example of the data is shown in Table 1. When analysing the data, we decided not to work with single months but with quarters, since less subtle division is equally effective. Moreover, we assume that it will be possible to observe the effect of the environment and weather (seasonality) upon failure occurrence during single quarters. In order to demonstrate a suitable reliability measure, we decided to use failure frequency (failure rate/failure intensity) as a pseudo-variable. We also suppose that the form of the hazard rate function will help us to estimate retrospectively the form of a probability distribution function, or a probability mass function.

The total course of failure rate during a single quarter is shown in Fig. 1. The course is completed by trend smoothing curves: lowess and cubic spline, which are very similar.

Next, in Fig. 2 we used box plots to show a preliminary analysis of a time series based on robust estimates of yearly or quarterly failure rate levels (thick lines in the boxes illustrate a median value) and on robust estimates of failure rate variability during single years or quarters (see the box containing 75 % of the values of a given period). The left graph (panel) shows that the failure frequency decreases over time, which indicates that the system condition 'gets improved', perhaps due to gradual modernisation of

Table 1
Example of field data – number of failures F_t during a respective month and failure frequencies (failure rate) ff.

Year	2000		2001		2002		2003	
Month	F_t	ff	F_t	ff	F_t	ff	F_t	ff
Jan.	6		10		6		13	
Feb.	12	0.5	11	0.389	11	0.344	7	0.311
Mar.	27		14		14		8	
Apr.	23		10		4		13	
May	18	0.538	12	0.341	4	0.121	4	0.209
Jun.	8		9		3		2	
Jul.	12		16		7		4	
Aug.	8	0.348	12	0.413	10	0.261	8	0.250
Sep.	12		10		7		11	
Oct.	17		10		6		12	
Nov.	24	0.630	5	0.239	6	0.283	7	0.337
Dec.	17		7		14		12	
Total	184		126		92		101	

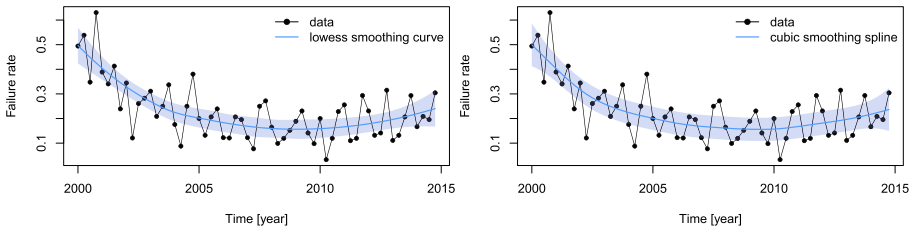

Fig. 1. Failure rate data of observed system: left with lowess trend line and 95% confidence intervals; right with cubic spline and 95% confidence intervals.

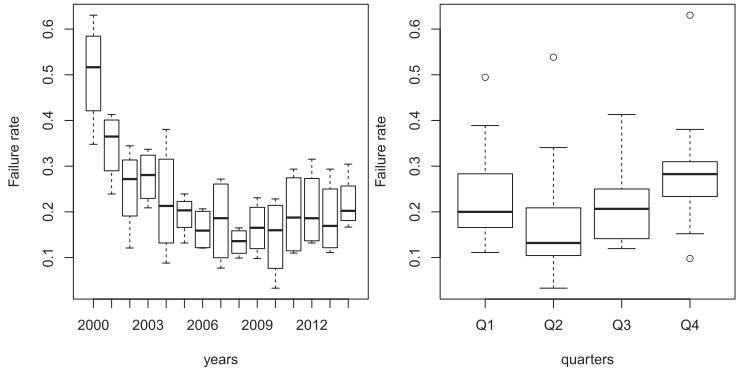

Fig. 2. Visualisation of robust estimates of failure rate levels and of a failure rate variability for each year (left panel) and quarter (right panel).

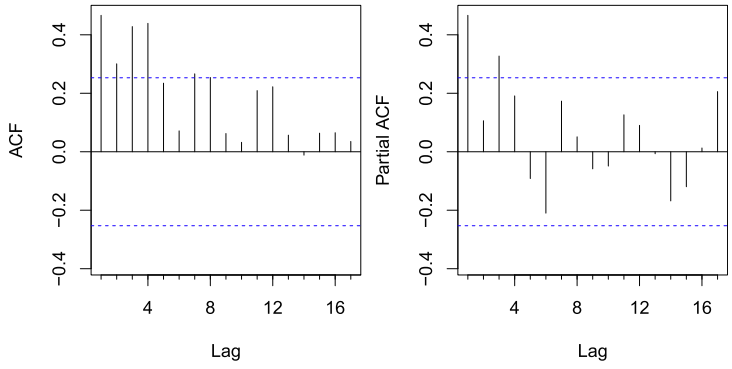

Fig. 3. Course of autocorrelation function during respective quarters (left) and partial autocorrelation function (right).

network lines, and that the failure rate variability fluctuates during single years. The right graph (panel) shows that the failure rate is lowest in the second quarter and highest in the fourth quarter, whereas the failure rate variability during the quarters/seasons is more or less the same.

Having in mind the following mathematical operations, we perform a data analysis with respect to an autocorrelation function (ACF) and a partial correlation (Partial ACF). The main goal of correlation analysis is to find out whether and how such time series can be predicted from its own history, and based on this knowledge a suitable model can be chosen. The results shown in Fig. 3 clearly show that the data are independent.

4. Modelling methodology

When modelling the system behaviour from the reliability point of view, we will follow the approach illustrated in Fig. 4. The first four steps have already been described above, and the remaining steps will be described in view of the applied theory.

Using the recorded data and the calculated failure rate, we set a time series which will be modelled with a dynamic backpropagation recursive algorithm. As demonstrated below, there are a few possible variants suitable for use, as the results of the performed AIC and BIC tests and residual analyses show. The nature of the data and their visualised properties indicate that they might be non-homogeneous Poisson data of a Markov type.

4.1. Theory

4.1.1. Model of the reliability measure

Albeit we recognise some typical or specific forms of time series data driven models, such as auto regressive methods (AR), moving average methods (MA) and their combinations, such as ARMA and/or ARIMA (auto regressive integrated moving average) methods, etc. we tend to work here with a more specialised form which generalises all the above mentioned ones – the Kalman recursor. We further develop the initial benefits of the Kalman recursor. The Kalman recursions associated with state-space representations are important tools when analysing data of time series type in various areas [5,27]. These procedures were originally developed to control systems of linear type. The state-space model is ideal for representing

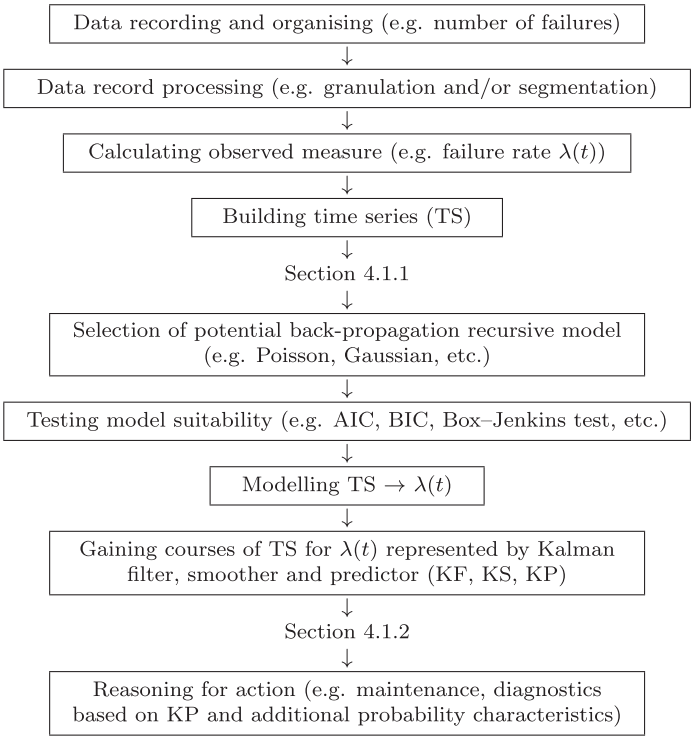

Fig. 4. Methodology of elaboration of time series data on WDN failure.

some types of time series. Such a form induces that the series $\{Y_t, t = 1, 2, \ldots\}$ satisfies an equation of the form

$$Y_t = G_t X_t + W_t, \quad t = 1, 2, \ldots, \tag{1}$$

where

$$X_{t+1} = F_t X_t + V_t, \quad t = 1, 2, \ldots \tag{2}$$

The Eq. (2) describes the system state X_t evolving at time t (a $v \times 1$ vector) based on matrices $F_1, F_2, \ldots$($v \times v$ matrices), which are known, plus vectors of random type which are in sequence $X_1, V_1, V_2, \ldots$ Eq. (1). It then defines an observation sequence, Y_t, which is obtained by a linear transformation of X_t and the addition of a random noise vector, $W_t, t = 1, 2, \ldots$

Below, there are some initial rules needed when analysing equations of both state (2) and observation (1):

- $F_1, F_2, \ldots$ is a sequence of given $v \times v$ matrices.
- $G_1, G_2, \ldots$ is a sequence of given $w \times v$ matrices.
- $\{X_1, (V_t', W_t'), t = 1, 2, \ldots\}$ represents random vectors which are in an orthogonal sequence. Moreover, they have finite second moments. (If matrix $E(XY')$ is zero, X and Y, which are random vectors, are claimed to be orthogonal and we write $X \perp Y$.)
- $EV_t = 0$ and $EW_t = 0$ for all t.
- $E(V_t V_t') = Q_t, E(W_t W_t') = R_t, E(V_t W_t') = S_t$, where $\{Q_t\}$, $\{R_t\}$ and $\{S_t\}$ are specified sequences of $v \times v, w \times w$ and $v \times w$ matrices respectively.

As we are aware of more potential approaches in time series modelling, both statically and dynamically, and especially in using Kalman recursors, we select and apply suitable forms on our data structures. For later comparative reasons, we first use typical time series models of dynamic character. They are called the Local Linear Trend (LLT) model and the Basic Structural Model (BSM). These models follow the rules:

a) LLT Model

Series, which is observed:

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim NID(0, \sigma_\epsilon^2) \tag{3}$$

Level of latent type:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim NID\left(0, \sigma_\eta^2\right) \tag{4}$$

Drift of latent type:

$$\beta_t = \beta_{t-1} + \xi_t, \quad \xi_t \sim NID(0, \sigma_\xi^2) \tag{5}$$

b) BSM Model

Series, which is observed:

$$y_t = \mu_t + \gamma_t + \epsilon_t, \quad \epsilon_t \sim NID(0, \sigma_\epsilon^2) \tag{6}$$

Level of latent type:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim NID\left(0, \sigma_\eta^2\right) \tag{7}$$

Drift of latent type:

$$\beta_t = \beta_{t-1} + \xi_t, \quad \xi_t \sim NID(0, \sigma_\xi^2) \tag{8}$$

Latent seasonal:

$$\begin{aligned} \gamma_{1,t} &= -\sum_{j=1}^{s-1} \gamma_{j,t-1} + \omega_t, \quad \omega_t \sim NID(0, \sigma_\omega^2) \\ \gamma_{2,t} &= \gamma_{1,t-1} \\ &\vdots \\ \gamma_{s-1,t} &= \gamma_{s-2,t-1} \end{aligned} \tag{9}$$

As the form and origin of our data are very specific, we have to modify the Kalman recursor and to develop new forms. The models of greatest concern to us, however, are the newly created

'Generalised Dynamic Linear Models' (GDLM) to be applied further. As we have mentioned previously, the origin and the nature of our data are by no means sufficient; therefore, it is necessary for the models to fit the situation. First, we will examine the Gaussian type model from the exponential family (GDLM-G).

c) GDLM-G

The observed series:

$$y_t \sim p(y_t|\theta_t) = p(y_t|\mu_t + \gamma_t), \quad EY_t = g^{-1}(\theta_t), \tag{10}$$

with g standing for a function which typically is called a "function of link type". Next, θ_t standing for a predictor which typically is called "linear": specifically, for the distribution of Gauss type: $g(\theta_t) = \theta_t$.

Linear predictor:

$$\theta_t = \mu_t + \gamma_t \tag{11}$$

Level of latent type:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim NID\left(0, \sigma_\eta^2\right) \tag{12}$$

Drift of latent type:

$$\beta_t = \beta_{t-1} + \xi_t, \quad \xi_t \sim NID(0, \sigma_\xi^2) \tag{13}$$

Latent seasonal:

$$\begin{aligned} &\gamma_{1,t} = -\sum_{j=1}^{s-1}\gamma_{j,t-1} + \omega_t, \quad \omega_t \sim NID(0, \sigma_\omega^2) \\ &\gamma_{2,t} = \gamma_{1,t-1} \\ &\vdots \\ &\gamma_{s-1,t} = \gamma_{s-2,t-1} \end{aligned} \tag{14}$$

Next, we will examine and then apply the Poisson type model from the exponential family (GDLM-P). In terms of the data origin and form, this model seems to be the best.

d) GDLM-P

The observed series:

$$y_t \sim p(y_t|\theta_t) = p(y_t|\mu_t + \gamma_t), \quad E(y_t|\theta_t) = g^{-1}(\theta_t), \tag{15}$$

with g standing for a function which typically is called a "function of link type". Next θ_t standing for a predictor which here is modified by Poisson distribution: $\theta_t = \log(\lambda_t)$, $\lambda_t = E(y_t|\theta_t) = D(y_t|\theta_t) = \exp\{\theta_t\}$.

Linear predictor:

$$\theta_t = \mu_t + \gamma_t \tag{16}$$

Level of latent type:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim NID\left(0, \sigma_\eta^2\right) \tag{17}$$

Drift of latent type:

$$\beta_t = \beta_{t-1} + \xi_t, \quad \xi_t \sim NID(0, \sigma_\xi^2) \tag{18}$$

Latent seasonal:

$$\begin{aligned} &\gamma_{1,t} = -\sum_{j=1}^{s-1}\gamma_{j,t-1} + \omega_t, \quad \omega_t \sim NID(0, \sigma_\omega^2) \\ &\gamma_{2,t} = \gamma_{1,t-1} \\ &\vdots \\ &\gamma_{s-1,t} = \gamma_{s-2,t-1} \end{aligned} \tag{19}$$

Professionals may have an issue with the correct selection of a suitable model. When it comes to deciding which model is suitable for use, we consider the data nature, the results of AIC and BIC tests, and the residual analysis.

4.1.2. Model of failures occurrence

A relatively extensive data set is available; therefore, we are able to calculate a hazard rate function. This can be used to approximate a probability density function to estimate the probability of the observed system time to failure. Figs. 1 and 2 show that the failure rate is not a constant function over time. We can then assume that times to failure form a counting process. This is a non-homogeneous Poisson process (NHPP) whose rate of failure occurrence depends on time [43]. Moreover, the available data are aggregated, thus the information about the number of failures over fixed intervals of time is available. [14] proved that the estimates for aggregated data are consistent – i.e. for an interval length approaching zero, the estimation for the aggregated data approaches the estimation for the data with accurate failure times.

For this reason, we use modelling and calculation based on the following assumptions [9]. Let $\{N(t), t \geqslant 0\}$ be a counting process. Its mean function, $\Lambda(t)$, is defined as $\Lambda(t) = E[N(t)]$, i.e. $\Lambda(t)$ is the expected number of failures in the interval $(0, t]$. Rate of occurrence of failures (rocof) is defined as the derivative of the mean function: rocof $= \lambda(t) = \Lambda'(t)$. The most widely used models are the power law model and the log-linear model [43,35,32]:

- The rocof of the power law model is defined as

  $$\lambda(t) = kbt^{b-1} = \alpha t^\beta \quad \text{for } k, b > 0 \quad \text{and} \quad t \geqslant 0.$$

  This model is very flexible and can model both increasing ($b > 1$) and decreasing ($0 < b < 1$) failure rates. The time to the first failure follows the Weibull distribution with k and b representing scale and shape parameters, respectively.
- The rocof of the log-linear model (also known as the Cox-Lewis model) is defined by a formula

  $$\lambda(t) = \exp(\alpha + \beta t) \quad \text{for} \alpha, \beta \in \mathbb{R} \text{ and } t \geqslant 0.$$

  The first failure time follows a Gompertz distribution. Similarly, the system has decreasing and increasing failure rates for $\beta < 0$ and $\beta > 0$, respectively.

Specific probability distributions correspond to these failure intensity models. The power law model is connected with the Weibull distribution, $Wb(b, k)$, with the density function:

$$f(t) = kb^k t^{k-1} \exp\left(-(bt)^k\right).$$

The log-linear model is usually connected with the Gompertz distribution with the density function:

$$f(t) = \beta\eta \exp(\eta + \beta t) \exp(-\eta \exp(\beta t)).$$

Usually, statistical assessment includes a probability density function of the number of failures.

5. Results and discussion

In this part we follow step-by-step the structure and methodology laid out in the previous sections. As the mathematical procedure may be not simple for practitioners, we suggest that following both the flow chart in Fig. 4 and then the theory mentioned in Section 4.1 – especially in Sections 4.1.1 and 4.1.2 can be very useful and helpful for all readers to understand the calculated principles of WDN reliability and investigated failures. In order to get an idea of the failure rate course, we first introduce

the results achieved with the LLT and BSM models mentioned in Section 4.1.1, and for this we apply the expressions (3)–(5), or (6)–(9). There are many mathematical programs that can be used for this work, but the results we obtained both in graphical and numerical forms have been achieved in R, which is free, very powerful and reliable software [42].

The numerical solution of the system of non-linear equations, which is applied when calculating the all Kalman recursor parameters, is further used to perform data convergence tests. In a similar manner, for empirical data the estimation of variability is carried out with values: for LLT: 0.006 176 431 and for BSM: 0.004 322 15. The lower value shows the nearness of the level/state of latent theoretical type to our model.

Using Eqs. (3)–(5) and (6)–(9), we form series of observed variables which are modelled by the Kalman filter (KF) for both the LLT and BSM model, see Fig. 5. Numerical results of the constructed KF, which are displayed graphically in Fig. 5, are shown further in Table 2.

As we possess relevant data, the seasonality can also be considered and studied. Thanks to the BSM, we are able to model the effect of seasonality upon the failure rate course during single quarters, see Fig. 6. The graph clearly shows that climatic conditions and temperature fluctuations typical for a winter season have a significant impact on the failure rate.

The decomposition of the KF into its elements (for details please see subSection 4.1.1), such as level or residuals, is better seen in Fig. 7, which displays these elements of our studied measure – the failure rate.

We use the Shapiro-Wilk test to assess the normality of the data. With resulting statistics and p-values: for LLT: $W = 0.979$, p-value $= 0.4154$; for BSM: $W = 0.9747$, p-value $= 0.2448$, assumption for observed data normality is not rejected. This can be seen in Fig. 8, where Q-Q plots and density estimates are shown. Also the graphical outcomes of this test were chosen due to better illustrative understanding of the results.

We can expect that the data might form some level of autocorrelation. From the residual test, auto-covariance function (ACF) and Ljung-Box statistics with corresponding p-values which are depicted in Fig. 9, we can later deduce that there might be some level of autocorrelation in the studied data.

As a next step, we calculate and construct the Kalman smoother (KS) for model types LLT and BSM and for studied empirical data – see Fig. 10 – which we use for detecting and smoothing a general trend in the data. Again, as the graphical form is the most

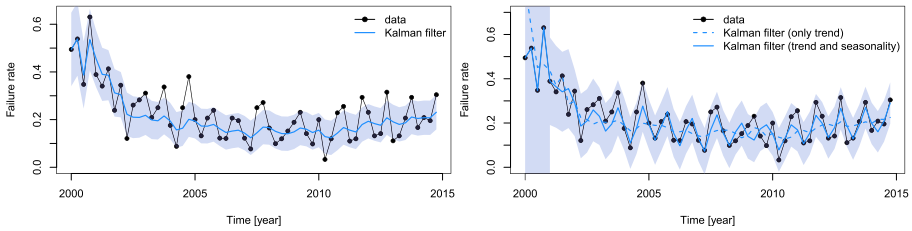

Fig. 5. Water distribution network LLT model (left), and BSM model with seasonal effects (right): Kalman filter of observed series – failure rate with 95% confidence thresholds.

Table 2
Examples of calculated time series for observed data – failure rate.

Quarter	LLT			BSM			GDLM-P		
	KF	KS	KP	KF	KS	KP	KF	KS and their trends	KP
1/2000	0.495	0.477		0.199	0.497		0.4959	0.5336	
							0.4328	0.5305	
2/2000	0.538	0.457		0.547	0.474		0.5735	0.3851	
							0.2403	0.5050	
3/2000	0.386	0.437		0.561	0.451		0.3592	0.4534	
							0.3550	0.4613	
...	...	...		...	...		...	...	
1/2014	0.205	0.216		0.198	0.211		0.1951	0.2075	
							0.1929	0.2063	
2/2014	0.208	0.221		0.217	0.216		0.1631	0.1620	
							0.2141	0.2125	
3/2014	0.207	0.226		0.216	0.221		0.2086	0.2123	
							0.2119	0.2160	
4/2014	0.231	0.231		0.225	0.225		0.2928	0.2928	
							0.2207	0.2207	
1/2015			0.2362			0.2290			0.2261
									0.2248
2/2015			0.2414			0.1774			0.1746
									0.2289
3/2015			0.2466			0.2303			0.2291
									0.2331
4/2015			0.2518			0.3078			0.3149
									0.2374
1/2016			0.2570			0.2466			0.2432
									0.2418
2/2016			0.2622			0.1950			0.1878
									0.2462

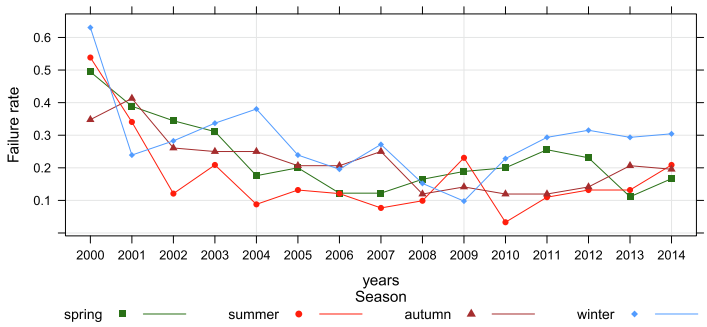

Fig. 6. Water distribution network BSM model: Seasonal influence on observed time series – failure rate.

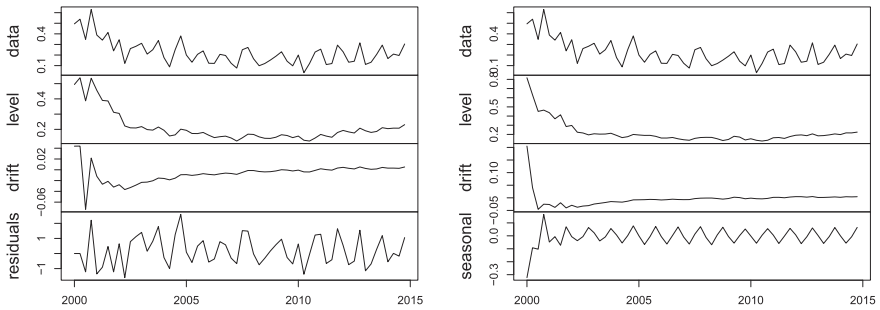

Fig. 7. Water distribution network LLT (left) and BSM (right) model: Elements of the Kalman filter of observed series – failure rate.

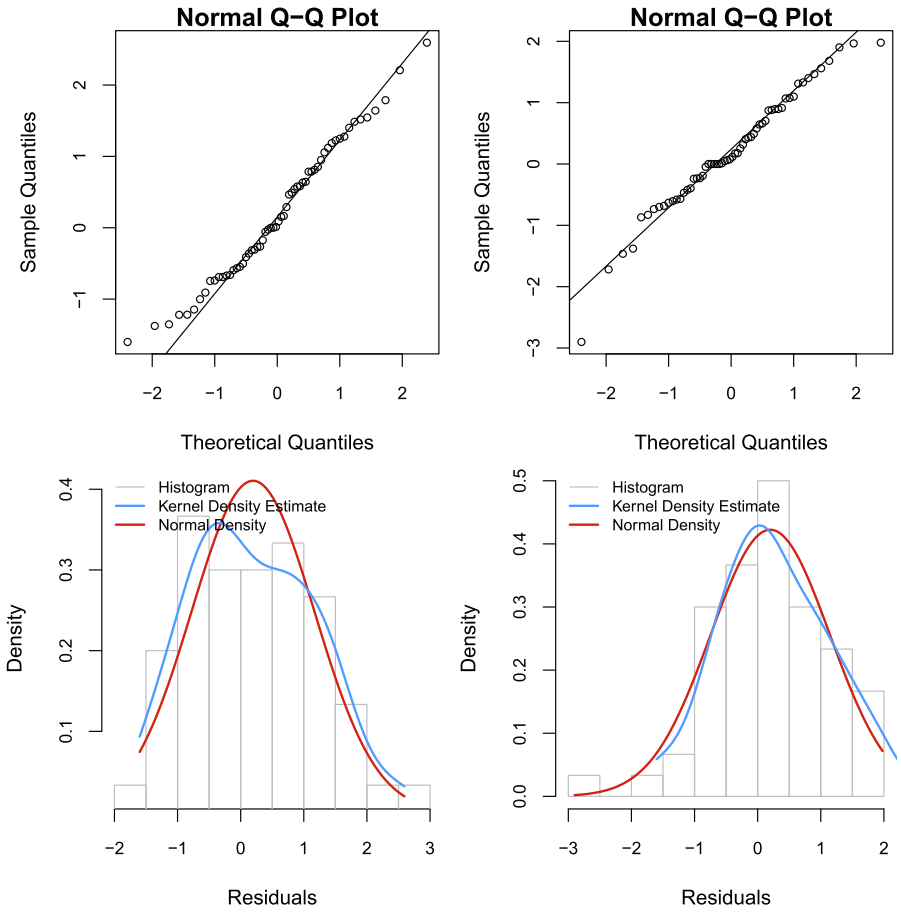

Fig. 8. Water distribution network LLT (left) and BSM (right) model: Kalman filter residuals Q-Q normal plot for empirical data – failure rate.

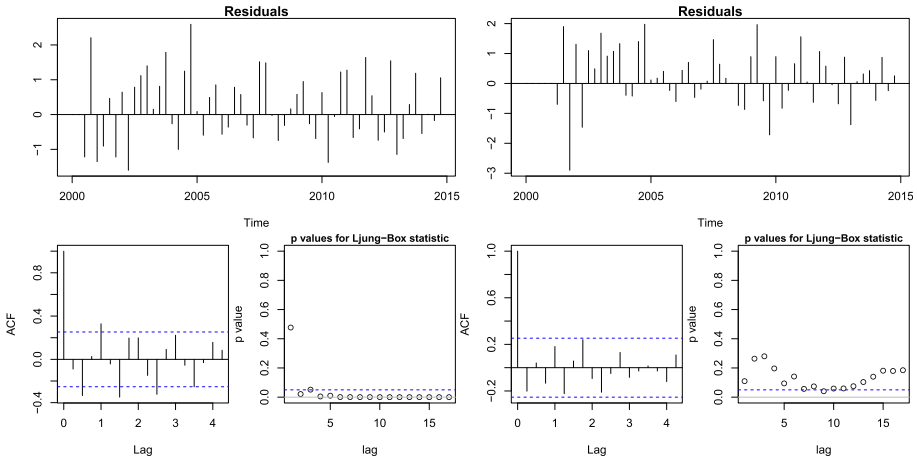

Fig. 9. Water distribution network LLT (left) and BSM (right) model: Kalman filter diagnostic plot for residuals and empirical data – failure rate.

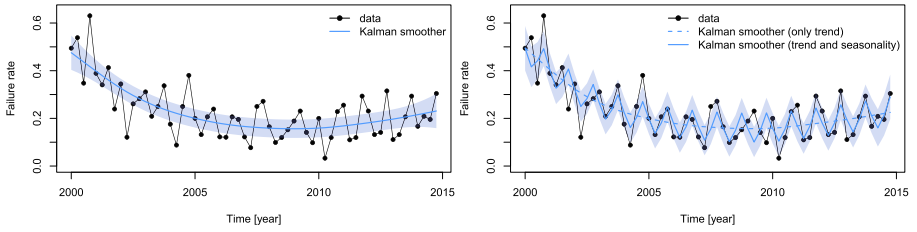

Fig. 10. Water distribution network LLT model (left) and BSM model with seasonal effects (right): Kalman smoother for observed series – failure rate accompanied with 95% thresholds of confidence.

descriptive, we can see the comparison of KF and KS of both LLT and BSM in Fig. 11.

To make the visual assessment of the Kalman smoother clearer, we decompose it into its respective components, which can be seen in Fig. 12.

Bearing in mind the practical point of view, we construct the Kalman predictor (KP) for the observed Kalman smoother data to depict the future development of the data. The prognosis is always an important and useful part especially for practical implementations. Therefore, we decide to take 10% of the empirical data as a

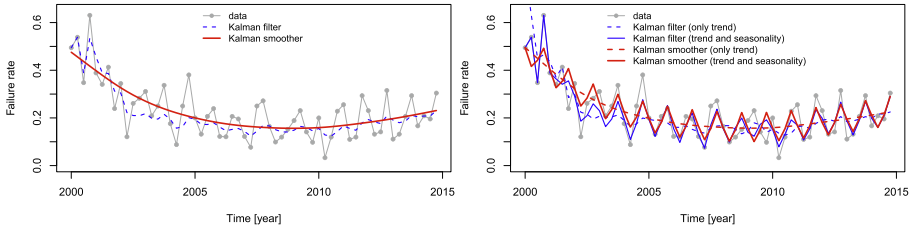

Fig. 11. Water distribution network LLT model (left) and BSM model with seasonal effects (right): Comparison of Kalman filter and smoother for empirical data – the failure rate.

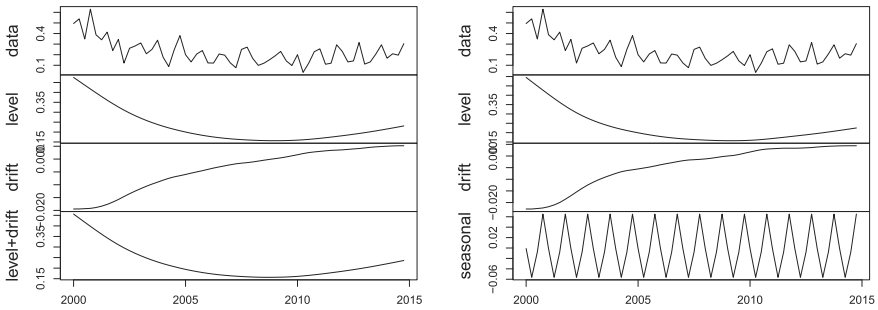

Fig. 12. Water distribution network LLT (left) and BSM (right) model: Kalman smoother decomposition for observed series – failure rate.

number of forward prediction steps. Fig. 13 shows the predicted trend as well as the assumed variability based on a normal distribution.

As the form of empirical data is very specific and typical dynamic models of time series (such as LLT and BSM) are not sufficient, we have to create new forms of time series models such as GDLM-G and GDLM-P. Therefore, the next step is to introduce the results achieved with the models of our major concern, i.e. GDLM-G and mainly GDLM-P mentioned in Section 4.1.1. The expressions (15)–(19) were used to achieve the results.

Similarly, as in previous paragraphs, we estimate the variability of the observed data of the GDLM-P. Since a Poisson model variability relates to the estimated mean value (a filter, a smoother or a predictor) and time – which is the typical characteristic of the Poisson distribution, it is in this case a function (variability series) and not one number. Again we construct the Kalman filter (KF) and Kalman smoother (KS) using the GDLM-P of the observed series – our empirical data converted into the failure rate, which is depicted in Fig. 14. Numerical results of KF and KS are summarised in Table 2 for detailed presentation.

We add the decomposition GDLM-P KF and GDLM-P KS in Fig. 15.

Also in this case we perform the Shapiro-Wilk test to assess normality. The resulting statistics value for GDLM-P: $W = 0.97418$ and its p-value = 0.2817 does not lead to rejection the fact that observed data may come from a normal distribution. Other tests including tests on normality of residuals in the form of a Q-Q plot, ACF test and Ljung-Box statistic were performed. The outcomes can be seen graphically in Fig. 16.

For practical reasons, it is convenient to be able to predict the behaviour of the observed system with a view to identifying failure occurrence. This is really helpful in field operations of a studied system. Therefore, we generate the Kalman predictor of the observed series and the GDLM-P model. Again, we focus on approximately 10% prediction of the original series. Outcomes in graphical form can be seen in Fig. 17. The detailed numerical outcomes are summarised in Table 2 for better comparison and understanding.

For the LLT, BSM and GDLM-P models, it is also important to predict the parameters of variability $\widehat{\sigma}^2_\epsilon$ – standard deviations of the empirical times series error (3), (6), (15), the error dispersion $\widehat{\sigma}^2_\eta$ at the latent level (4), (7), (17), the error dispersion $\widehat{\sigma}^2_\xi$ in latent drift (5), (8) and the error dispersion $\widehat{\sigma}^2_\omega$ in latent seasonal (9), (19).

For LLT we get:

$$\widehat{\sigma}^2_\epsilon = 6.176513 \cdot 10^{-3}, \quad \widehat{\sigma}_\epsilon = 7.859079 \cdot 10^{-2},$$
$$\widehat{\sigma}^2_\eta = 7.471000 \cdot 10^{-11}, \quad \widehat{\sigma}_\eta = 8.643482 \cdot 10^{-6},$$
$$\widehat{\sigma}^2_\xi = 5.639518 \cdot 10^{-6}, \quad \widehat{\sigma}_\xi = 2.374767 \cdot 10^{-3}.$$

For BSM we get:

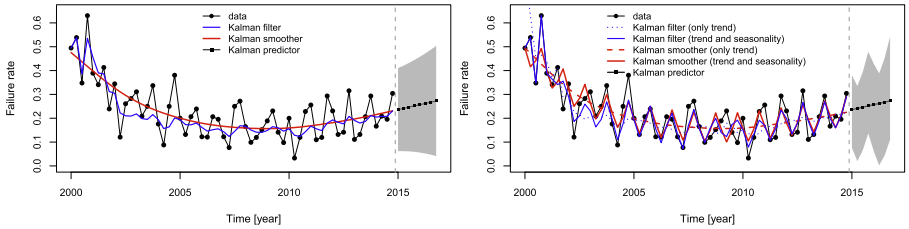

Fig. 13. Water distribution network LLT model (left) and BSM model with seasonal effects (right): Kalman filter, smoother with 95% confidence intervals and predictor (several points forward) where estimated variability is in grey for empirical data – the failure rate.

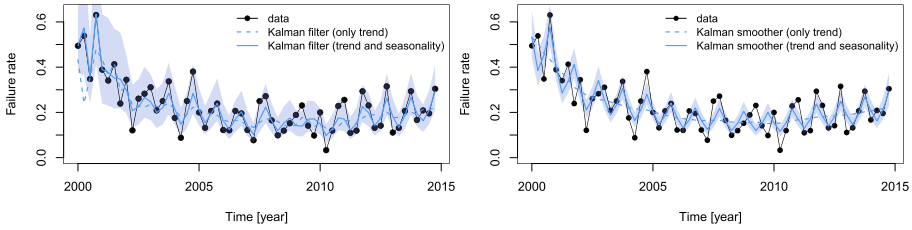

Fig. 14. Water distribution network observed series – failure rate: GDLM-P KF model (left) and GDLM-P KS model (right) – both with seasonal effects and with 95% confidence thresholds.

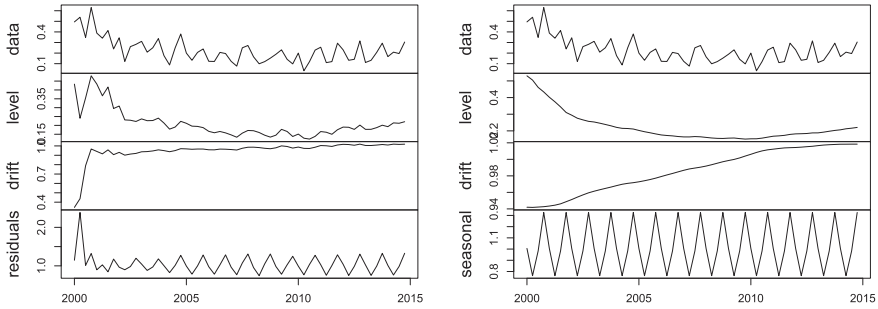

Fig. 15. Water distribution network observed series – failure rate: decomposed GDLM-P KF model (left) and decomposed GDLM-P KS model (right).

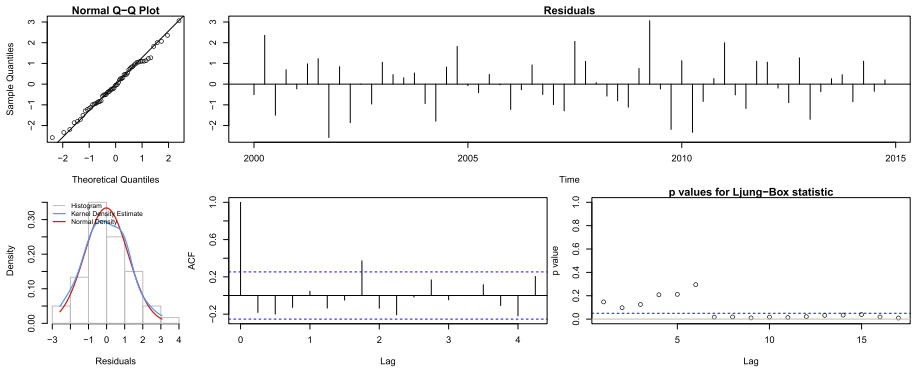

Fig. 16. Water distribution network observed series – failure rate – GDLM-P: Normality Q-Q plot test of residuals (left) and diagnostic plots of residuals, ACF and Ljung-Box (right).

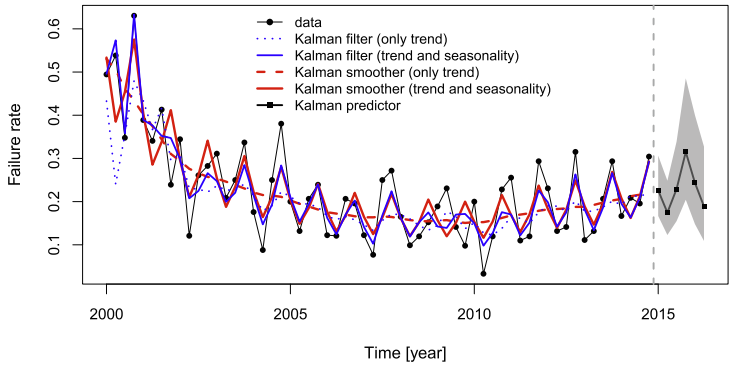

Fig. 17. Water distribution network observed series – failure rate – GDLM-P: Kalman predictor also with seasonal effects and 95% confidence interval.

Table 3
The expected number of failures according to applied models in respective quarters.

Quarter	LLT		BSM		GDLM-P	
	Failure rate	F_t	Failure rate	F_t	Failure rate Failure rate-trend	F_t F_t-trend
1/2015	0.2362	21	0.2290	21	0.2261 0.2248	20 20
2/2015	0.2414	22	0.1774	16	0.1746 0.2289	15 21
3/2015	0.2466	23	0.2303	21	0.2291 0.2331	21 21
4/2015	0.2518	23	0.3078	28	0.3149 0.2374	28 21
1/2016	0.2570	23	0.2466	22	0.2432 0.2418	21 21
2/2016	0.2622	24	0.1950	18	0.1878 0.2462	17 22

$$\widehat{\sigma}^2_{\epsilon} = 4.325\,816 \cdot 10^{-3}, \quad \widehat{\sigma}_{\epsilon} = 6.577\,093 \cdot 10^{-2},$$
$$\widehat{\sigma}^2_{\eta} = 2.266\,026 \cdot 10^{-9}, \quad \widehat{\sigma}_{\eta} = 4.760\,280 \cdot 10^{-5},$$
$$\widehat{\sigma}^2_{\xi} = 7.079\,431 \cdot 10^{-6}, \quad \widehat{\sigma}_{\xi} = 2.660\,720 \cdot 10^{-3},$$
$$\widehat{\sigma}^2_{\omega} = 2.506\,000 \cdot 10^{-11}, \quad \widehat{\sigma}_{\omega} = 5.006\,023 \cdot 10^{-6}.$$

For GDLM-P we get:

$$\widehat{\sigma}^2_{\epsilon} = 0.002\,939\,463, \quad \widehat{\sigma}_{\epsilon} = 0.054\,216\,813,$$
$$\widehat{\sigma}^2_{\eta} = 0.000\,063\,158, \quad \widehat{\sigma}_{\eta} = 0.007\,947\,201,$$
$$\widehat{\sigma}^2_{\xi} = 0.000\,000\,378, \quad \widehat{\sigma}_{\xi} = 0.000\,614\,572.$$

The achieved results clearly show that there are certain differences between the variability and standard deviations of LLT, BSM and the newly developed GDLM-P models. However, this does not mean that significant deviations occur in the numerical values of the failure amount results as stated below.

Next, in Table 2, we introduce the parameter values of the calculated time series for the respective models. Several numerical values calculated for the Kalman filter, smoother and predictor KF, KS and KP are displayed in following table.

Based on the time series results in single models, we can calculate the expected number of failures during single quarters. The results are shown in Table 3.

The results achieved are interpreted as follows. It is remarkable that there are small deviations between the used models and their practical interpretation. We could say that during the 1st quarter of 2015, we can expect a water pipe failure every four days on average. During the 2nd quarter of 2015, we could expect a failure every 4.5 to 5.5 days, etc. This information is very important for

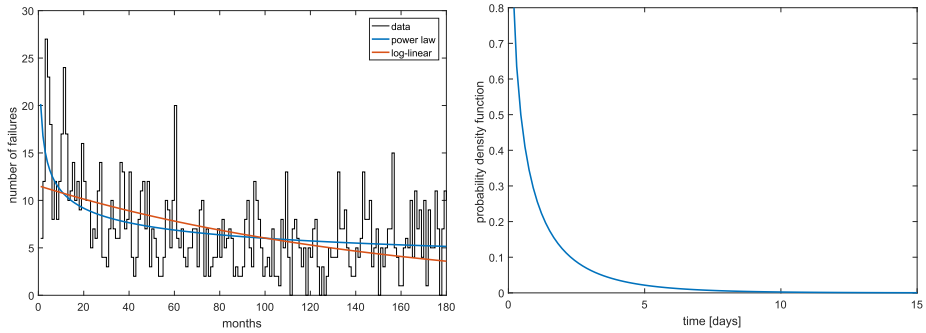

Fig. 18. Left: Models of failure intensity (grey) – power law model (blue) and log-linear model (orange); Right: Weibull probability density function corresponding to the power law model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

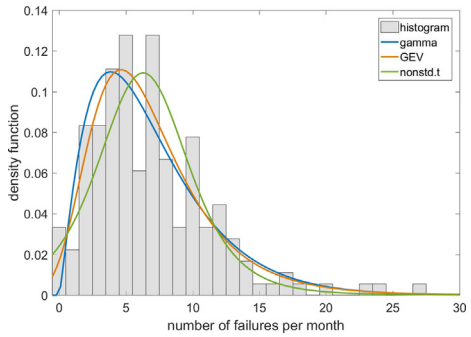

Fig. 19. Histogram of numbers of water distribution network failures per month with fitted non-rejected parametric distributions: the gamma (blue), the generalised extreme value (GEV, orange) and the generalised Student distribution (gen.t, green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the managers who plan and arrange maintenance and water supply in case of breakdown. The crisis manager does not know exactly where the failures will occur, but they can estimate how many failures could occur and how many inhabitants may be affected. Based on this information, asset and manpower sources can be optimised, which of course saves time and money.

The course of the power law model and the log-linear model (see Section 4.1.2) compared with the course of the recorded data and the calculated failure frequency are shown in Fig. 18. This approach was taken in order to create a possible probability density function which describes and may represent the studied system behaviour. Estimated parameters of single models used for further estimation of the density of the observed system time to failure are

- the power law model: $\widehat{\alpha} = 0.6627$ and $\widehat{\beta} = -0.2629$, i.e.

  $$\widehat{\lambda}(t) = 0.6627t^{-0.2629},$$

- the log-linear model: $\widehat{\alpha} = -0.9699$ and $\widehat{\beta} = -0.0065$, i.e.

  $$\widehat{\lambda}(t) = \exp(-0.9699 - 0.0065t).$$

Corresponding probability distribution to the power model is the Weibull distribution with parameters $\widehat{k} = 0.7371, \widehat{b} = 0.8990$. The Weibull density function acquired from our power-law calculation is graphically displayed in Fig. 18. This distribution is used to estimate time to failure of the analysed mains and we strongly believe it may be very realistic. The Gompertz distribution is connected with the log-linear model, however the parameters are negative $\widehat{\beta} = -0.0065, \widehat{\eta} = -58.5168$. The obtained function is not a probability density function – the distribution does not have a proper distribution function [24,18] and cannot be used as a probability model.

As we have a relatively significant data set and mathematical tools available, we also construct the probability density function describing the distribution of the number of failures per month. The histogram of the numbers of failures is in Fig. 19. We compared the empirical distribution to some relevant parametric distributions typically used in this praxis of WDN. Using the Kolmogorov-Smirnov test, we have not rejected the hypotheses for the gamma distribution, the generalised extreme value distribution and the generalised Student distribution.

We have also estimated the parameters of non-rejected distributions which again may be practically very useful. For the gamma distribution with density function $f(t) = \left[b^a\Gamma(a)\right]^{-1}t^{a-1}\exp(-t/b), x > 0$, the estimated parameters – using the moment method – are: $\widehat{a} = 2.2860$ and $\widehat{b} = 3.0062$. This estimation allows us to provide neither confidence intervals (CI) nor prediction intervals (PI).

For the generalised extreme value distribution with density function $f(t) = s^{-1}\exp\left(-(1+k(t-m)/s)^{-1/k}\right)(1+k(t-m)/s)^{-1-1/k}$ for $(1+k(t-m)/s) > 0$, the estimated parameters – using the maximum likelihood estimates (MLE) – are: $\widehat{k} = 0.0475$ with $\mathrm{CI} = (-0.0611, 0.1561)$, $\widehat{s} = 3.3228$ with $\mathrm{CI} = (2.9424, 3.7523)$ and $\widehat{m} = 4.7914$ with $\mathrm{CI} = (4.2444, 5.3385)$.

For the generalised Student distribution (also known as the t location-scale distribution) with density function $f(t) = \Gamma((v+1)/2)/\left[s(v\pi)^{1/2}\Gamma(v/2)\right]\cdot\left(\left(v+(t-m)^2/s^2\right)/v\right)^{-(v+1)/2}, \quad t\in\mathbb{R}$, the estimated parameters are: $\widehat{m} = 6.2930$ with $\mathrm{CI} = (5.6665, 6.9195)$, $\widehat{s} = 3.4677$ with $\mathrm{CI} = (2.9239, 4.1126)$ and $\widehat{v} = 4.8033$ with $\mathrm{CI} = (2.4843, 9.2871)$.

6. Conclusion

In this article we applied a special form of time series modelling using a generalised Kalman recursor. Although there are known forms of Kalman filters (such as LLM, BSM, etc.), we had to create new forms of the generalised Kalman approach. In our case, we focused on generalised dynamic models based on Gauss and Poisson distributions. This decision was based on the empirical data form and the studied system. The main intention of our article is to introduce a possible and promising application of analytical and modelling principles in the specific forms of field data records. Especially for a system – such as WDN – which is considered as an element of critical infrastructure but where at the same time the operational data available are very poor and lack more detailed information. Therefore, we created and applied very modern

approaches of dynamic linear models based on Kalman recursion. The form of the field data and their properties directed our effort to the non-homogeneous Poisson field of dynamic linear models of time series, which is not a very common approach to studying time series. Therefore, based on the above, we hope that this article will be of use for other professionals and practitioners in field operations of WDN.

In our future work, we plan to investigate emerging break points in data structures. Our intention is to also apply other advanced tools of dynamic linear models such as advanced non-parametric predictions. For such an idea, we can already find quite a few inspirational examples of system diagnostics including non-homogeneous Poisson data and some interesting principles of WDN availability assessment, see [26,25,21,1,49,20,31].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This article was prepared with the partial support of development intentions UO-FVT-K202-DZRO 'MOBAUT' and UO-FVL-K102-DZRO 'AERO'.

References

[1] E. Abrahamsen, F. Asche, M. Milazzo, An evaluation of the effects on safety of using safety standards in major hazard industries, Saf. Sci. 59 (2013) 173–178.
[2] A. Arsenio, P. Dheenathayalan, R. Hanssen, J. Vreeburg, L. Rietveld, Pipe failure predictions in drinking water systems using satellite observations, Struct. Infrastruct. Eng. 11 (2015) 1102–1111, https://doi.org/10.1080/15732479.2014.938660.
[3] Z. Asadi, R. Melchers, Extreme value statistics for pitting corrosion of old underground cast iron pipes, Reliab. Eng. Syst. Saf. 162 (2017) 64–71, https://doi.org/10.1016/j.ress.2017.01.019.
[4] R. Barrela, C. Amado, D. Loureiro, A. Mamade, Data reconstruction of flow time series in water distribution systems – a new method that accommodates multiple seasonality, J. Hydroinf. 19 (2017) 238–250, https://doi.org/10.2166/hydro.2016.192.
[5] P. Brockwell, R. Davis, Time Series: Theory and Methods, Springer, New York, 2006.
[6] W. Deng, G. Wang, A novel water quality data analysis framework based on time-series data mining, J. Environ. Manage. 196 (2017) 365–375, https://doi.org/10.1016/j.jenvman.2017.03.024.
[7] M. Firat, M. Turan, M. Yurdusev, Comparative analysis of fuzzy inference systems for water consumption time series prediction, J. Hydrol. 374 (2009) 235–241, https://doi.org/10.1016/j.jhydrol.2009.06.013.
[8] R. Francis, S. Guikema, L. Henneman, Bayesian belief networks for predicting drinking water distribution system pipe breaks, Reliab. Eng. Syst. Saf. 130 (2014) 1–11, https://doi.org/10.1016/j.ress.2014.04.024.
[9] M.L. Gámiz, K. Kulasekera, N. Limnios, B. Lindqvist, Applied Nonparametric Statistics in Reliability, Springer, London, 2011.
[10] B. García-Mora, A. Debón, C. Santamarí, A. Carrión, Modelling the failure risk for water supply networks with interval-censored data, Reliab. Eng. Syst. Saf. 144 (2015) 311–318, https://doi.org/10.1016/j.ress.2015.08.003.
[11] Y.L. Gat, P. Eisenbeis, Using maintenance records to forecast failures in water networks, Urban. Water. J. 3 (2000) 173–181.
[12] Y.L. Gat, R. Fenner, Kalman filtering of hydraulic measurements for burst detection in water distribution systems, J. Pipeline Syst. Eng. 2 (2011) 14–22, https://doi.org/10.1061/(ASCE)PS.1949-1204.0000070.
[13] B. Harding, T. Walski, Long time-series stimulation of water quality in distribution systems, J. Water. Res. Plan. Manage. 126 (2000) 199–209, https://doi.org/10.1061/(ASCE)0733-9496(2000) 26:4(199).
[14] S. Henderson, Estimation for nonhomogeneous poisson processes from aggregated data, Oper. Res. 31 (2003) 375–382.
[15] R. Herz, Exploring rehabilitation needs and strategies for water distribution networks, J. Water. Supply Res. Technol. 47 (1998) 275–283.
[16] H. Hwang, K. Lansey, D. Quintanar, Resilience-based failure mode effects and criticality analysis for regional water supply system, J. Hydroinf. 17 (2015) 193–210, https://doi.org/10.2166/hydro.2014.111.
[17] D. Jung, K. Lansey, Burst detection in water distribution system using the extended kalman filter, Proc. Eng. 70 (2014) 902–906, https://doi.org/10.1016/j.proeng.2014.02.100.
[18] N. Kolev, Characterizations of the class of bivariate gompertz distributions, J. Multivariate Anal. 148 (2016) 173–179, https://doi.org/10.1016/j.jmva.2016.03.004.
[19] M. Kutylowska, H. Hotlos, Failure analysis of water supply system in the polish city of glogow, Eng. Fail. Anal. 41 (2014) 23–29, https://doi.org/10.1016/j.engfailanal.2013.07.019.
[20] L. Letting, Y. Hamam, A. Abu-Mahfouz, Estimation of water demand in water distribution systems using particle swarm optimization, Water 9 (2017) 593, https://doi.org/10.3390/w9080593.
[21] R. Lisi, G. Consolo, G. Maschio, M. Milazzo, Estimation of the impact probability in domino effects due to the projection of fragments, Process. Saf. Environ. 93 (2015) 99–110, https://doi.org/10.1016/j.psep.2014.05.003.
[22] K. Manandhar, X. Cao, F. Hu, Attack detection in water supply systems using kalman filter estimator, in: 35th IEEE Sarnoff Symposium, 2012, pp. 1–6, https://doi.org/10.1109/SARNOF.2012.6222737.
[23] A. Manuco, M. Compare, A. Salo, E. Zio, T. Laakso, Risk-based optimization of pipe inspections in large underground networks with imprecise information, Reliab. Eng. Syst. Saf. 152 (2016) 228–238, https://doi.org/10.1016/j.ress.2016.03.011.
[24] A. Marshall, I. Olkin, Life Distributions: Structure of Nonparametric, Semiparametric, and Parametric Families, Springer, New York, 2007.
[25] D. Mazurkiewicz, Tests of extendability and strength of adhesive-sealed joints in the context of developing a computer system for monitoring the condition of belt joints during conveyor operation, Eksploat. Niezawodn. 3 (2010) 34–39.
[26] D. Mazurkiewicz, Computer-aided maintenance and reliability management systems for conveyor belts, Eksploat. Niezawodn. 16 (2014) 377–382.
[27] W. Meeker, A. Luis, Statistical Methods for Reliability Data, Wiley, New York, 1998.
[28] S. Mounce, J. Gaffney, S. Boult, J. Boxall, Automated data-driven approaches to evaluating and interpreting water quality time series data from water distribution systems, J. Water. Res. Plan. Manage. 141 (2015), https://doi.org/10.1061/(ASCE)WR.1943-5452.0000533.
[29] S. Mounce, R. Mounce, J. Boxall, Novelty detection for time series data analysis in water distribution systems using support vector machines, J. Hydroinf. 13 (2011) 672–686, https://doi.org/10.2166/hydro.2010.144.
[30] S. Mounce, R. Mounce, T. Jackson, J. Austin, J. Boxall, Pattern matching and associative artificial neural networks for water distribution system time series data analysis, J. Hydroinf. 16 (2014) 617–632, https://doi.org/10.2166/hydro.2013.057.
[31] A.D. Nardo, M.D. Natale, C. Giudicianni, G. Santonastaso, V. Tzatchkov, J. Varela, Economic and energy criteria for district meter areas design of water distribution networks, Water 9 (2017) 463, https://doi.org/10.3390/w9080593.
[32] NIST/SEMATECH, 2012. e-handbook of statistical methods. on-line.http://www.itl.nist.gov/div898/handbook/ (accessed 24.10.2017). .
[33] I. Okeya, Z. Kapelan, C. Hutton, D. Naga, Online burst detection in a water distribution system using the kalman filter and hydraulic modelling, Proc. Eng. 89 (2014) 418–427, https://doi.org/10.1016/j.proeng.2014.11.207.
[34] C. Ossai, B. Boswell, I. Davis, Stochastic modelling of perfect inspection and repair actions for leak-failure prone internal corroded pipelines, Eng. Fail. Anal. 60 (2016) 40–56, https://doi.org/10.1016/j.engfailanal.2015.11.030.
[35] S. Park, H. Jun, B. Kim, G. Im, Modeling of water main failure reates using the log-linear rocof and the power law process, Water. Resour. Manage. 22 (2008) 1311–1324, https://doi.org/10.1007/s11269-007-9227-3.
[36] L. Perelman, J. Arad, M. Housh, A. Ostfeld, Event detection in water distribution systems from multivariate water quality time series, Environ. Sci. Technol. 46 (2012) 8212–8219, https://doi.org/10.1021/es3014024.
[37] I. Piegdon, B. Tchorzewska-Cieslak, D. Szpak, The use of geographical information system in the analysis of risk of failure of water supply network, in: M. Pawlowska, L. Pawlowski (Eds.), Environmental Engineering V, Proceedings of the 5th National Congress of Environmental Engineering, CRC/Balkema, Leiden, 2017, pp. 7–14, https://doi.org/10.1201/9781315281971-3.
[38] K. Pietrucha-Urbanik, Failure prediction in water supply system – current issues, in: W. Zamojski, J. Mazurkiewicz, J. Sugier, T. Walkowiak, J. Kacprzyk (Eds.), Theory and Engineering of Complex Systems and Dependability. Advances in Intelligent Systems and Computing, Springer, Cham, 2015, pp. 351–358.
[39] K. Pietrucha-Urbanik, A. Studzinski, Selected issues of costs and failure of pipes in an exemplary water supply system, Rocz. Ochr. Sr. 18 (2016) 616–627.
[40] K. Pietrucha-Urbanik, A. Studzinski, Case study of failure simulation of pipelines conducted in chosen water supply system, Eksploat. Niezawodn. 19 (2017) 317–323, https://doi.org/10.17531/ein.2017.3.1.
[41] O. Pozos-Estrada, A. Sanchez-Huerta, J. Brena-Naranjo, A. Pedrozo-Acuna, Failure analysis of a water supply pumping pipeline system, Water 8 (2016), https://doi.org/10.3390/w8090395.
[42] R Core Team, 2015. R: A Language and Environment for Statistical Computing. https://www.R-project.org/. .
[43] M. Rausand, A. Hoyland, System Reliability Theory: Models, Statistical Methods, and Applications, Wiley, New Jersey, 2004.
[44] E. Renaud, Y.L. Gat, M. Poulton, Using a break prediction model for drinking water networks asset management: From research to practice, in: International water association, 4th Conference on Strategic Asset Management, 2011, p. 10.
[45] U. Shamir, C. Howard, An analytic approach to scheduling pipe replacement, J. Am. Water. Resour. Assoc. 5 (1979) 248–258.

[46] G. Williams, G. Kuczera, Framework for forensic investigation of associations between operational states and pipe failures in water distribution systems, J. Water. Res. Plan. Manage. 142 (2016), https://doi.org/10.1061/(ASCE)WR.1943-5452.0000623.

[47] D. Winkler, M. Haltmeier, M. Kleidorfer, W. Rauch, F. Tscheikner-Gratl, Pipe failurea modelling for water distribution networks using boosted decision trees, Struct. Infrastruct. Eng. 14 (2018) 1402–1411, https://doi.org/10.1080/15732479.2018.1443145.

[48] B. Wols, P. van Thienen, Impact of climate on pipe failure: predictions of failures for drinking water distribution systems, Eur. J. Transp. Infrastruct. 16 (2016) 240–253.

[49] J. Zischg, M. Mair, W. Rauch, R. Sitzenfrei, Enabling efficient and sustainable transitions of water distribution systems under network structure uncertainty, Water 9 (2017) 715, https://doi.org/10.3390/w9090715.