

Flow forecasting for leakage burst prediction in water distribution systems using long short-term memory neural networks and Kalman filtering

Lauren McMillan^{a,*}, Jawad Fayaz^{a,b}, Liz Varga^a

^a Infrastructure Systems Institute, Department of Civil, Environmental and Geomatic Engineering, UCL, London, UK

^b School of Computing, Engineering & Digital Technologies, Teesside University, Teesside, UK

ARTICLE INFO

Keywords:

Water flow forecasting
Leakage prediction
Long short-term memory
Recurrent neural networks
Kalman filter
Self-healing systems

ABSTRACT

Reducing pipe leakage is one of the top priorities for water companies, with many investing in higher quality sensor coverage to improve flow forecasting and detection of leaks. Most research on this topic is focused on leakage detection through the analysis of sensor data from district metered areas (DMAs), aiming to identify bursts after their occurrence. This study is a step towards the development of 'self-healing' water infrastructure systems. In particular, machine learning and deep learning-based algorithms are applied to forecasting the anomalous water flow experienced during bursts (new leakage) in DMAs at various temporal scales, thereby aiding in the health monitoring of water distribution systems. This study uses a dataset of over 2,000 DMAs in North Yorkshire, UK, containing flow time series recorded at 15-minute intervals for a period of one year. Firstly, the method of isolation forests is used to identify anomalies in the dataset, which are cross referenced with entries in the water mains repair log, indicating the occurrence of bursts. Going beyond leakage detection, this research proposes a hybrid deep learning framework named FLUIDS (Forecasting Leakage and Usual flow Intelligently in water Distribution Systems). A recurrent neural network (RNN) is used for mean flow forecasting, which is then combined with forecasted residuals obtained through real-time Kalman filtering. While providing expected day-to-day flow demands, this framework also aims to issue sufficient early warning for any upcoming anomalous flow or possible leakages. For a given forecast period, the FLUIDS framework can be used to compute the probability of flow exceeding a pre-defined threshold, thus allowing decision-making for any necessary interventions. This can inform targeted repair strategies that best utilize resources to minimize leakages and disruptions. The FLUIDS framework is statistically assessed and compared against the state-of-practice minimum night flow (MNF) methodology. Based on the statistical analyses, it is concluded that the proposed framework performs well on the unobserved test dataset for both regular and leakage water flows.

1. Introduction

With rising population levels putting pressure on water supplies, efficient distribution of this increasingly scarce resource is crucial if the needs of consumers are to be met. For 2018–2019, in England and Wales, an average of 3170 million liters (21% of the water entering the public supply) was lost daily as leakage. This equates to wastage of 53 liters per person per day (PR19 final determinations: Securing cost efficiency technical appendix 2019). Such high levels of leakage result not only in higher energy use (and associated carbon emissions) in water treatment (Spedaletti et al., 2022), but also undermines efforts to reduce consumption on the demand side, as consumer confidence in their water supplier is reduced. This is compounded by the fact that, despite efforts

by the water companies to improve leakage management, a significant proportion of bursts in distribution systems are reported by consumers rather than detected by the companies themselves. This puts increased pressure on supply-based solutions for rising water consumption, such as the construction or expansion of water storage and treatment infrastructure, many of which are carbon-intensive and will set the water industry back in its goal of achieving net-zero by 2030 (Net Zero 2030 Routemap, 2020). Instead, a sustainable approach to reduce supply-based wastage via leakage is needed (Ávila, Sánchez-Romero, López-Jiménez & Pérez-Sánchez, 2022), which will in-turn impact consumer behavior by building trust. Hence, more proactive and reliable approaches are required for leakage management in water distribution systems (Proactive approach to leaks required to meet tough Ofwat

* Corresponding author at: Room 119, Chadwick Building, Gower St, London, WC1E 6AE, UK
E-mail address: Lauren.mcmillan.19@ucl.ac.uk (L. McMillan).

<https://doi.org/10.1016/j.scs.2023.104934>

Received 14 March 2023; Received in revised form 14 June 2023; Accepted 10 September 2023

Available online 12 September 2023

2210-6707/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

targets, 2020; Zanfei, Menapace & Righetti, 2023).

Hence, a priority for Ofwat, the economic regulator of the water industry in England and Wales, is to reduce leakage across water distribution networks, ensuring a more significant proportion of water is available to meet the population's needs and improving consumer confidence in the water supply. The standard practice for water utility companies in the UK is to divide the water distribution network into district meter areas (DMAs). DMAs represent isolated water network areas, typically serving up to 2000 households, where the flow is measured at the inlet and outlet. Leakage management in the UK is usually performed at the DMA level (Morrisson, 2004).

In the field of leakage management of water flow distribution pipes, leakage detection has been a critical research subject (Romano, Kapelan & Savić, 2014; Chan, Chin & Zhong, 2018; Puust, Kapelan, Savić & Koppel, 2010; Aksela, Aksela & Vahala, 2009; Mounce, Boxall & Machell, 2010). Sensors at the DMA inlet and outlet record flow data at regular time-intervals, monitoring the flow behavior. This data can be fed into leakage detection models that seek to identify leaks from changes in the flow profile over a set time period. Many of these models are trained using examples of normal flow data and flow during leakage bursts. The burst examples are typically obtained by matching the timestamps of abnormal flow patterns to pipe repair records or reports of visible leakage from consumers to water companies. Alternatively, the data can be simulated through a hydrant flush event designed to mimic a leakage burst (Birek, Petrovic & Boylan, 2014). Some studies do not use data from real water distribution networks and instead extract pressure data from simulation software-based network models (Leu and Bui, 2016).

The most common methods for identifying leaks utilize the concept of minimum night flow (MNF) (García, Cabrera, and Cabrera, 2012). This technique recognizes that water usage during nighttime hours is less variable compared to the daytime. Hence, the average nightly minimum over a specified window can be used as a baseline for comparison with new flow data, with a significant variation (relative to a pre-defined threshold) indicating a leakage (Malithong, Gulphanich, and Suesut, 2005; Mounce, Boxall, and Machell, 2007). However, these techniques are not highly reliable as MNF methodologies have to deal with several uncertainties. Accurate use of MNF relies upon having sufficient knowledge to estimate several parameters, including active night users, leakage exponent (which varies with system pressure), and the hour-to-day factor (Amoatey, Obiri-Yeboah, and Akosah-Kusi, 2021). Reliable estimation of these parameters typically requires pressure data in addition to the flow data. Selection of the best time window for MNF is an additional consideration. It has been shown that minimum error does not correspond with the selected night flow window but with the hour in which average demand applies (García, Cabrera, and Cabrera, 2012). While it is often the responsibility of trained operators to identify leakage from MNF, a significant proportion of leaks are reported to the water companies by their customers (Mounce, Boxall, and Machell, 2007).

More recent work conducted in the leakage management and detection domain has explored the potential of using machine learning and deep learning tools. These include artificial neural networks (ANNs) (Mounce and Machell, 2006; Aksela, Aksela, and Vahala, 2009; Romano, Kapelan, and Savić, 2014; Zanfei et al., 2022), support vector machines (Geberemariam, Juran, and Shahrour, 2014; Kang et al., 2018; Gao, Yang, and Hu, 2010), Kalman filters (KFs) (Ye and Fenner, 2011; Jung and Lansey, 2015), and wavelet analysis (Romano, Kapelan, and Savić, 2014). With sufficient quality and quantity of training data, these methods have demonstrated strong performance in leakage identification (Mounce and Machell, 2006). Recent work has also seen machine learning-based leakage detection built into larger models that act as digital twins of water systems (Wu et al., 2023), recognizing that intelligent management of water systems is a key component of the drive towards sustainable smart cities (Oberascher, Rauch, and Sitzenfrie, 2022). Apart from leakage detection, another critical domain in the field

of leakage management is the forecasting or prediction of leakage. Unlike leakage detection, which is concerned with the identification of bursts from flow data after they have occurred, leakage prediction/forecasting aims to anticipate anomalous flow before it occurs, thereby enabling early warning of potential leakage within a given forecasting period. This allows preventative maintenance to be scheduled, which can act to repair pipes before any water is lost as leakage. While the field of leakage detection has observed several dedicated studies, leakage prediction/forecasting has received significantly less attention from the research community due to its complexity. Leakage forecasting at a regional level has been conducted over various time periods ranging from weeks to a year. For example, Birek et al. (2014) utilize an evolving fuzzy algorithm on historic leakage levels and repair data across nine regions consisting of aggregated DMA areas to forecast the future rates of monthly leakage (Birek, Petrovic, and Boylan, 2014). Studies on leakage forecasting at the individual pipe level have analyzed pipe properties, such as diameter, age, and material, as well as other factors, including soil type, ground movement, and traffic loading, to assess their impact on leakage likelihood (Leu and Bui, 2016; Jing and Zhi-Hong, 2012; Barton, Farewell, Hallett, and Acland, 2019).

In recent years, the forecasting of water flow data at a DMA level has gained attention (Hutton and Kapelan, 2015; Mounce, 2013). Typically, these studies have primarily focused on predicting regular water demand rather than specifically addressing leakage prediction (Pandey, Bokde, Dongre, and Gupta, 2021; Kavya, Mathew, Shekar, and P, 2023). Water demand forecasting aims to estimate expected water usage, and thus it mainly focuses on forecasting typical non-leakage flow. On the other hand, leakage prediction requires forecasting anomalous flow, which can indicate potential leakage incidents (Geelen, Yntema, Molenaar, and Keesman, 2021). While water demand forecasting is valuable for resource planning, leakage prediction can significantly improve asset repair strategies and enhance system efficiency by reducing water loss (McMillan and Varga, 2022).

However, there have been some studies that attempt to detect leakage by forecasting expected non-leakage flow levels using Bayesian forecasting methods (Hutton and Kapelan, 2015; Geelen, Yntema, Molenaar, and Keesman, 2021). These studies compare the forecasted flow levels with incoming flow data, and a significant difference is considered indicative of leakage (Hutton and Kapelan, 2015; Geelen, Yntema, Molenaar, and Keesman, 2021). It has been suggested that machine learning techniques, particularly ANNs, have the potential to outperform baseline methods in forecasting flow data at the DMA level (Mounce, 2013). Recent research has indicated that long short-term memory (LSTM)-based neural networks offer superior performance in demand forecasting, surpassing other time series forecasting models for predicting typical short-term water demand in a single DMA case study (Kavya, Mathew, Shekar, and P, 2023). Furthermore, recent work on the forecasting of time series climate data suggests that using an information theory based loss function (Sayeed et al., 2021) can improve performance over the traditional loss functions seen to date in water demand forecasting (Kavya, Mathew, Shekar, and P, 2023).

Hybrid forecasting methods, which combine various forecasting techniques with error (residual) forecasting modeling, have shown high levels of accuracy and the ability to forecast time series with different characteristics (de Oliveira, Silva, and de Mattos Neto, 2022). Applying these methods to the water sector has clear benefits, and some studies have already applied hybrid methods to typical water demand forecasting (Bata, Carriveau, and Ting, 2020; Pandey et al., 2021). The effectiveness of residual forecasting in improving time series forecasting of water demand has been demonstrated at both the regional (Chen, Long, Bai, and Zhang, 2019) and DMA levels (Brentan et al., 2018), with the KF being the preferred method for residual forecasting in these studies. However, it should be noted that these studies have focused solely on demand forecasting and do not address the forecasting of leakage flow (Chen, Long, Bai, and Zhang, 2019; Brentan et al., 2018). Furthermore, the LSTM-based forecasting method (Kavya, Mathew,

Shekar, and P, 2023) and the residual forecasting approach with KF (Brentan et al., 2018) have so far been applied only at the scale of a single DMA and have not been combined or applied to large datasets, such as the thousands of DMAs managed by each water company (Kavya, Mathew, Shekar, and P, 2023; Brentan et al., 2018). Therefore, there is a need to explore the potential of combining these methods and apply them to a large dataset to harness their benefits on a broader scale (such as the thousands of DMAs managed by each water company).

Although there is clearly significant potential in this area, there has not yet been a study that uses real sensor data and sophisticated data-driven machine-learning and deep-learning techniques to forecast, at any geographic level, the anomalous flow that indicates bursts/leakages. By forecasting anomalous flow, rather than contrasting a forecast of expected flow with incoming data, early warning can be provided for leakage, facilitating faster repair. An accurate forecast of leakage flow can also provide an estimate of expected water loss, which can inform the prioritization of repair jobs. This places leakage forecasting within a bigger system of self-healing leakage management, which considers the processes of anticipation, detection, and repair (McMillan and Varga, 2022).

Hence, the study presented herein proposes a hybrid deep learning framework named FLUIDS (Forecasting Leakage and Usual flow Intelligently in Distribution Systems) for leakage forecasting at a DMA level. FLUIDS is based on LSTM recurrent neural networks (RNNs) and KF which utilizes the recorded flow data up to the current time step (t) to forecast flow for n future time steps ($t + n$). The framework is trained and tested using an extensive database of ~ 2000 DMAs from North Yorkshire, UK, containing 15-minute interval flow data for a year. The trained framework is statistically validated for high goodness-of-fit and forecasting power. Furthermore, due to the data-driven nature of the proposed FLUIDS framework, it can be efficiently trained for other DMAs using the flow data available to the water utility companies and hence can be effectively utilized for proper water resource management.

2. Dataset

The dataset used in this study is provided by Yorkshire Water, the utility company responsible for water supply and distribution in North Yorkshire, UK. For water distribution management, Yorkshire Water divides this region into over 2000 DMAs. Fig. 1 presents the locations of the ~ 2000 DMAs used in this study.

For each DMA, the net flow data (in liters/second) is available at 15-minute intervals for a year, from April 2016 to April 2017. Yorkshire Water assigns each flow data point a validity code; 'V' for valid, 'I' for invalid, or 'M' for missing, as it is not uncommon for a DMA to have periods of missing or invalid data due to faults in the sensors. However,

these faults are infrequent and represent less than 5% of the DMA flow data provided for this study. Fig. 2a shows the full year of flow data for one exemplar DMA. It can be observed that the magnitude of flow remains broadly consistent throughout the year, with slight seasonal fluctuations as expected. Fig. 2b shows a standard week of valid flow data from the same DMA and the daily fluctuations can be easily seen in the data indicating lower nightly usages. On a 24-hour scale, minimums are seen during the night hours, with peaks occurring during the morning and late afternoon that correspond with a large proportion of the population leaving for and returning from work/school. The example in Fig. 2a also contains very short periods of missing data and several very short 'spikes', where the flow magnitude jumps well above the values typically seen. The provided dataset also contains the repair log of the DMAs with their repair dates. Over 5000 repairs are reported in the repair log provided for this study, covering over 1600 unique DMAs. The dataset provided by Yorkshire Water doesn't assign a reason for the conducted repairs. It is assumed that the majority of repairs are in response to leakage or damage to the pipes. Repairs are typically prompted either by customer leakage reports or identification of unusual flow data by Yorkshire Water operators. While leaks that are customer-reported and visible at surface level are often tackled within a few hours or days, leaks that are not visible may take several weeks to be repaired. This delay between leakage and repair means that a comparison of flow and repair logs alone is insufficient for verifying the leakage instances. Instead, a method is needed for the identification of flow data that corresponds to leakage and the timing of which can then be compared to recorded repairs. In the absence of widespread metering, repair data is the best alternative for the validation of the models.

3. Dataset pre-processing

In an ideal scenario, the FLUIDS framework would undergo training using a dataset comprising confirmed leakages, as well as confirmed non-leakage flow, extracted from a complete dataset devoid of any missing or invalid water flow data. However, this ideal scenario is unrealistic due to the presence of various issues and errors inherent in real-time sensor data collection. Therefore, it becomes necessary to statistically complete the available water flow data and carefully select appropriate examples of water flow data that accurately represent bursts/leakage events and periods of regular/non-leakage flow. This section outlines the pre-processing steps taken to generate the essential inputs for training the proposed framework.

The sensor data obtained from the DMAs often contain missing and invalid flow data points thereby requiring statistical completion of the flow data. In this study, Kalman smoothing is used to complete the dataset. Once the flow data is completed, an anomaly detection algorithm, such as isolation forests, is applied to identify outlier points corresponding to pipe leakage. It should be noted that the repair logs may not precisely align with the timestamps of the leakages (as repair process may be conducted hours to weeks after the leakage event), making it necessary to employ external algorithms to statistically label the most probable leakage points. After identifying the outliers, which represent the leakage points, the study selects appropriate time-series examples of both leakage data (LKG) and non-leakage data (NLKG) for approximately 2000 DMAs. This process results in a total of approximately 10,000 flow data series that are utilized for training.

3.1. Handling missing and invalid data

Before training the proposed FLUIDS framework, any errors in the flow data must be identified and replaced. Errors may be in the form of missing or invalid data. The data is deemed invalid if it has a value of less than zero (as flow values cannot be negative) or if it has a preassigned 'invalid' flow validity code. Furthermore, missing data is identified by the presence of 'NaN' values or by the 'missing' flow validity code.

The need for pre-processing due to the prevalence of missing or

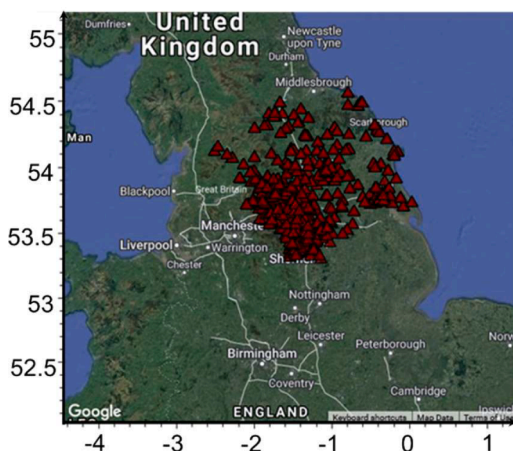


Fig. 1. DMA locations.

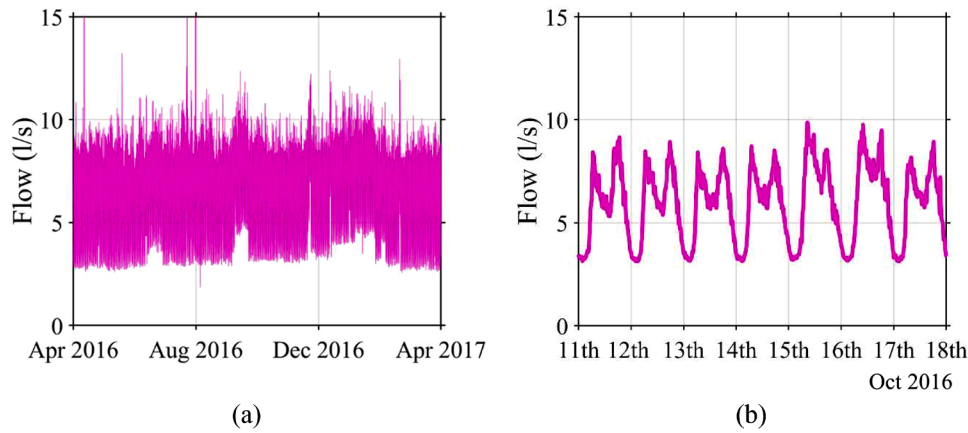


Fig. 2. Flow data from DMA 586 for (a) a full year and (b) a typical week.

erroneous data in water flow time series is a recognized issue (Moretti et al., 2022), and several methods have been proposed for dealing with this issue, including filling missing data with preceding flow values (Wu and He, 2021; Xue et al., 2022), but there is no singular method that is recognized as the standard in this field (O'Halloran and Jarrett, 2006). In order to rectify these issues in the flow data, this study uses Kalman smoothing to replace invalid data or complete the missing data (Shumway and Stoffer, 1982). Kalman smoothing is able to capture the time-varying behavior present in dynamic models by updating the estimates based on available past and future measurements. This allows for more accurate completion of missing data points, even when the system's characteristics are evolving (e.g., changes in the physical properties of the water pipes due to aging, corrosion, etc.). Kalman smoothing is also able to effectively handle missing data that follows an unpredictable pattern in terms of the frequency and length of missing sections (as is the case for the data used in this study) (Shumway and Stoffer, 2017). By smoothing the time series, not just completing the missing sections, Kalman smoothing reduces the impact of noise and outliers, helping to reveal underlying trends in the water flow data and thereby enhancing the data completeness. Thus Kalman smoothing is an ideal candidate model for handling missing or invalid time series data (Menéndez García et al., 2022; Skarlatos et al., 2023).

The smoothing process involves estimating the 'state' of the time-series data (in this case, for the flow data) before and after the missing portion of the data, to perform appropriate interpolation of the observations. This is done by first performing a forward pass through the time-series data with a KF. The KF is a state-space method that models the observation and the state of a given time-series data using Eqs. (1) and 2, respectively.

$$y_t = x_t + v_t \quad (v_t \sim N(0, \sigma^2)) \quad (1)$$

$$x_t = \theta x_{t-1} + w_t \quad (w_t \sim N(0, \tau^2)) \quad (2)$$

where y_t is the observed (or measured) value at time t , x_t represents the underlying state, θ is a tuning parameter, and v_t and w_t are noise components that are assumed to be normally distributed with a mean of 0 and standard deviations of σ and τ , respectively.

The KF is a simple dynamic Bayesian network, which estimates the underlying state at each time-step t recursively, based on a series of observed measurements assumed to be a linear combination of the state and noise (Masreliez and Martin, 1977). The KF process consists of a prediction step, where the underlying state and co-variance are estimated, and an update step, where information from the observed measurement (at time-step t) is used to revise state and co-variance estimates (Durbin and Koopman, 2012). Kalman smoothing is a post-processing method where, for a given window ($t = 1, 2, \dots, T$) in addition to a forward pass with KF, a backward recursive pass estimates each past state,

taking into account information from after the smoothing window ($t > T$) (Shumway and Stoffer, 2017). Kalman smoothing is an improvement over KF alone, as smoothing can refine estimates of previous states in the light of later observations (Briers, Doucet, and Maskell, 2009). By incorporating both past and future observations, taking into account the uncertainty and noise present in the measurements, Kalman smoothing ensures that the completed data points are good representation of the state and trends observed in the data especially in the vicinity of the missing data (Shumway and Stoffer, 2017). Hence, using Kalman smoothing for data replacement, any erroneous (i.e., invalid and missing data) flow data is rectified, and complete flow data is available to train the proposed FLUIDS framework. Sections of missing or invalid flow, as defined by the flow validity codes assigned by Yorkshire Water or by the presence of NaNs, zeros, or negative values in the flow data, are replaced by the process of Kalman smoothing.

A preliminary analysis of the initial dataset used in this study shows that ~95% of sections of missing or invalid flow data have under 480 datapoints (equivalent to 5 days compared to 1 year of total available water flow data), with over 85% of sections containing less than 96 datapoints (equivalent to 24 h). The median and mode of the missing or invalid sections are observed to be 5 datapoints (equivalent to 75 min). Hence such small a ratio of missing/invalid data is not expected to affect the Kalman smoothing process. It should also be noted that from the total available flow data only ~10,000 randomly selected groupings representing leakage and non-leakage flow are used to train the framework, and so the impact of missing or invalid data is further restricted (this is explained in Section 3.3).

Fig. 3 shows a short section of flow data from an exemplar DMA (a) before and (b) after Kalman smoothing. The replaced section of flow, which was missing in this instance, smoothly connects the preceding and subsequent data, producing a flow profile that follows the expected pattern for this section. The fluctuation from the overall flow curve is no more than is seen in the observed adjacent data. Thus, Kalman smoothing can replace erroneous sensor data with realistic values based on the available non-erroneous data. This allows for complete flow data to be provided to the leakage identification model and ensures that the anomaly detection stage of the framework is targeting leakage rather than erroneous data.

3.2. Anomalous flow detection and leakage labeling

Once the flow data has been through pre-processing to replace missing or invalid data, it is essential to label the leakage data appropriately to train the FLUIDS framework. Since leakage labels are not directly available from the provided dataset, based on existing literature (Mounce, Boxall, and Machell, 2007; Mounce et al., 2013) it is assumed that some groups of outliers in the flow data are indicative of leakage. To

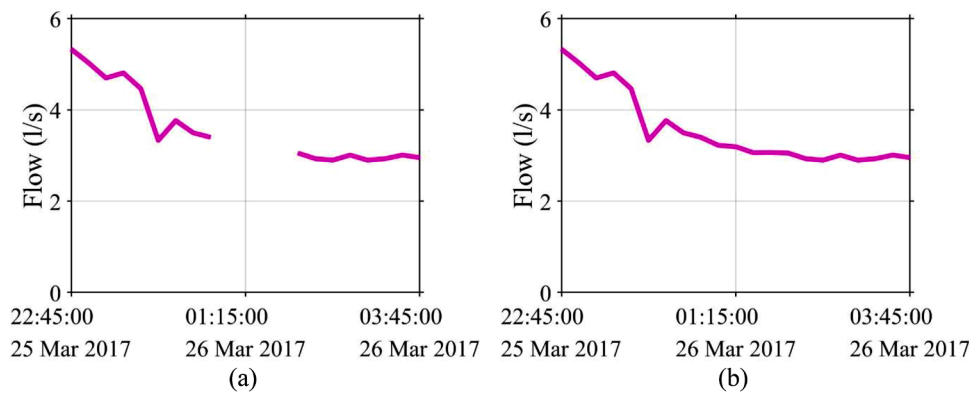


Fig. 3. Section of missing data (a) before and (b) after Kalman smoothing, DMA 586.

ascertain that the identified outliers highly likely represent leakage, additional steps were adopted based on a review of relevant literature and discussions with experts at Yorkshire Water. As there may be other causes of outlier flow, further criteria must be met in order for an outlier grouping to be defined as probable leakage. First, this section verifies the assumption that many outliers correspond to leakage by checking the time difference between the detected outlier and the nearest repair log, as it is expected that the pipe repair would have been conducted quickly after the leakage burst. The following section then details the additional criteria that must be met for outliers to be selected as LKG groupings, which are used to train the proposed framework.

It is important to acknowledge that this holds true only for leakages that are eventually detected by the water company, either through customer reports or flow monitoring. Leakage that occurs gradually and remains undetected for an extended period may not have a corresponding repair log and might not be identified as an outlier, especially if the leakage began before April 2016 (the start of the study period). Consequently, the outliers selected to train this framework are more likely to represent new bursts or leakage events rather than background leakage. If data on background leakage were to become available, the framework could be trained using relevant examples. However, such data is not accessible for the current study. Nevertheless, the analysis of $|Z|$ values, as presented in the results section of this study, confirms that the points selected during the data pre-processing stage indeed represent genuine outliers.

Outliers are identified using a tree-based unsupervised machine learning algorithm known as isolation forests (Liu, Ting, and Zhou, 2008). The concept behind isolation forests is that the outliers are typically easier to isolate than normal data points, based on their rarity and different attribute values. In terms of decision trees, this places

outliers closer to the root node than expected data points. The outliers/anomalies are distinguished from the non-outliers/anomalies based on a hyperparameter called contamination fraction which sets the classification threshold (Liu, Ting, and Zhou, 2012). In this study, a contamination fraction of 0.005 is selected. The anomaly detection process allows the data to be labeled as outliers and non-outliers. Outliers can then be further analyzed to establish whether they are indicative of potential leakage.

Outliers are detected in the completed flow data for all DMAs using the isolation forests algorithm. Fig. 4 presents the outliers detected inflow data of two DMAs. The dashed lines in the figures also show the dates on which the corresponding DMAs were repaired, based on the repair log. It can be observed from the figures that the algorithm performs well in identifying both extreme outliers and extended periods of unusual flow rates. The detected outliers, particularly extreme outliers, seem to correlate well with repair dates. Though some repair dates are observed to be away from the outlier data, this can be due to the repairs being conducted for reasons other than pipe leakage, which are not of interest in this study. It is further observed that the algorithm also flags some other unusual flow data points that do not appear to be leakages. Hence, it is crucial to identify outlier groups so that only extended periods of irregular flow are flagged as possible leakage, while isolated individual outliers are discarded. For this reason, LKG groupings are required to be a minimum of 20 outliers in length. The literature supports this approach in the field, suggesting that anomalous flow shorter than a few hours in length is likely not leakage but sensor error, fire-fighting, or an industrial event (Mounce, Boxall, and Machell, 2007).

Identifying accurate leakage data points is essential in developing a reliable flow forecasting tool that can accurately anticipate bursts. The assumption of using the outliers as the proxy for real leakage is further

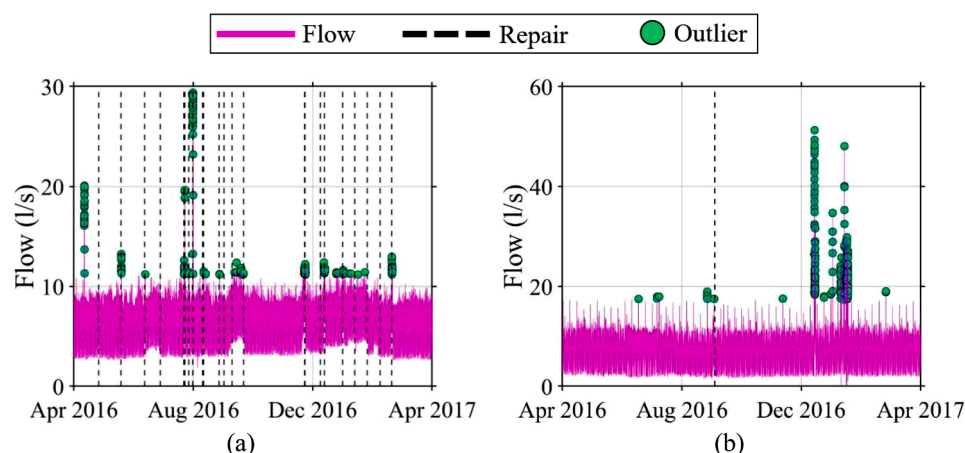


Fig. 4. Outlier identification and repair dates for DMAs: (a) 586, (b) 1316.

validated by comparing the time differences between the outliers and repair dates from the repair log. Some fluctuation between the leakage and repair times is expected, as the time taken to respond to suspected leakage can vary depending on factors such as accessibility, size, and visibility. Also, since the repair logs only contain the dates of repair (rather than exact timestamps), outliers identified on the same day or closely before or after the repair can be used to validate the leakage labels conducted via isolation forests. Fig. 5 shows the time difference between each repair and the closest outlier before and after the repair for DMA 586, which has undergone 22 repairs during the year. It is noted that outliers largely correspond well with the recorded repairs, with a significant number of repairs occurring within 24 hours of an outlier. Those repairs that occur within 24 hours of an outlier likely represent the visible leakages reported by the customers. Almost all recorded repairs in the example shown in Fig. 5 took place less than 16 days after a period of outlier flow, with many taking place less than ten days after an outlier. This falls well within the repair timescale that would be expected for less urgent, non-visible leakage or leakage on land requiring permissions for access. These findings confirm that repair data is the best proxy for confirmed bursts.

3.3. Data preparation

Once the outliers have been identified, these need to be grouped into LKG and NLKG data in order to train the forecasting model. Outliers that sit very close together (within a few hours of each other) are likely to indicate a single burst rather than two distinct bursts, so these are grouped together. Hence, if the outliers are within six hours of each other, they are placed within the same outlier grouping, with interim data points also labeled as outliers. This is done for all outliers for each DMA flow record. A minimum length of outlier grouping is chosen, as short periods of anomalous flow or unique outlier points may indicate causes other than leakage (such as anomalous usage). Additionally, these groupings must be of sufficient length to train the LSTM-RNN. The length of each outlier grouping is computed, and the ones with fewer than 20 data points are ignored for this study. The literature suggests that shorter periods of anomalous flow tend to represent firefighting or industrial events rather than leakage (Mounce, Boxall, and Machell, 2007). The finally selected outlier groupings represent the potential leakage groups. This LKG flow data is expected to be forecasted by the final trained FLUIDS framework, thus preceding flow data is required to be used as the input. Hence the outlier/leakage (LKG) groupings must also be preceded by a sufficient amount of non-outlier datapoints to serve as an input to the LSTM-RNN. To have sufficient data for training, this input data needs to be equal to or greater than the LKG data in length. LKG groupings that do not satisfy this requirement are discarded. The maximum length of input data is set to 672 data points, representing a week's flow data. This is deemed sufficiently long to give a representative sample of flow before an outlier. A total of 3409 LKG groupings

are selected with these criteria. Since the RNNs, or any ANN-based model, require a set number of input and output data points, the length of LKG groupings and the preceding input data must be the same for all examples. The maximum length of LKG data is observed to be 335 points. Hence all inputs need to have 672 data points, and the outputs need to be 335 in length. This consistency is obtained by zero-padding, where zeros are added before the flow data for inputs and after the flow data for outputs (i.e., LKG data). Zero padding ensures that all LKG groupings have the same total length, which is required to train the LSTM-RNN. This process prepares the leakage dataset. An example of a LKG grouping is provided in Fig. 6a.

As the proposed FLUIDS framework aims to forecast both regular flow data and leakage flow data, a selection of non-outlier/non-leakage data (NLKG) is also selected and merged with LKG data to train the LSTM-RNN. The NLKG groupings are equal in length to the LKG groupings (i.e., 335 data points) and are similarly divided into input and output sections. However, as the NLKG groups are more prevalent in the dataset than the number of LKG groupings, to avoid any data bias all the NLKG groupings cannot be used to train the proposed FLUIDS framework. A significant difference in the sample sizes of the two groupings can cause considerable bias in tuning the models. Hence, random samples of NLKG groupings are obtained from the ~2000 DMA flow datasets with different sampling ratios between the NLKG and LKG groupings. Based on the performance of the RNN (discussed in section 3.2.2), a sample size with NLKG samples equivalent to two times the number of LKG samples was used for further study. An example of a NLKG grouping is shown in Fig. 6b. In summary, both LKG and NLKG groupings contain preceding flow data which is used as inputs to the LSTM-RNN, and the outputs consists of the following outlier grouping in case of LKG data or following regular flow in case of NLKG data.

To ensure the input data does not contain any erroneous datapoints (possibly due to undetected sensor error), variance checks are performed on the input data section of both LKG and NLKG groupings, and those with exceptionally high or low coefficients of variation (COV) are discarded. Any LKG group with a coefficient of variation (COV) below 0.1 or greater than 10 for the input section of data is discarded. Furthermore, for NLKG groupings, the output section of the flow data is also required to have a COV between 0.1 and 10. This is done to ensure that the selected non-outlier information is error-free and doesn't contain any non-detected peaks or portions of unbalanced flow (any unexpected/unlabelled malfunctioning). In total, ~10,000 LKG and NLKG groupings are used as the final dataset, with 672 points of inputs and 335 points of outputs. This dataset is then used to train and test a hybrid forecasting model for leakage prediction.

4. Methodology

The general procedure for training the proposed FLUIDS framework is illustrated in Fig. 7. The details are explained in the following sub-

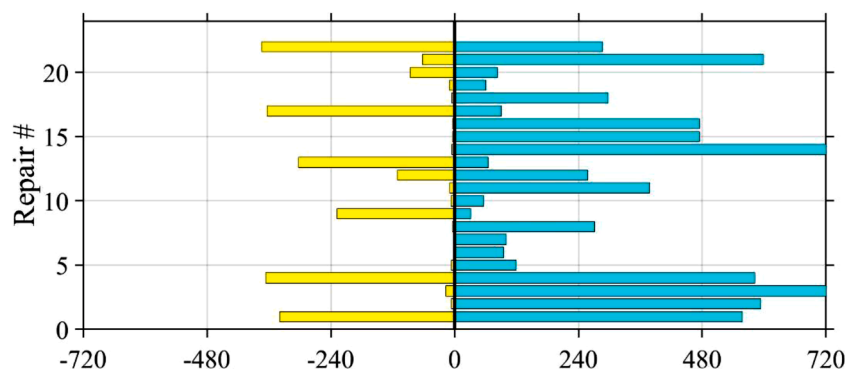


Fig. 5. Time difference between repairs and closest outliers (hrs) for DMA 586.

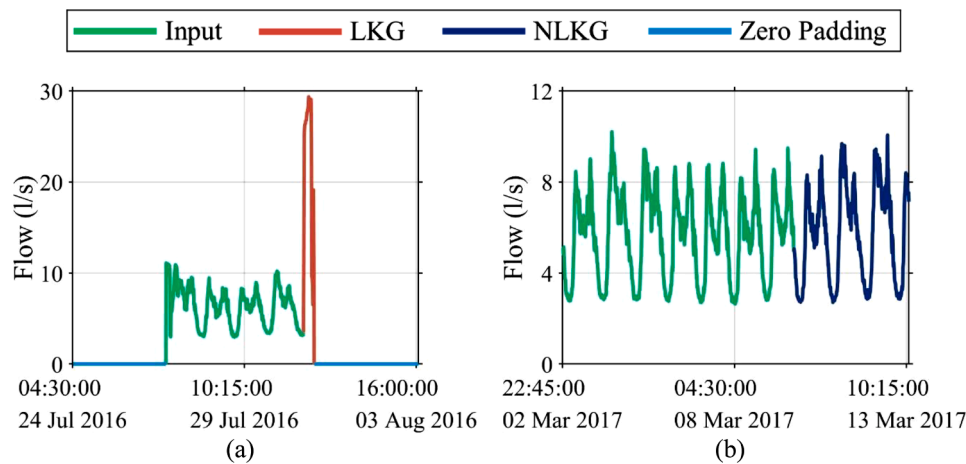


Fig. 6. Examples of groupings for (a) LKG data: DMA 586 outlier 1, (b) randomly selected NLKG data: DMA 586 group 1.

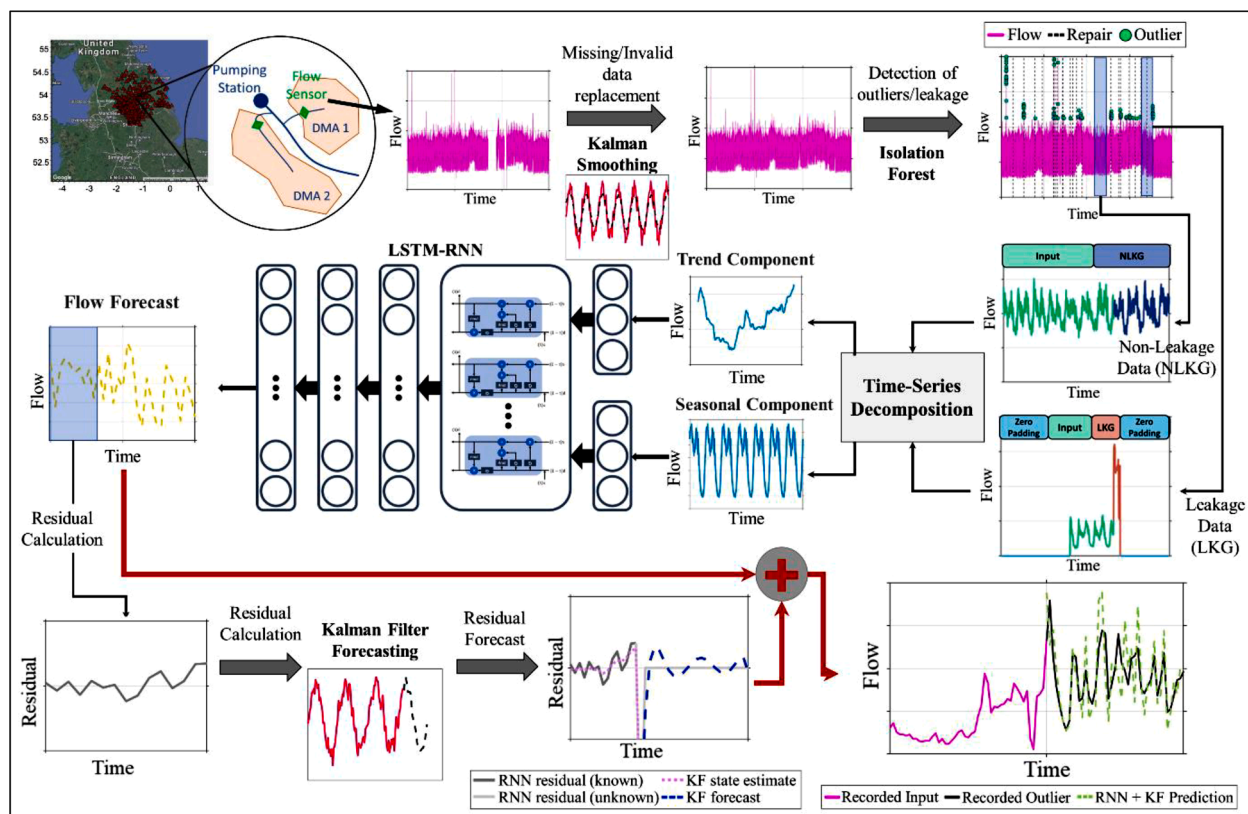


Fig. 7. Procedure for training the proposed FLUIDS framework.

sections. First the selected 10,000 flow data series undergo a time-series decomposition to separate their trend and seasonal components. These decomposed components serve as inputs for training a LSTM-RNN, which aims to forecast the mean flow data. It is important to note that both LKG and NLKG data are included in the training dataset, enabling the LSTM-RNN to predict flow behavior during burst periods, rather than solely focusing on regular non-leakage flow. The forecasted mean flow data from the LSTM-RNN is then used to calculate the residuals, representing the differences between the predicted and recorded flow values. These residuals are subsequently utilized to train a boosting KF, which enables real-time forecasting of future residuals, further enhancing the accuracy of predictions. Finally, the forecasted residuals are combined with the mean forecast generated by the LSTM-RNN to

derive the final predictions. Consequently, the FLUIDS framework consists of two main components: 1) LSTM-RNN, trained to forecast the expected flow data for $t + n$ time points using $t - m$ recorded flow values (where t is the current time step, n is the number of future timesteps, and m is the number of previous timesteps), and 2) KF, which provides real-time estimates of residuals, thereby refining the LSTM-RNN's predictions and aligning them closer to the true flow behavior.

As state-space models like KF are widely based on the assumption of stationary time series, this assumption is validated in this study using an augmented Dickey-Fuller unit hypothesis test (Mushtaq, 2011). The test has the null hypothesis that the time series is non-stationary and has a unit root. For the flow data in this study, p -values are observed to be well below the significance level of 0.05, and the null hypothesis is rejected.

This indicates that the flow data can be deemed stationary for state-space and other time-series modeling techniques.

4.1. Flow forecasting

4.1.1. Mean flow forecasting

Effective forecasting of anomalous flow data can allow leakage to be anticipated, facilitating a more efficient approach to leakage management and system maintenance. By forecasting the anomalous flow data, itself, rather than forecasting regular flow data and comparing this to incoming anomalous data, earlier warning can be provided of potential bursts and estimated water loss calculated. In this study, the ANN-based model is used to forecast the mean flow for a future period of time. It is well known that a stationary time series typically consists of two general components: i) trend and ii) seasonal (Shumway and Stoffer, 2017). The trend component represents the general pattern of the time-series data over the entire time duration. In contrast, the seasonal component refers to the cyclic repetition of a trend within a specific time period. Neural networks can potentially struggle to model seasonality directly from time series data (Nelson, Hill, Remus, and O'Connor, 1999). Hence, to address this issue, additive time series decomposition is used to break down the input data of both LKG and NLKG groupings into a trend, seasonal component, and the remaining noise (Cleveland and Tiao, 1976). In this case, the trend represents the general pattern of flow data over the input time window, while the seasonal component reflects the fluctuations in flow during the 24 hours. Using the time-series decomposition (Eq. (3)), the input flow time series y_t is decomposed into a trend, seasonal, and noise components, where T_t and S_t are the trend and seasonal components at timestep t , and ε is the noise in the data which is assumed to be normally distributed with a mean of 0 and a standard deviation of δ .

$$y_t = T_t + S_t + \varepsilon (\sim N(0, \delta^2)) \quad (3)$$

In this study, the decomposed trend and seasonal components are used as inputs to train an LSTM-RNN, which outputs a forecast of the NLKG/LKG flow section of the groupings. RNNs are a class of ANN developed for modeling time series data (Rumelhart, Hinton, and Williams, 1986). RNNs allow the output of a neural network layer at time-step $t - 1$ to be used as inputs for the same neural network layer for the following time-step t . This forms a directed graph and allows the transfer of 'memory' between adjacent time steps so that the output of the neural network layer at a given time step is dependent on prior elements within the time series.

Although RNNs can handle dependencies between individual steps in a time series, they suffer from issues with long-term dependencies and vanishing gradients (Hochreiter, 1998). As a result, RNNs struggle to learn if asked to use outputs from previous time steps many steps back (time lags) as inputs for estimating the current time step. Selecting an LSTM architecture for the RNN can solve these challenges. Each LSTM layer contains a set of recurrently connected blocks, with one or more recurrently-connected memory cells and three multiplicative gates regulating information flow (Graves and Schmidhuber, 2005). A cell state transfers relative information down the sequence chain and between LSTM blocks - the 'memory' of the network. In each LSTM cell, a forget gate passes on information from previous outputs and the current input at time-step t and decides what data to keep in the cell state. An input gate decides how the current input should be used to update the cell state and modify the memory, and an output gate uses the input and the memory of the cell to decide the output for the current time step. Thus LSTM cells act as information processing units and provide a route for 'memory' to pass beyond adjacent cells, enabling the RNN to bridge long time lags steps (Hochreiter and Schmidhuber, 1996; Gers, Schmidhuber, and Cummins, 2000). While other deep learning-based models such as feed-forward networks, gated recurrent unit (GRUs), etc., have been known to work well some time-series applications, for

long-memory time-series such as water flow, those models also suffer from vanishing gradient phenomenon (Cahuantzi, Chen, and Güttel, 2023). Hence in this study, LSTM-RNNs are used to develop the flow forecasting model.

The LKG and NLKG groupings are split into train and test sets such that 80% of both groupings are used in training and 20% of both groups are used for testing the LSTM-RNN. LSTM-RNNs of various configurations and hyperparameters are developed and trained. In particular, index of agreement (IA) (Willmott, 1981; Willmott et al., 1985), described in Equation 4, is used as the loss function for training and testing an LSTM-RNN, where O is the recorded output flow data and P is the RNN predicted flow data, and n and i represent the total number of forecasted timesteps and the timestep of interest, respectively. A valuable tool for the comparison of model performance, IA gives a single bounded metric for pattern characterization and comparison, yet also incorporates information on the magnitude of deviations into this metric and has therefore been widely applied to the assessment of model-produced estimates of time-series data (Duveiller, Fasbender, and Meroni, 2016). When compared to traditional loss functions such as mean squared error (MSE), IA is shown to achieve better performance on time series data due to the reduction of bias in both high and low values rather than the average bias (Sayeed et al., 2021).

$$IA = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (4)$$

The various architectures are tested using the train set, and the best performing combination of LSTM-RNN architecture and hyperparameters is then selected. The chosen LSTM-RNN configuration, once trained, provides mean flow forecasting.

4.1.2. Residual forecasting

Since the LSTM-RNN is a pre-trained model, it cannot adjust to fluctuations in real-time; hence, to further improve the predictions, a boosting concept (Chen and Guestrin, 2016) is utilized to model the early residuals observed between the LSTM-RNN-based flow forecast and the real-time sensor recording of corresponding flow data. The early residuals are then used to estimate the future residuals for the remaining uncertain period of the output. Thus, modeling of residuals can improve overall forecast accuracy by providing an estimate of the error expected from the LSTM-RNN due to any real-time fluctuations. This study uses KF as the boosting model for forecasting the residuals in real-time (Harvey, 1990). The KF is a Bayesian method for sequentially estimating the states of a dynamic system where the state evolution and measurement models are linear and Gaussian (Kovvali, Banavar, and Spanias, 2013). The recursive nature of the KF enables it to model continuously changing systems. KFs also do not need to hold much memory and thus can be run very quickly, making them ideal for real-time applications (Maybeck, 1990). In the residual forecasting model, the KF algorithm first uses Kalman smoothing to estimate the state of the observed residuals and then forecasts residuals for a pre-defined forecast period. Forecasting uses the observation and state equations recursively (Morrison and Pike, 1977). Given an initial estimate at time t , the KF first performs a prediction step, estimating the state at time $t + 1$, as well as the uncertainty of this prediction. Once the observed value at $t + 1$ is received, a correction step is performed. A calculation for Kalman gain adjusts the weights given to the incoming observations and current-state estimate. The prediction and uncertainty estimates are then updated based on this new information, and the cycle repeats for the next time step (Durbin and Koopman, 2012). This continues for known observations at time-steps $t = 1, 2, \dots, T$. Having provided the KF with sufficient known observations to tune parameters such as the co-variance estimate and the Kalman gain, the KF can be used to forecast for a defined period of future time-steps ($t > T$), where observations are unknown. This process is repeated in real-time for complete predictions. As more recorded values are received, the KF model can

update residual predictions to reflect this new information. The final flow forecast from the proposed FLUIDS framework is based on the addition of the mean forecast from the LSTM-RNN and the forecasted residuals from the KF. Fig. 8 shows the processing of incoming flow data through the proposed framework and final flow forecasting.

5. Results

This section presents the results of the trained framework. The mean and residual components of the forecasting method are presented first separately and then in combination, and the accuracy of the model is assessed. Furthermore, a comparison of the results of this study to a simple MNF method of leakage detection is presented, to demonstrate the benefits of the FLUIDS framework over traditional methods.

5.1. Mean flow forecasting

Fig. 9 illustrates the outcome of time series decomposition on exemplar input data. The trend component shows the overall pattern of change in flow across a week, while the seasonal component captures the daily flow pattern. This typical pattern, with twice-daily peaks and a significant drop overnight, reflects typical water consumption over 24 h and is seen in most input data. The trend component is more variable across LKG/NLKG input groupings, as this is affected by factors such as which (and when) days of the week appear in the input data and if and how leakage is reflected in the input data. In order to ensure all relevant

patterns are considered, these two components are separately input into the RNN.

The 10,227 input and LKG/NLKG data groupings are randomly split into train (80%) and test (20%) sets while making sure train and test sets consists of same ratio of LKG and NLKG data. The training is conducted with 10% cross-validation. After hyperparameter tuning, the best performing final LSTM-RNN architecture in terms of *IA* is shown in Fig. 10. In particular, the LSTM-RNN network is trained using stochastic gradient descent (Kiefer and Wolfowitz, 1952) with Adam optimizer (Kingma and Ba, 2017) and *IA* (Willmott, 1981; Willmott et al., 1985) as the loss function. Since the values of *IA* range from 0 to 1, with 1 being the best match and 0 as the worst match, the loss function is used negatively to allow gradient descent rather than ascent.

The trained RNN uses flow data's trend and seasonal components as inputs and predicts the flow for the future 335-time-steps (i.e., LKG/NLKG data for 335 15-minute intervals). *IA* values are calculated for each grouping to assess how well these predictions align with the observed LKG/NLKG data. The left part of Fig. 11 presents the distribution of *IA* values for all the groupings. Overall, this *IA* profile indicates good performance by the RNN, with predicted values and observed flow in good agreement. The vast majority of groupings have an *IA* value over 0.5, with a first peak between 0.5 and 0.6 and a second, more prominent peak between 0.8 and 0.9. The reason for these peaks can be due to the differences in the 'type' of grouping, so factors that vary between groupings are further investigated.

Due to different magnitudes of outliers, LKG groupings vary signifi-

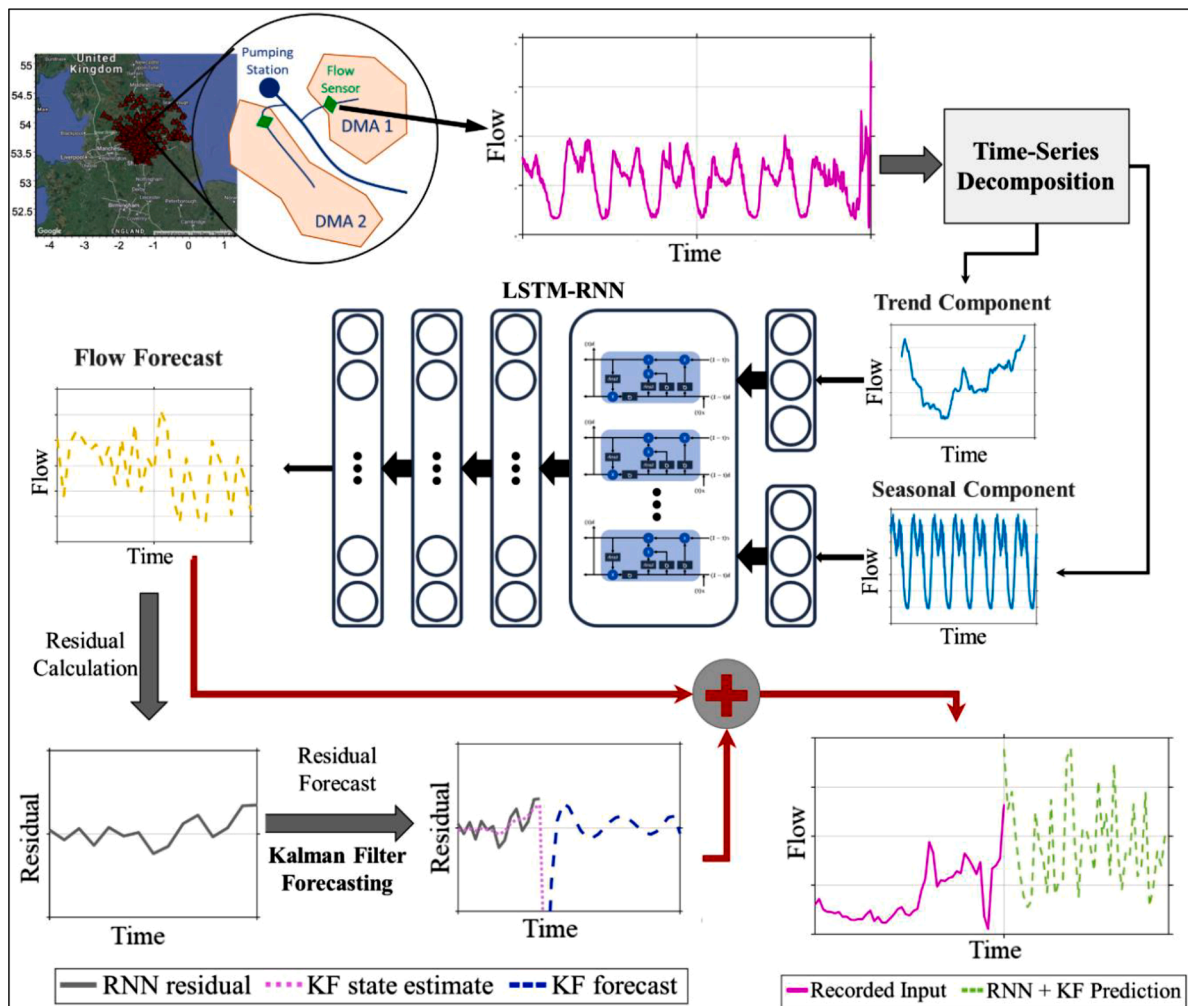


Fig. 8. Proposed FLUIDS framework for flow forecasting.

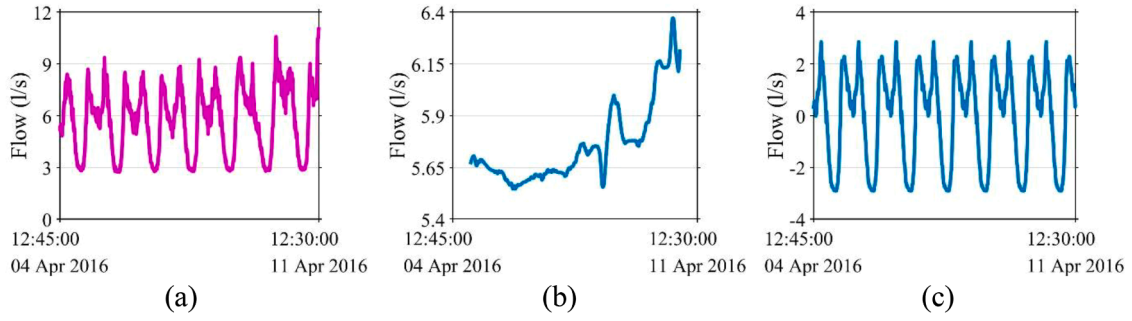


Fig. 9. DMA 586 outlier 1 (a) input data, (b) trend component, (c) seasonal component.

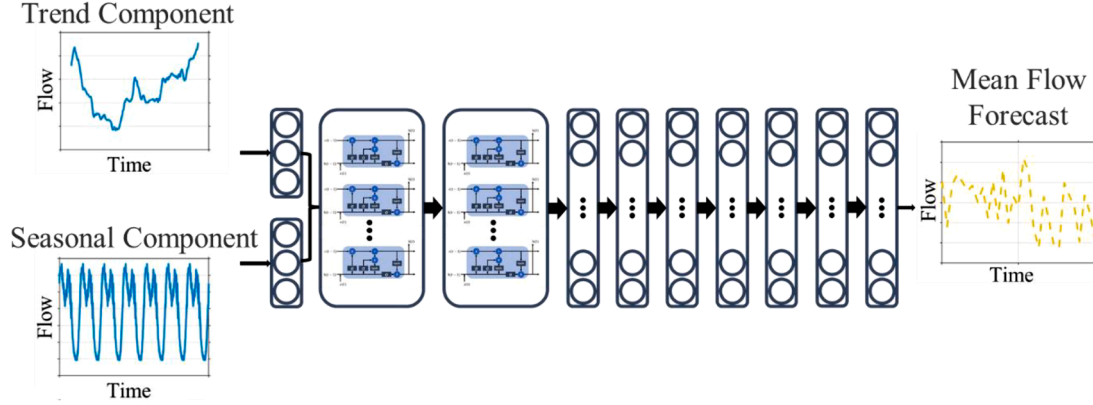


Fig. 10. Architecture of the trained LSTM-RNN.

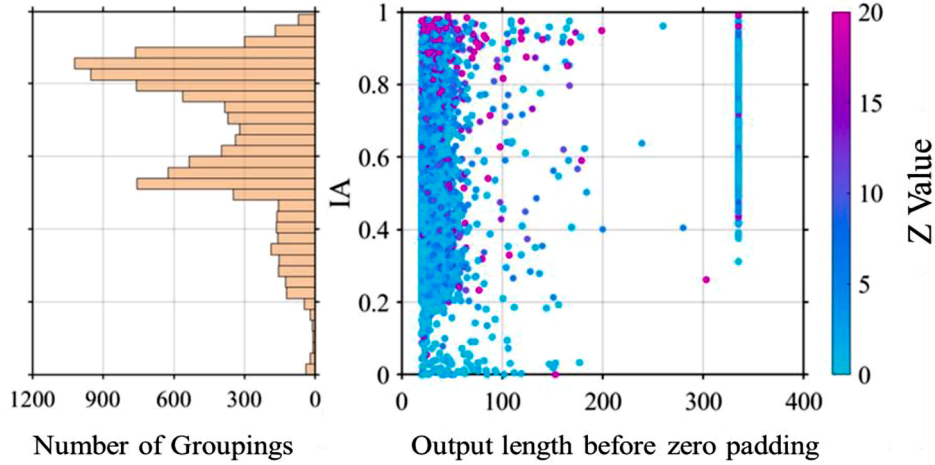


Fig. 11. IA values, $|Z|$ values, and output lengths (without zero padding) of all groupings.

cantly in length and volatility (i.e., the magnitude of LKG flow compared to the magnitude and variance of preceding input flow). Hence, it is necessary to ensure that the RNN predictions are not biased towards LKG groupings with lower volatility compared to the high volatility groupings. This is done by computing the $|Z|$ values for each output grouping (LKG/NLKG groupings) using Eq. (5). In this equation, $peak_{out}$ and $median_{in}$ are the largest value in the LKG/NLKG section of flow and the median value in the input flow, respectively, while σ_{in} is the standard deviation of the input data. The $|Z|$ value thus compares the size of the output peak to the size and variability of the preceding input data. As can be observed from Fig. 6, the peak of NLKG flow is not as high as the peak of LKG flow when compared to the median of the preceding input flow. Also, peak flows can vary significantly based on the burst level

within the LKG groupings. Therefore, computing the $|Z|$ values and comparing them against the corresponding IA values allow the detection of any unintended bias in the model.

$$|Z| = \frac{|(peak_{out} - median_{in})|}{\sigma_{in}^2} \quad (5)$$

Furthermore, since the LKG/NLKG parts of the groupings vary in length and are zero-padded (as described in previous sections), it is essential to check any potential bias in the LSTM-RNN predictions concerning the non-zero padded length of the output data. The right side of Fig. 11 shows the IA values for all the $\sim 10,000$ examples compared to the output data's $|Z|$ values and non-zero padded length. The color of each dot represents the $|Z|$ value of each grouping. It is observed that

25% of the selected LKG groupings have a $|Z|$ value greater than 5, indicating a low probability (less than 0.00001) of randomly having $|Z| > 5$, and the median $|Z|$ value of LKG groupings is 2.1, with a probability of $|Z| > 2.1$ being less than 0.035. This confirms that the pre-processing method has captured genuine outliers in the dataset. Additionally, the additional criteria for LKG group selection ensure the capture of flow patterns typical of leakage, characterized by a significant spike in flow data that surpasses the fluctuations in preceding data.

As the most extended LKG group was 335 data points in length, many groupings possess this length without any zero padding (especially NLKG data). While higher values of IA are observed across the different LKG group lengths, the concentration of higher $|Z|$ values in the top left of the plot suggests that the proposed model performs particularly well on leakages with large flow magnitudes and shorter LKG lengths. This may indicate that the preceding flow data for such LKG groups follow a more identifiable pattern captured by the LSTM-RNN. While this is an interesting hypothesis, it is beyond the scope of this paper to statistically investigate. Conversely, the lowest IA scores are seen in LKG groups with common $|Z|$ values, suggesting that the LSTM-RNN struggles to forecast accurately if the peak values are small and the variability in the preceding flow is high.

To check the patterns of the LSTM-RNN predictions against the recorded output flow, Fig. 12 shows four quantile cases from the test data (20% of the dataset). Each of these cases represent a period of anomalous flow data flagged as LKG, rather than a period of regular, non-anomalous flow. While there are some fluctuations, the LSTM-RNN forecasts largely follow the overall pattern of the recorded flow data in both magnitude and direction of change (increase/decrease). In some areas, the flow forecast fluctuates more than the recorded data. In most

cases, as seen in the examples shown for the 50th, 75th, and 99th IA percentiles in Fig. 12, this fluctuation tends to be distributed relatively evenly above and below the flow profile of the recorded outlier, suggesting that the overall pattern of flow is well captured. All four examples in Fig. 12 see a higher forecast value for the first forecast flow datapoint than the recorded flow value. In terms of leakage management, that the LSTM-RNN is more inclined to an initial overestimate than an underestimate means that leakage is less likely to be missed by the model. However, particularly at the higher IA percentiles, the overestimation and underestimation in the prediction across the entire forecast window seem well balanced. Therefore, the model could be expanded to offer an accurate prediction of the quantity of water loss via leakage. It is observed that even at 25th percentile, the IA value exceeds 0.5, with IA rising to over 0.7 at the 50th percentile. Forecast accuracy, and thus IA values, can be expected to improve with the addition of residual forecasting using the Kalman forecasting.

5.2. Residual forecasting

As the LSTM-RNN model is trained to estimate the mean flow using the known trend and stationary components of the preceding flow, the weights of the LSTM-RNN network are pretrained. They are expected to perform the flow forecasting with known causality. However, to improve the proposed FLUIDS framework's real-time performance, the residuals obtained in real-time are further used to develop the state-space model using KF to appropriately forecast the future residuals. Using the pre-trained LSTM-RNN, the flow forecast is obtained from current time t to n time-steps ahead to $t + n$ time, and then as the true values of flow are observed in real-time for time-steps t to $t + k$, where

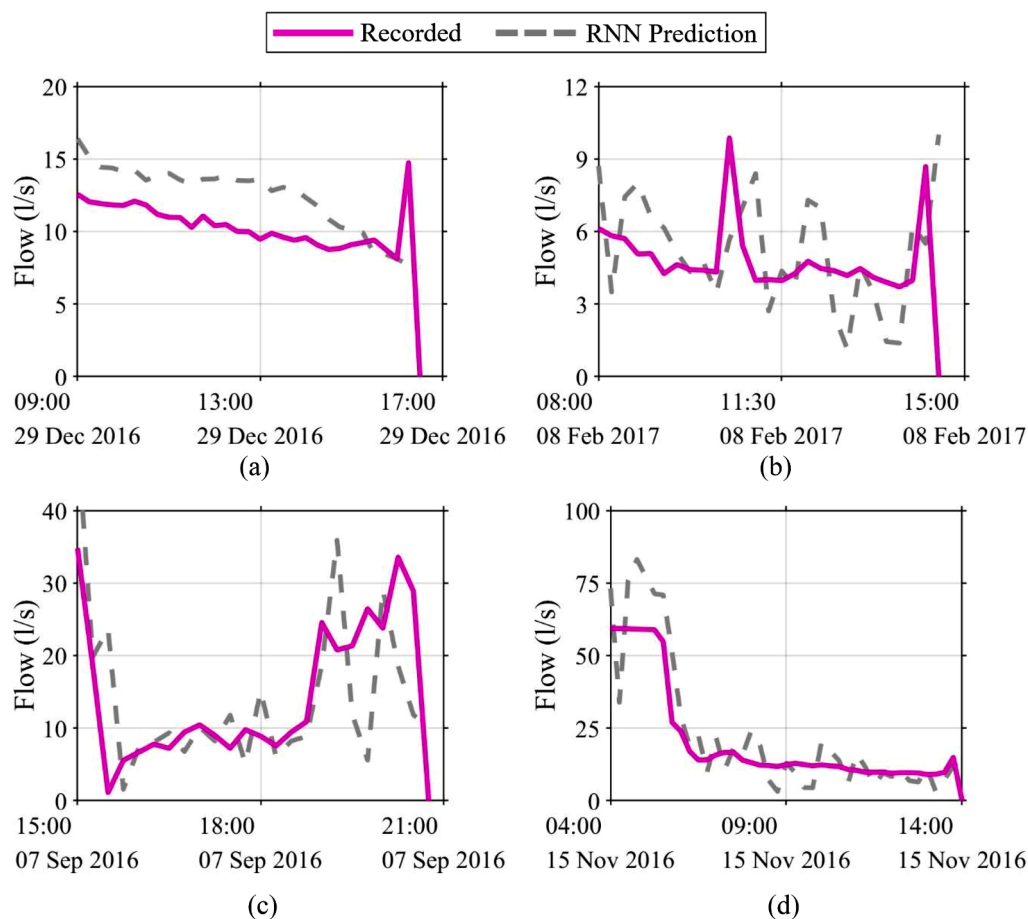


Fig. 12. Examples of true vs LSTM-RNN predicted flow for: (a) 25th percentile: DMA 1672 outlier 3 ($IA = 0.5362$), (b) 50th percentile: DMA 1999 outlier 2 ($IA = 0.7098$), (c) 75th percentile: DMA 189 outlier 4 ($IA = 0.8307$), (d) 99th percentile: DMA 1085 outlier 2 ($IA = 0.9554$).

$k < n$, KF is used to model the residuals by finding the difference between the LSTM-RNN forecast and the recorded flow. Due to the recursive nature of KF estimates, this process is expected to provide the FLUIDS framework with real-time deviations of the data and improve the accuracy of the hybrid forecasting system. The results of KF-based forecasting of the LSTM-RNN residuals are presented for an exemplar DMA in Fig. 13. This example chooses a prediction window of six hours (24 data points).

As the observed values are considered to be the sum of the underlying state plus noise, KF is performed on known residuals (time-steps $\leq t$) before forecasting so that the prediction can be based on the estimated state rather than the observed values. KF is then used to obtain a forecast for n time-steps ahead. As more data points for the flow are recorded, the residuals are computed, and the updated model is used to forecast the residuals for a future time window. Fig. 13 shows the estimated states for $t =$ (a) 24, (b) 48, and (c) 60 and KF forecast for $n = 24$ during a period of LKG flow, demonstrating how the KF is used to provide a forecast of residuals with a real-time rolling time window. It can be observed that forecasting power is improved as more residuals are provided to the model. The KF demonstrates strong performance in both state estimation of known residual data and forecasting the unknown residual data. The state estimation step smooths the observed data, with the estimated states showing less volatility than the observed residual values.

Similarly, while the forecast can predict changes in the overall trend of the residual data, many peaks in the observed residual data appear less extreme in the forecasted data. Although huge spikes in residual data may be underestimated in the forecast, the KF effectively captures the overall pattern of residuals. Therefore, adding residual forecasting to mean flow forecasting will allow real-time updates to forecasting and improve the accuracy of the final combined prediction.

5.3. Final flow forecasting

Finally, the results of the LSTM-RNN and KF predictions are combined to obtain a final flow forecast. This is presented for an example case of LKG from the test set in Fig. 14. Note that the x-axis is split to provide greater detail for the forecasted section of flow. In this example, 60 residuals are provided for a forecast window of 24 residuals. It can be observed that the combined forecast appears to match the recorded outlier well. The forecast anticipates the fluctuations in flow for the outlier period, during which the daily water consumption pattern is much less precise than for the input data. The forecast also matches the peak of the outlier well in both magnitude and time, though both peaks and dips can be overestimated in the prediction.

Captured in this example is both an elevated daytime and nighttime flow, relative to the input data. The prediction also captures the significant drop in flow from day to night, despite the leakage. This drop corresponds to the overnight period often used to calculate MNF in other

leakage identification studies. The forecast captures both the reduction in water usage from day to night and the elevated nighttime flow level indicative of leakage in studies using MNF. In shorter outlier groupings where the nighttime period is not represented within the outlier grouping, extending the forecast beyond the outlier period may be beneficial to verify whether the predicted minimum remains elevated compared to overnight periods in the input data, which would be expected during leakage. As this prediction shows strong agreement with the recorded data throughout the outlier period, not just the overnight section, and the increased MNF is accurately anticipated by the forecast, it is shown that this method, unlike many traditional leakage identification methods, does not necessitate a full overnight period of flow data to identify leakage. Instead, anomalous flow behavior can be accurately determined and anticipated during daytime hours. This allows for more rapid flagging of leakage and thus can facilitate more timely and less disruptive repairs. For the LKG group shown in Fig. 14, the LSTM-RNN mean flow forecast has an *IA* of 0.8466. When combined with the residual forecast, however, the *IA* for this group rises to 0.9240. This improvement demonstrates the value of this hybrid modeling method.

5.4. Comparison with minimum night flow (MNF)

Finally, the groupings flagged as LKG are analysed for burst detection using the traditional MNF methodology. It should be noted that, unlike the FLUIDS framework proposed herein, MNF is only a detection methodology and thus can only be applied to leakage detection and does not possess any flow prediction or forecasting capability. Nevertheless, this exercise is conducted to showcase the efficacy of the FLUIDS framework proposed in this study.

The MNF method compares flow data for a given time of the day where the recorded flow is expected to have minimal variation with flow data for the same time of day during the preceding days or months. The time of day chosen is generally during the night, and so the flow during this time is termed ‘night flow’ (NF). MNF requires a preceding “window period” of regular flow (flow with no leakage events/bursts occurring) to calculate the average NF, against which new NF can be compared. While the size of the window period varies in the literature, it is generally agreed that a larger window leads to a better representation of typical NF behavior for a given DMA. The range of window periods in the literature vary from three days to six months (Amoatey, Obiri-Yeboah, and Akosah-Kusi, 2021; Lee, Lee, and Lee, 2022; Tabesh, Yekta, and Burrows, 2009; Hamilton and McKenzie, 2014; Huang et al., 2018). Similarly, the nightly hours used to calculate NF vary in literature, beginning as early as 12am and ending as late as 5am (Lee, Lee, and Lee, 2022), but the hours of 2am to 4am are a common selection in pre-existing research (Mushtaq, 2011). The median value of flow during these hours over the window period is found, as MNF is sensitive to fluctuations or anomalies. The median value of flow data is often used to

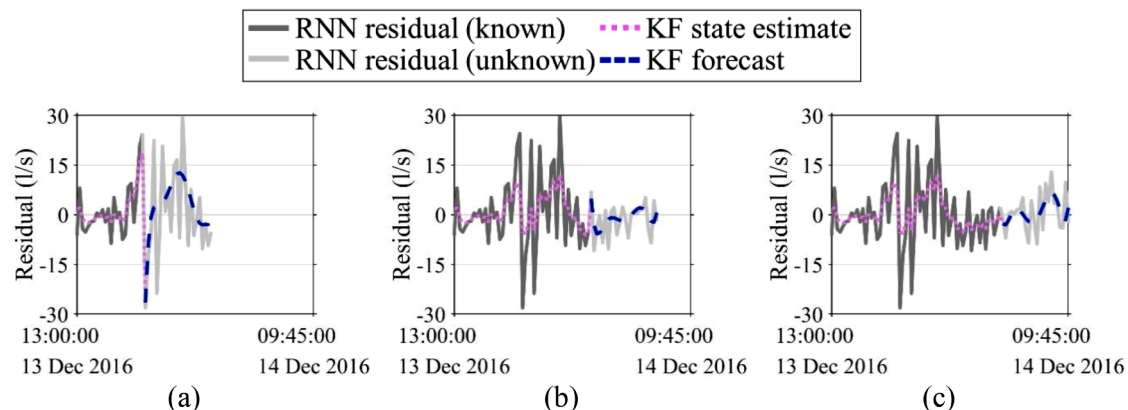


Fig. 13. Residual plots for 24 steps ahead, DMA 1316 outlier 1 with: (a) 24, (b) 48, and (c) 60 inputs.

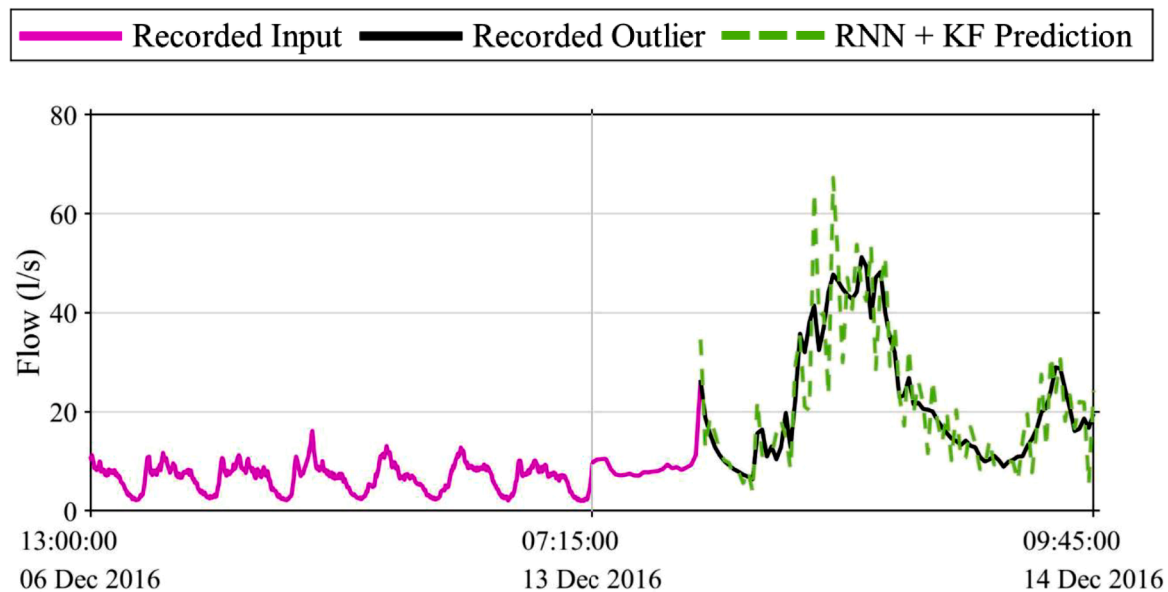


Fig. 14. Forecast for DMA 1316 outlier 1 ($IA = 0.9240$).

limit the effects of erroneous data. A significant deviation from this value during the same hours of the following night is taken to indicate new leakage under MNF analysis. The deviation is computed in terms of percentage deviation between the NF of input data and the NF during the night of interest using Eq. (6), where d is the 24-hour day of interest and i is the size of the window period in days.

$$\text{Percentage deviation} = \frac{\text{median}(NF_d)}{\text{median}(NF_{d-1}, NF_{d-2}, \dots, NF_{d-i})} \times 100 \quad (6)$$

Due to the limitations of the dataset, this study uses three days of window period of preceding flow. Based on this criterion, around 1300 LKG group had sufficient length of input data to be used for MNF analysis (other LKG groups were discarded as they didn't contain enough regular flow points between the outlier data). This requirement of long preceding flow is also a limitation of MNF as compared to the proposed FLUIDS framework, which is able to operate with minimal input data length (set to a minimum of six hours for this study and can be easily altered and retrained for shorter windows). The hours of 2 am to 4 am were chosen for this study for the NF period. The extent to which the 2 am to 4 am flow of each LKG grouping (with a sufficient amount of preceding flow data) deviates from the MNF of the preceding three nights is found using Eq. (6). This method was also repeated using mean values, rather than median values, to see if there were significant differences. The number of LKG groupings meeting or exceeding a selection of percentage deviation thresholds are shown in Table 1.

It is observed that the MNF classifies under 50% of the true LKG groupings even at a low 10% percentage deviation threshold and can

provide an accuracy of $\sim 31\%$ for a percentage deviation threshold of 50%. Although using the mean leads to slightly higher accuracies than using the median, the differences are very small. This may be due to a poor representation of typical NF being captured in the limited input data. In contrast, the proposed FLUIDS framework produces forecasts of groups identified as LKG with a median IA value over 0.7 (with a lower quartile of 0.54 and an upper quartile of 0.83) even prior to the addition of residual forecasting. Again, it should be noted that while the MNF only detects the leakages, the proposed FLUIDS framework provides a point-to-point estimate of the future flow time-series. Hence a high mean IA in the forecast indicates a better prediction power which can be easily extended to better classification/detection power given a relevant threshold. Selection of the threshold for the proposed framework can be subjectively decided by the users or tested thoroughly on a case-by-case basis which is beyond the scope of this study. Thus, in addition to being able to flag outliers at any time of the day, the proposed FLUIDS framework is able to accurately forecast flow during periods of outlier flow, even in cases where MNF analysis would not flag the flow data as indicative of a burst.

6. Conclusions

In England and Wales, over 20% of water put into public supply does not make it to the taps of consumers (PR19 final determinations: Securing cost efficiency technical appendix 2019), a percentage that is not particularly unusual in the developed world (Kingdom, Liemberger, and Marin, 2006). The impacts of such high levels of leakage, other than the obvious wastage of water, include energy wastage from the treatment of water lost as leakage, additional costs for operators, negative customer experiences, and disincentivising water-saving behavior in customers (Funding approaches for leakage reduction - report for Ofwat 2019). All of these act as barriers to the development of a truly sustainable water network, able to handle the demands of growing urban population. While detection methods have improved, it remains true that a majority of bursts are reported by customers, rather than detected by water companies (Proactive approach to leaks required to meet tough Ofwat targets 2020). To drive down leakage and build customer trust, a proactive approach is required that can provide rapid and accurate information on leakage at a localised level, facilitating efficient and targeted repair strategies. The proposed hybrid machine learning-based framework (named FLUIDS) is able to forecast flow behavior at the DMA

Table 1
MNF analysis of LKG groupings (1338 total groups).

Percentage deviation	Method for finding average			
	Median No. of LKG groups meeting MNF detection threshold	Percentage of LKG groups meeting MNF detection threshold	Mean No. of LKG groups meeting MNF detection threshold	Percentage of LKG groups meeting MNF detection threshold
+10%	584	43.7	589	44.0
+25%	487	36.4	497	37.1
+50%	414	30.9	417	31.2

level to a high degree of accuracy. Able to forecast both regular (non-leakage) flow and anomalous flow indicative of a burst, the proposed framework can be used to estimate the likelihood that flow in a given DMA will exceed a threshold that can be defined by the user, allowing for intelligent and proactive leakage management.

The data-driven FLUIDS framework is trained and tested on a large dataset containing records for 12-month of flow for over 2000 DMAs in the Yorkshire region. Bursts are first identified using an anomaly detection algorithm and checked against logged repairs; isolation forests is the algorithm chosen for this study. Flow forecasting is performed for both LKG and NLKG data, with an LSTM-RNN giving a mean flow forecast and KF used to provide real-time residual forecasting. The LSTM-RNN alone is able to generate forecasts with a median IA of over 0.7, with the IA further improved by the addition of residual forecasting. Performance is particularly strong for LKG groupings with high $|Z|$ values, indicating more extreme bursts. This demonstrates that the framework is able to accurately forecast both regular (non-leakage) flow data and leakage flows.

Leakage prediction via forecasting of anomalous flow is a relatively unexplored field of study, and it is hoped that the proposed FLUIDS framework will demonstrate the potential of anticipatory leakage management. FLUIDS framework can detect and forecast anomalous flow data even with limited available preceding flow data and regardless of time-of-day, which can improve the proportion of leaks detected first by water companies rather than customers. The balanced distribution of forecasting error in the results of this study also indicates that the FLUIDS framework could be expanded to provide accurate quantification of water loss during leakage. Accurate prediction of leakage can allow time-efficient and cost-efficient preventative maintenance, reducing water loss and customer disruption and taking an important step towards sustainable and systemic smart water management (Oberascher, Rauch, and Sitzenfrie, 2022; McMillan and Varga, 2022).

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.

Data availability

Data will be made available on request.

References

- Aksela, K., Aksela, M., & Vahala, R. (2009). Leakage detection in a real distribution network using a SOM. *Urban Water Journal*, 6(4), 279–289. <https://doi.org/10.1080/15730620802673079>
- Amoatey, P. K., Obiri-Yeboah, A. A., & Akosah-Kusi, M. (2021). Impact of active night population and leakage exponent on leakage estimation in developing countries. *Water Practice and Technology*, 17(1), 14–25. <https://doi.org/10.2166/wpt.2021.124>
- Ávila, C. A. M., Sánchez-Romero, F.-J., López-Jiménez, P. A., & Pérez-Sánchez, M. (2022). Improve leakage management to reach sustainable water supply networks through by green energy systems. Optimized case study. *Sustainable Cities and Society*, 83, Article 103994. <https://doi.org/10.1016/j.scs.2022.103994>
- Barton, N. A., Farewell, T. S., Hallett, S. H., & Acland, T. F. (2019). Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. *Water Research*, 164, Article 114926. <https://doi.org/10.1016/j.watres.2019.114926>
- Bata, M., Carriveau, R., & Ting, D. S.-K. (2020). Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model. *Smart Water*, 5(1), 2. <https://doi.org/10.1186/s40713-020-00020-y>
- Birek, L., Petrovic, D., & Boylan, J. (2014). Water leakage forecasting: The application of a modified fuzzy evolving algorithm. *Applied Soft Computing*, 14, 305–315. <https://doi.org/10.1016/j.asoc.2013.05.021>
- Brentan, B., Meirelles, G., Herrera, M., Luvizotto, E., & Izquierdo, J. (2018). Updating recurrent neural networks for near real-time water demand predictive models. In *International Congress on Environmental Modelling and Software* [Online]. Available <http://scholarsarchive.byu.edu/iemssconference/2018/Stream-B/30>.
- Briers, M., Doucet, A., & Maskell, S. (2009). Smoothing algorithms for state-space models. *Annals of the Institute of Statistical Mathematics*, 62(1), 61. <https://doi.org/10.1007/s10463-009-0236-2>
- Cahuntzi, R., Chen, X., & Güttel, S. (2023). A comparison of LSTM and GRU networks for learning symbolic sequences. *arXiv*. <https://doi.org/10.48550/arXiv.2107.02248>, 04.
- Chan, T. K., Chin, C. S., & Zhong, X. (2018). Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection. *IEEE Access: Practical Innovations, Open Solutions*, 6, 78846–78867. <https://doi.org/10.1109/ACCESS.2018.2885444>
- Chen, G., Long, T., Bai, Y., & Zhang, J. (2019). A forecasting framework based on kalman filter integrated multivariate local polynomial regression: Application to urban water demand. *Neural Process Letter*, 50(1), 497–513. <https://doi.org/10.1007/s11063-019-10001-3>
- Chen T., & Guestrin C. 'XGBoost: A scalable tree boosting system', Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- Cleveland, W. P., & Tiao, G. C. (1976). Decomposition of seasonal time series: A model for the census X-11 program. *Journal of the American Statistical Association*, 71(355), 581–587. <https://doi.org/10.2307/2285586>
- de Oliveira, J. F. L., Silva, E. G., & de Mattos Neto, P. S. G. (2022). A hybrid system based on dynamic selection for time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 33(8), 3251–3263. <https://doi.org/10.1109/TNNLS.2021.3051384>
- Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. OUP Oxford.
- Duveiller, G., Fasbender, D., & Meroni, M. (Jan. 2016). Revisiting the concept of a symmetric index of agreement for continuous datasets. *Scientific reports*, 6(1). <https://doi.org/10.1038/srep19401>. Art. no. 1.
- Funding approaches for leakage reduction - report for Ofwat, PwC, Final Determinations /PR19, Dec. 2019. [Online]. Available: <https://www.ofwat.gov.uk/wp-content/uploads/2019/12/PwC-%E2%80%93Funding-approaches-for-leakage-reduction.pdf>.
- Gao, X., Yang, S., & Hu, Y. (2010). Leakage forecasting for water supply network based on GA-SVM model. In *Proceedings of the 2010 Symposium on Piezoelectricity, Acoustic Waves and Device Applications* (pp. 206–209). <https://doi.org/10.1109/SPAWDA.2010.5744304>
- García, V.J., Cabrera, E., & Cabrera, J. 'The minimum night flow method revisited', pp. 1–18, Apr. 2012, doi: 10.1061/40941(247)35.
- Geberemariam, T., Juran, I., & Shahrou, I. (Jan. 2014). Virtual DMA municipal water supply pipeline leak detection and classification using advance pattern recognizer multi-class SVM. *Journal of Pattern Recognition Research*, 9, 25–42. <https://doi.org/10.13176/11.548>
- Geelen, C. V. C., Yntema, D. R., Molenaar, J., & Keesman, K. J. (2021). Burst detection by water demand now casting based on exogenous sensors. *Water Resources Management*, 35(4), 1183–1196. <https://doi.org/10.1007/s11269-021-02768-9>
- Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451–2471. <https://doi.org/10.1162/089976600300015015>
- Graves, A., & Schmidhuber, J. (2005). Framework phoneme classification with bidirectional LSTM networks. In , 4. *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005* (pp. 2047–2052). <https://doi.org/10.1109/IJCNN.2005.1556215>
- Hamilton, S., & McKenzie, R. S. (2014). *Water management and water loss*. London: IWA Publishing.
- Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781107049994>
- Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 06(02), 107–116. <https://doi.org/10.1142/S0218488598000094>
- Hochreiter, S., & Schmidhuber, J. (1996). LSTM can solve hard long time lag problems. In *Proceedings of the 9th International Conference on Neural Information Processing Systems*, in NIPS '96 (pp. 473–479). Cambridge, MA, USA: MIT Press.
- Huang, P., et al. (2018). Real-Time burst detection in district metering areas in water distribution system based on patterns of water demand with supervised learning. *Water*, 10(12). <https://doi.org/10.3390/w10121765>. Art. no. 12.
- Hutton, C., & Kapelan, Z. (Jan. 2015). Real-time burst detection in water distribution systems using a bayesian demand forecasting methodology. *Procedia Engineering*, 119, 13–18. <https://doi.org/10.1016/j.proeng.2015.08.847>
- Jing, K., & Zhi-Hong, Z. (Jan. 2012). Time prediction model for pipeline leakage based on grey relational analysis. *Physics Procedia*, 25, 2019–2024. <https://doi.org/10.1016/j.phpro.2012.03.344>
- Jung, D., & Lansey, K. (2015). Water distribution system burst detection using a nonlinear kalman filter. *Journal of Water Resources Planning and Management*, 141(5), Article 04014070. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000464](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000464)
- Kang, J., Park, Y.-J., Lee, J., Wang, S.-H., & Eom, D.-S. (2018). Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems. *IEEE Transactions on Industrial Electronics*, 65(5), 4279–4289. <https://doi.org/10.1109/TIE.2017.2764861>
- Kavya, M., Mathew, A., Shekar, P. R., & P, S. (2023). Short term water demand forecast modelling using artificial intelligence for smart water management. *Sustainable Cities and Society*, 95, Article 104610. <https://doi.org/10.1016/j.scs.2023.104610>
- Kiefer, J., & Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(4), 462–466.
- Kingdom, B., Liemberger, R., & Marin, P. (2006). *The challenge of reducing non-revenue water in developing countries-how the private sector can help: A look at performance-based service contracting* Accessed: Feb. 16, 2023. [Online]. Available <https://openknowledge.worldbank.org/handle/10986/17238>.

- Kingma, D. P., & Ba, J. (Jan. 29, 2017). Adam: A method for stochastic optimization. *arXiv*. <https://doi.org/10.48550/arXiv.1412.6980>
- Kovvali, N., Banavar, M., & Spanias, A. (2013). An introduction to kalman filtering with MATLAB examples. *Synthesis Lectures on Signal Processing*, 6(2), 1–81. <https://doi.org/10.2200/S00534ED1V01Y201309SPR012>
- Lee, S. S., Lee, H.-H., & Lee, Y.-J. (Jan. 2022). Prediction of minimum night flow for enhancing leakage detection capabilities in water distribution networks. *Applied Sciences*, 12(13). <https://doi.org/10.3390/app12136467>. Art. no. 13.
- Leu, S.-S., & Bui, Q.-N. (2016). Leak prediction model for water distribution networks created using a bayesian network learning approach. *Water Resour Manage*, 30(8), 2719–2733. <https://doi.org/10.1007/s11269-016-1316-8>
- Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining* (pp. 413–422). <https://doi.org/10.1109/ICDM.2008.17>
- Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data*, 6(1), 3. <https://doi.org/10.1145/2133360.2133363>, 1–3:39.
- Malithong, P., Gulpanich, S., & Suesut, T. 'Water loss control in DMA monitoring system used wireless technology', 제어로봇시스템학회:학술대회논문집, pp. 773–777, 2005.
- Masreliez, C., & Martin, R. (1977). Robust bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Transactions on Automatic Control*, 22(3), 361–371. <https://doi.org/10.1109/TAC.1977.1101538>
- Maybeck, P. S. (1990). The Kalman filter: An introduction to concepts. *Autonomous robot vehicles* (pp. 194–204). New York, NY: Springer. https://doi.org/10.1007/978-1-4613-8997-2_15. I. J. Cox and G. T. Wilfong.
- McMillan, L., & Varga, L. (2022). Towards self-healing in water infrastructure systems. In *Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction* (pp. 1–9). <https://doi.org/10.1680/jsmic.22.00006>
- Menéndez García, L. A., et al. (Jan. 2022). A method of pruning and random replacing of known values for comparing missing data imputation models for incomplete air quality time series. *Applied Sciences*, 12(13). <https://doi.org/10.3390/app12136465>. Art. no. 13.
- Moretti, M., Fiorillo, D., Guercio, R., Giugni, M., De Paola, F., & Sorgenti degli Uberti, G. (2022). A preliminary analysis for water demand time series. *Environmental Sciences Proceedings*, 21(1). <https://doi.org/10.3390/envirosci2022021007>. Art. no. 1.
- Morrison, G. W., & Pike, D. H. (1977). Kalman filtering applied to statistical forecasting. *Management Science*, 23(7), 768–774.
- Morrison, J. (2004). Managing leakage by district metered areas: A practical approach. *Water*, 21, 44–46.
- Mounce, S., Boxall, J., & Machell, J. (Jan. 2007). An artificial neural network/Fuzzy Logic system for DMA flow meter data analysis providing burst identification and size estimation. In *Proc. Water Management Challenges in Global Change* (p. 320).
- Mounce, S. R. (2013). A comparative study of artificial neural network architectures for time series prediction of water distribution system flow data. *Machine learning in water systems - AISB convention 2013* (pp. 5–12). UK: Sheffield: University of Exeter Accessed: Jun. 18, 2022. [Online]. Available <https://eprints.whiterose.ac.uk/83574/>.
- Mounce, S. R., Boxall, J. B., & Machell, J. (2010). Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. *Journal of Water Resources Planning and Management*, 136(3). [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000030](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000030). Art. no. 3.
- Mounce, S. R., & Machell, J. (2006). Burst detection using hydraulic data from water distribution systems with artificial neural networks. *Urban Water Journal*, 3(1), 21–31. <https://doi.org/10.1080/15730620600578538>
- Mounce, S. R., Mounce, R. B., Jackson, T., Austin, J., & Boxall, J. B. (2013). Pattern matching and associative artificial neural networks for water distribution system time series data analysis. *Journal of Hydroinformatics*, 16(3), 617–632. <https://doi.org/10.2166/hydro.2013.057>
- Mushtaq, R. 'Augmented dickey fuller test'. Rochester, NY, Aug. 17, 2011. doi: 10.2139/ssrn.1911068.
- Nelson, M., Hill, T., Remus, W., & O'Connor, M. (1999). Time series forecasting using neural networks: Should the data be deseasonalized first? *Journal of Forecasting*, 18(5), 359–367. [https://doi.org/10.1002/\(SICI\)1099-131X\(199909\)18:5:359:AID-FOR746.3.0.CO;2-P](https://doi.org/10.1002/(SICI)1099-131X(199909)18:5:359:AID-FOR746.3.0.CO;2-P)
- Net Zero 2030 Routemap. (2020). *Water UK* Accessed: Mar. 06, 2023. [Online]. Available <https://www.water.org.uk/routemap2030/wp-content/uploads/2020/11/Water-UK-Net-Zero-2030-Routemap.pdf>.
- Oberascher, M., Rauch, W., & Sitzenfrei, R. (Jan. 2022). Towards a smart water city: A comprehensive review of applications, data requirements, and communication technologies for integrated management. *Sustainable Cities and Society*, 76, Article 103442. <https://doi.org/10.1016/j.scs.2021.103442>
- O'Halloran R.J. & Jarrett, R. 'On-line monitoring of water distribution systems : Data processing and anomaly detection', 2006, Accessed: Jun. 09, 2023. [Online]. Available: <https://publications.csiro.au/rpr/pub?list=BR0&pid=procite:a97053ed-658d-4c7b-8462-2e9ef7f205f>.
- Pandey, P., Bokde, N. D., Dongre, S., & Gupta, R. (2021). Hybrid models for water demand forecasting. *Journal of Water Resources Planning and Management*, 147(2), Article 04020106. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001331](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001331)
- PR19 final determinations: Securing cost efficiency technical appendix. (2019). *Ofwat* Accessed: Jun. 29, 2021. [Online]. Available <https://www.ofwat.gov.uk/wp-content/uploads/2019/12/PR19-final-determinations-Securing-cost-efficiency-technical-appendix.pdf>.
- Proactive approach to leaks required to meet tough Ofwat targets. (2020). *Global Water Intelligence*, 21(7) Accessed: Feb. 16, 2023. [Online]. Available <https://www.globalwaterintel.com/global-water-intelligence-magazine/21/7/smart-water-watch/proactive-approach-to-leaks-required-to-meet-tough-ofwat-targets>.
- Puust, R., Kapelan, Z., Savić, D. A., & Koppel, T. (2010). A review of methods for leakage management in pipe networks. *Urban Water Journal*, 7(1), 25–45. <https://doi.org/10.1080/15730621003610878>
- Romano, M., Kapelan, Z., & Savić, D. A. (2014). Automated detection of pipe bursts and other events in water distribution systems. *Journal of Water Resources Planning and Management*, 140(4), 457–467. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000339](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000339)
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088). <https://doi.org/10.1038/323533a0>. Art. no. 6088.
- Sayed, A., et al. (2021). A novel CMAQ-CNN hybrid model to forecast hourly surface-ozone concentrations 14 days in advance. *Scientific reports*, 11(1). <https://doi.org/10.1038/s41598-021-90446-6>. Art. no. 1.
- Shumway, R. H., & Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the Em algorithm. *Journal of Time Series Analysis*, 3(4), 253–264. <https://doi.org/10.1111/j.1467-9892.1982.tb00349.x>
- Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications: With R examples. *Springer texts in statistics*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-52452-8>
- Skarlatos, K., Bekri, E. S., Georgakellos, D., Economou, P., & Bersimis, S. (Jan. 2023). Projecting annual rainfall timeseries using machine learning techniques. *Energies*, 16(3). <https://doi.org/10.3390/en16031459>. Art. no. 3.
- Spedaletti, S., Rossi, M., Comodi, G., Cioccolanti, L., Salvi, D., & Lorenzetti, M. (2022). Improvement of the energy efficiency in water systems through water losses reduction using the district metered area (DMA) approach. *Sustainable Cities and Society*, 77, Article 103525. <https://doi.org/10.1016/j.scs.2021.103525>
- Tabesh, M., Yekta, A. H. A., & Burrows, R. (2009). An integrated model to evaluate losses in water distribution systems. *Water Resour Manage*, 23(3), 477–492. <https://doi.org/10.1007/s11269-008-9284-2>
- Willmott, C. J., et al. (1985). Statistics for the evaluation and comparison of models. *Journal of Geophysical Research: Oceans*, 90(C5), 8995–9005. <https://doi.org/10.1029/JC090iC05p08995>
- Willmott, C. J. (1981). On the validation of models. *Physical Geography*, 2(2), 184–194. <https://doi.org/10.1080/02723646.1981.10642213>
- Wu, Z. Y., et al. (2023). High fidelity digital twin-based anomaly detection and localization for smart water grid operation management. *Sustainable Cities and Society*, 91, Article 104446. <https://doi.org/10.1016/j.scs.2023.104446>
- Wu, Z. Y., & He, Y. (2021). Time series data decomposition-based anomaly detection and evaluation framework for operational management of smart water grid. *Journal of Water Resources Planning and Management*, 147(9), Article 04021059. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001433](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001433)
- Xue, M., Chew, A. W. Z., Cai, J., Pok, J., Kalfarisi, R., & Wu, Z. Y. (2022). Improving near real-time anomaly event detection and classification with trend change detection for smart water grid operation management. *Urban Water Journal*, 19(6), 547–557. <https://doi.org/10.1080/1573062X.2022.2058565>
- Ye, G., & Fenner, R. A. (2011). Kalman filtering of hydraulic measurements for burst detection in water distribution systems. *Journal of Pipeline Systems Engineering and Practice*, 2(1), 14–22. [https://doi.org/10.1061/\(ASCE\)PS.1949-1204.0000070](https://doi.org/10.1061/(ASCE)PS.1949-1204.0000070)
- Zanfei, A., Menapace, A., Brentan, B. M., Righetti, M., & Herrera, M. (2022). Novel approach for burst detection in water distribution systems based on graph neural networks. *Sustainable Cities and Society*, 86, Article 104090. <https://doi.org/10.1016/j.scs.2022.104090>
- Zanfei, A., Menapace, A., & Righetti, M. (Jan. 2023). An artificial intelligence approach for managing water demand in water supply systems. In *1136. IOP Conference Series: Earth and Environmental Science*, Article 012004. <https://doi.org/10.1088/1755-1315/1136/1/012004>