Environmental Science and Pollution Research (2024) 33:1992–2006
https://doi.org/10.1007/s11356-024-32355-5

SUSTAINABLE TECHNOLOGIES FOR ENVIRONMENTAL HEALTH AND SAFETY

Screening of representative rainfall event series for long-term hydrological performance evaluation of grassed swales

Feikai Yang[1,2,3] · Dafang Fu[1,2] · Chris Zevenbergen[3] · Floris C. Boogaard[4,5] · Rajendra Prasad Singh[1,2]

Received: 27 October 2023 / Accepted: 2 February 2024 / Published online: 23 February 2024

Abstract

Evaluation of the hydrological performance of grassed swales usually needs long-term monitoring data. At present, suitable techniques for simulating the hydrological performance using limited monitoring data are not available. Therefore, current study aims to investigate the relationship between saturated hydraulic conductivity (K_s) fitting results and rainfall characteristics of various events series length. Data from a full-scale grassed swale (Enschede, the Netherlands) were utilized as long-term rainfall event series length (95 rainfall events) on the fitting outcomes. Short-term rainfall event series were extracted from these long-term series and used as input in fitting into a multivariate nonlinear model between K_s and its influencing rainfall indicators (antecedent dry days, temperature, rainfall, rainfall duration, total rainfall, and seasonal factor (spring, summer, autumn, and winter, herein refer as 1, 2, 3, and 4). Comparison of short-term and long-term rainfall event series fitting results allowed to obtain a representative short-term series that leads to similar results with those using long-term series. A cluster analysis was conducted based on the fitting results of the representative rainfall event series with their rainfall event characteristics using average values of influencing rainfall indicators. The seasonal index (average value of seasonal factors) was found to be the most representative short rainfall event series indicator. Furthermore, a Bayesian network was proposed in the current study to predict if a given short-term rainfall event series is representative. It was validated by a data series (58 rainfall events) from another full-scale grassed swale located in Utrecht, the Netherlands. Results revealed that it is quite promising and useful to evaluate the representativeness of short-term rainfall event series used for long-term hydrological performance evaluation of grassed swales.

Keywords Long-term hydrological performance · Grassed swale · Saturated hydraulic conductivity · Rainfall event series · Bayesian network · Representativeness

Introduction

Grassed swales are vegetated shallow ditches, generally suitable along the sides of urban roads, impermeable sites, parks, and green spaces (Lu et al. 2023). They can be coupled with or even replace the traditional urban stormwater drainage system. Stormwater runoff which flows through a grassed swale could be reduced and purified through infiltration, settlement, and filtration of soil media and vegetation (Zhao et al. 2016; Gong et al. 2019; Saudo-Fontaneda et al. 2020). The first application of grassed swales dates back to the 1940s. These facilities were originally designed for the prevention of soil erosion (Fardel et al. 2019). The USA has established the stormwater management function of BMP (best management practices) measures including grassed swales through the federal Clean Water Act in 1972 (Board 2009). Grassed swales particularly exhibit effective

Responsible Editor: Philippe Garrigues

Highlights
- Bayesian network-based method was built for long-term monitoring and evaluation of grassed swales.
- Saturated hydraulic conductivity (K_s) of a grassed swale was correlated with rainfall event indicators.
- Nonlinear fitting results of K_s between short and long rainfall event series were compared to screen representative series.
- The minimum representative rainfall event series length was 60 events.
- Most representative rainfall event indicator was the seasonal index.

Extended author information available on the last page of the article

runoff reduction in response to small rainfall events (Saudo-Fontaneda et al. 2020; Shafique et al. 2018). They can reduce peak flow by 4 to 87% (Deletic and Fletcher 2006; Rujner et al. 2018) and total runoff by 15 to 82% (Winston et al. 2018). Significant individual differences in the hydrological performance of grassed swales have been reported in previous studies. These differences may be caused by factors such as the initial soil moisture content (Rujner et al. 2018), soil characteristics (Davis et al. 2012; Rujner et al. 2016), surface roughness, height and density of the grass cover (Deletic and Fletcher 2006), hydraulic conductivity (Besir and Cuce 2018; Rujner et al. 2018), compaction during construction (Gregory et al. 2006; Pitt et al. 2008), and maintenance during the long-term operation (Saudo-Fontaneda et al. 2020).

The saturated hydraulic conductivity of the filter medium is the main factor which controls the long-term hydrological performance of grassed swales. In absence of maintenance, they are prone to clog by particles and sediment during long-term operation (Xie et al. 2020). Previous work revealed that most rainwater pollutants are trapped in the upper soil layer (Dierkes and Geiger 1999). There are only few studies focused on the time-varying characteristic of the saturated hydraulic conductivity of grassed swales and other BMP facilities, such as bioretention cells and green roof (Wang et al. 2023). Coustumer et al. (2009) conducted on-site experiments on 37 biofilter systems in Australia to investigate the changes in the saturated hydraulic conductivity after 0.5 to 3 years of operation. Results revealed that the 40% of the biofilter systems had low saturated hydraulic conductivity (<50 mm/h), 43% had medium (50 to 200 mm/h), and 17% had high (>200 mm/h). Some biofilters with high initial saturated hydraulic conductivity (> 200 mm/h) showed significant reductions after prolonged operation. However, facilities with lower initial saturated hydraulic conductivity (< 20 mm/h) were less affected over prolonged operation. Haile et al. (2016) noted that the saturated hydraulic conductivity of biofilter media significantly decreased within 5 to 7 years of operation but remained within an acceptable range (212.4–504 mm/h). A bioretention system in Blacksburg, USA, exhibited good runoff reduction and infiltration capacity after 7 years of operation (Willard et al. 2017). Paus et al. (2014) explored the long-term performance of the infiltration function of bioretention systems and found that their infiltration capacity were sustained at a constant level for more than 6 years. These studies revealed no significant change in the long-term infiltration performance of bioretention systems. It may attribute to the maintenance activities and environmental variations, or it is hard to detect substantial alterations in shorter monitoring period. A grassed swale, with prolonged operation (9 to 14 years), entirely clogged due lack of proper maintenance measures (Al-Rubaei et al. 2015). Soil or media clogging is the key limiting factor for the service life of grassed swales and other similar BMPs (Kandra et al. 2014). Soil and vegetation also have a significant impact on the peak flow reduction and drainage performance of grassed swales (Saracoglu and Kazezyilmaz-Alhan 2023). On the one hand, there is a strong correlation between the presence of plants and the clogging process of the media, because vegetation reduces the runoff flow rate, leading to an accumulation of solid particles within the media. And on the other hand, during the long-term operation, dense grassland or vegetation coverage can improve the infiltration capacity of soil media (Hunt et al., 2012; Yousef et al. 1987). The growth of vegetation roots will form large pores and root channels in the media, thereby enhancing permeability and preventing clogging (Muerdter et al. 2018).

Previous study revealed that cold climates and repeated snowmelt processes can also affect the infiltration capacity of the filter medium in a long-term operational grassed swale (He et al. 2022). The snow could also affect the performance of grassed swales by concentrating the flow in narrow channels, which reduced the effective area of infiltration, but also led to the longer lag times and stored a portion of the runoff water within its pack (Zaqout and Andradóttir 2021). With the addition of snowmelt water, it is highly significant to quantify the combined effects of frozen soils, snowmelt, and rainfall on runoff from urban catchments during the melting period (Moghadas et al. 2018). Compared with other BMPs and urban green space, grassed swales are more likely to be affected by the snow melting process because grassed swales are usually implemented on the roadsides and become a road snow accumulation or deposit areas (Zaqout and Andradóttir 2021). When experiencing rain on snow events, an ice layer that prevents infiltration may form within the snow cover, which helps to generate instantaneous runoff during snow rainfall and snowmelt events (Garvelmann et al. 2015). While suffering from repeated cycles of freeze–thaw events, the infiltration rate could increase or decrease mainly due to the soil texture and initial water content, and ice may have formed in soil pores and causing lower infiltration rates after frequent freeze–thaw cycles (Fouli et al. 2013). However, repeated freezing and thawing processes could also change the stability of the media structure and the continuity of pores, thus producing preferential flow that may promote infiltration (Paus et al. 2016).

Long-term hydrological monitoring data is crucial for evaluating the long-term hydrological performance of grassed swales, clarifying the time-varying characteristics of their saturated hydraulic conductivity and the influencing factors that cause the observed changes. However, the optimal duration of the monitoring period and the number of rainfall events required to effectively evaluate or predict the long-term hydrological performance of grassed swales are still difficult to define (Yang et al. 2023). Brown et al.

(2013) employed DRAINMOD model to simulate the performance of four bioretention cells using 2-year monitoring data. Abualfaraj et al. (2018) collected monitoring data from a green roof over a period of 4 years and compared it with the predicted results of the EPA-SWMM model. Brown et al. (2013) used 1-year monitoring data to calibrate and validate the DRAINMOD model. A water balance approach, termed the Soil Water Apportioning Method (SWAM), to enable economic assessment of the long-term hydrologic performance of green roofs, was developed by Hakimdavar et al. (2016). It is utilized for the estimation of green roof runoff and evapotranspiration (ET) based solely on the measurements of local precipitation, substrate moisture, and the substrate maximum water storage capacity. Information on the assessment of the representativeness of existing monitoring data and the enhancement of subsequent monitoring strategies for evaluating and predicting the long-term hydrological performance of grassed swales and similar BMPs are lacking. Hence, it is crucial to develop a data-driven method for monitoring work.

Several relevant studies have investigated the appropriateness of employing various lengths of time series in rainwater harvesting research. For example, Mitchell (2010) assessed the efficiency of a rainwater harvesting system using rainfall data spanning 1, 10, and 50 years for simulation. The findings revealed that using 10 and 50 years of data yielded comparable results, while a 1-year time series exhibited significant variability. Geraldi and Ghisi (2017) concluded that a 10-year time series is long enough to lead to results similar to those obtained using a 30-year time series rainfall data. Geraldi and Ghisi (2018) evaluated the possibility of using different short-term time series in 13 cities around the world. The results showed that the representativeness was highly dependent on the rainfall characteristics, and a 15-year time series was adequate to lead to equivalent simulation results.

The statistical approach to identify the time series representativeness in research of rainwater harvesting can be used for the monitoring and evaluation of the long-term hydrological performance of grassed swales. This must be investigated by adopting a probabilistic perspective in the realm of statistics, outcomes can be represented with a degree of probability to indicate their validity (Geraldi and Ghisi 2019). Therefore, to forecast the forthcoming conditions as desired in long-term hydrological performance simulations, a Bayesian network can be used, which applies the Bayes theory through the relationship between nodes that are conditionally dependent. Bayesian networks are widely used in prediction, inference, diagnosis, decision risk, and reliability analysis (Borsuk et al. 2004; Geraldi and Ghisi 2019; Liu 2020; Li 2020). Bayesian networks can easily display the dependency relationship between the variables in the form of graphs and tables, which is easier for users and decision-makers to understand and utilize. Meanwhile, there might be a series of missing data due to some reason not being recorded; Bayesian networks can utilize information from other data to supplement missing data for improving the accuracy and reliability of the analysis. In addition, Bayesian networks can also remove the noise and errors from the collected data by establishing probability models for obtaining more accurate and reliable results (Marcot and Penman 2019). Therefore, the current study aims to evaluate the application of short event series instead of long event series for simulations of the saturated hydraulic conductivity of grassed swales using a Bayesian network. This study also aimed to investigate the relationship between the saturated hydraulic conductivity and the event (temperature, rainfall, and seasonal) characteristics of various event series lengths used in the simulation.

Material and methods

Study sites

Two full-scale grassed swales were selected as study sites in the current study. These grassed swales are located in Enschede and Utrecht in the Netherlands. The Enschede study site is the oldest large-scale grass swale and located in the eastern part of the Netherlands (further details available at https://www.climatescan.org/projects/211/detail). The monitoring period was divided into two stages, May 1999 to May 2002 and November 2020 to October 2021. The monitored hydrological variables included antecedent dry days, temperature, rainfall, duration of rainfall, total rainfall (the accumulated rainfall between first rainfall after construction and target rainfall), seasonal factors (spring, summer, autumn, and winter, herein refer as 1, 2, 3, and 4), and saturated hydraulic conductivity (K_s) of filter medium in grassed swales under a single rainfall event. A small rain station was installed on a building close to the swale for rainfall measurements. Two divers were used to measure the atmosphere and the surface water level in the swale (Fig. 1). The second swale (Utrecht) is located in the central part of the Netherlands (further details available at https://www.climatescan.org/projects/2523/detail). The monitoring period was from November 2020 to October 2021, and monitoring parameters were the same as the swale in Enschede. The rainfall data was obtained from city government, the inflow and outflow were measured through triangular weir, and a diver was used to measure the weir head (Fig. 1). The saturated hydraulic conductivity (K_s) of both sites was calculated by Ensemble Kalman filter method using the data recorded, and this work was conducted in our previous study (Yang et al. 2023).

Fig. 1 Location and monitoring equipment of two target grassed swales in the Netherlands

In this study, data from Enschede swale were used for statistical analysis and building the Bayesian network. Data from Utrecht swale were used for validation of the field application. The specific data are listed in Table S1 and Table S2.

Classification and characterization of the event series

Data collected from 95 events at the Enschede grassed swale were selected as reference for the long event series. Each

event was described by the six characteristic factors. These factors include saturated hydraulic conductivity, antecedent dry days, temperature, rainfall intensity, duration of rainfall, total rainfall, and seasonal factors. A dataset of 95 events was randomly sampled to generate a set of short event series with field counts of 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, and 90, and each contained 1000 series combinations. Jennings et al. (2010) proposed indicators to describe the magnitude (annual average rainfall), variability (rainfall standard deviation, and seasonal factors), and behavior of rainfall (rainfall duration, antecedent dry days before rainfall, and annual average dry days). These indicators were characterized as monthly, daily, and event average to describe the magnitude, variability, and behavior of the rainfall events (Istchuk and Ghisi 2022). Based on these, six indicators (average antecedent dry days, average temperature, average rainfall, average rainfall duration, average total rainfall, and seasonal index) were derived to characterize each event series according to the following formula:

$$AVG = \frac{\sum_{i=1}^{n}(R_i)}{n} \tag{1}$$

where AVG is the average value of characteristic factors from the events series, R_i is the value of characteristic factors from a single event, and n is the number of events in a series set.

The indicators were classified by establishing three categories: low, medium, and high. These categories discretized the event characteristics by clustering the event series according to the value of the indicators. The calculation of the category range considered the amplitude (maximum value minus minimum value) of the indicators observed in all event series. The overall minimum value was the lower limit of the "low" range, and the overall maximum value was the upper limit of the "high" range. The amplitude of the indicator was divided by three. The value obtained was added to the minimum value to obtain the upper limit of "low" and was subtracted from the maximum value to obtain the lower limit of "high." The range from the upper limit of "low" to the lower limit of "high" was the "medium" range.

Classification of representative event series

The short event series were divided into representative and non-representative, with a set of 95 events as the reference series. All event series shorter than 95 were considered short event series. This study classified representative and non-representative event series by comparing the similarity of the results of multivariate nonlinear fitting between short and long event series. In our previous study (Yang et al. 2023), K_s was considered as the dependent variable, and other characteristic factors were the independent variables. A quadratic function was used for multivariate nonlinear fitting. The optimization fitting formula of short event series was compared with the result from reference event series to determine their representativeness. The similarity index (Geraldi and Ghisi 2018) was used to calculate the similarity between short and long event series fitting results. It is commonly used to compare the differences in simulation results.

$$Sim = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{K_{sL}^{\,i} - K_{ss}^{\,i}}{K_{sL}^{\,i}}\right) \tag{2}$$

where Sim is the similarity index between short and long events series fitting results, $K_{sL}^{\,i}$ is the prediction K_s value of i^{th} event by the fitting formula obtained from reference event series, $K_{ss}^{\,i}$ is the prediction K_s value of i^{th} event by the fitting formula obtained from short event series, n is the event series length, and the value is 95 in this paper.

Geraldi and Ghisi (2017) noted that when the absolute value of the similarity index was less than 5%, the short event series can obtain results similar to those obtained using long event series, defined as a representative series. Meanwhile, there were many combinations from the same event series length. Therefore, when 90% of the event series were considered similar, this event series length was considered to provide the same result as the reference.

Statistics and analysis of characteristic indicators

Firstly, the statistical analysis used a simple sensitivity analysis (expected value versus predicted value graph) to evaluate the impact of event series length. The expected value was the indicator value of the reference series, represented as a straight line. Scatter point was the predicted value of the short event series indicator, distributed on both sides of the straight line. The root means square error (RMSE) and the mean absolute relative error (MARE) were calculated to compare which indicator has greater impact on the representativeness of short event series.

Secondly, representative event indicators were classified according to the "low," "medium," and "high" and was compared to see which category of each indicator has the higher representativeness.

$$RMSE = \sqrt{\frac{1}{n}\left(\sum_{i=1}^{n}\left(\theta_{pre,i} - \theta_{exp,i}\right)^2\right)} \tag{3}$$

$$MARE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\theta_{pre,i} - \theta_{exp,i}\right|}{\theta_{exp,i}} \tag{4}$$

where $\theta_{pre,i}$ is the predicted value of indicators, $\theta_{exp,i}$ is the expected of indicators, and n is the series length.

Establishment and validation of the Bayesian network

A Bayesian network was built to assist the decision-making. This network used the frequencies of representative series obtained through the statistical analysis to evaluate the representativeness of new short event series. The construction of Bayesian network was based on the Bayesian theory and the calculation of conditional probability.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{5}$$

where $P(A|B)$ is the posterior conditional probability of event A under the condition of event B, $P(B|A)$ is the posterior conditional probability of event B under the condition of event A, and $P(A)$ and $P(B)$ are the probability a priori of event A and B. Detailed approach and examples of Bayesian theory application are available in the literature (Borsuk et al. 2004; Hinesf and Landis 2014; Xiao et al. 2023).

The construction of the Bayesian network was performed in two stages: training and validation. The structure of the network had been configured using a Bernoulli Naive Bayes type and was configured by the input nodes that conditions the output node. The input nodes were the indicators of the event series (1–8), including the number of events, the number of months (the total number of months an event series covered), average antecedent dry days, average temperature, average rainfall, average rainfall duration, average total rainfall, and seasonal index. The output node (9) was the representativeness probability, classified as "Yes" or "No." Each node was discretized (in class) to calculate the conditional probability. The category of node 1 was defined as the number of events (5, 10,…, 90). Node 2 to 8 was divided into three categories: "low," "medium," and "high." Node 9 was classified as "yes" or "no." The representative event series and series characteristics were used to calculate the prior probability.

In the training phase, the prior probability was put into the node, and the network was set done. In the validation stage, the Bayesian network was submitted to a new dataset with known results, and the predicted results of the network were compared with the actual results. In this case, a bootstrap verification method was used, which involves resampling the dataset used for training. During the validation process, various samples of the short event series were put into the network (by setting the characteristics of the short event series on the input nodes), and the output values were recorded. If the probability of the output node predicting a "yes" category was greater than 50%, then the analyzed event series was considered representative by the network. It was compared with the known results, and a performance table (or confusion matrix) was made, which summarized all case results of the validation test. These cases include true-positives (TP, network predicted correct when the series was representative), false-negatives (FN, network predicted false when the series was not representative), true-negatives (TN, network predicted correct when the series is non-representative), and false-positives (FP, network predicted wrong when the series was representative). The performance indexes were determined from the performance table as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{6}$$

$$Error\ Rate = \frac{FN + FP}{TP + FN + FP + TN} \tag{7}$$

$$\text{True positive rate}(TPR\ or\ Sensitivity) = \frac{TP}{TP + FN} \tag{8}$$

$$\text{True negetive rate}(TNR\ or\ Specificity) = \frac{TN}{FP + TN} \tag{9}$$

The final performance analysis was performed using the receiver operating characteristic curve (ROC), which showed the sensitivity versus 1-specificity ratio under different thresholds. By analyzing ROC curve, it could be verified whether the results obtained from the network were from the trend or randomly generated.

Results and discussion

Classification of event series

The event series were characterized and classified using six indicators: average antecedent dry days, average temperature, average rainfall, average rainfall duration, average total rainfall, and seasonal index, as shown in Fig. 2. For each indicator, the scatter plot showed the difference between the short event series and the reference event series. The short event series were generated from the reference event series, and Fig. 2 shows that the short event series indicator converge into the reference series. In fact, in longer event series, more information on event indicators required to make event series more representative. Table 1 shows the classification of each indicator.

Classification of representative event series

The results obtained by short event series with reference series were compared, and short event series were divided into representative and non-representative event series. Figure 3 reveals a high similarity index and 95% confidence intervals at the initial stage and then converged into the

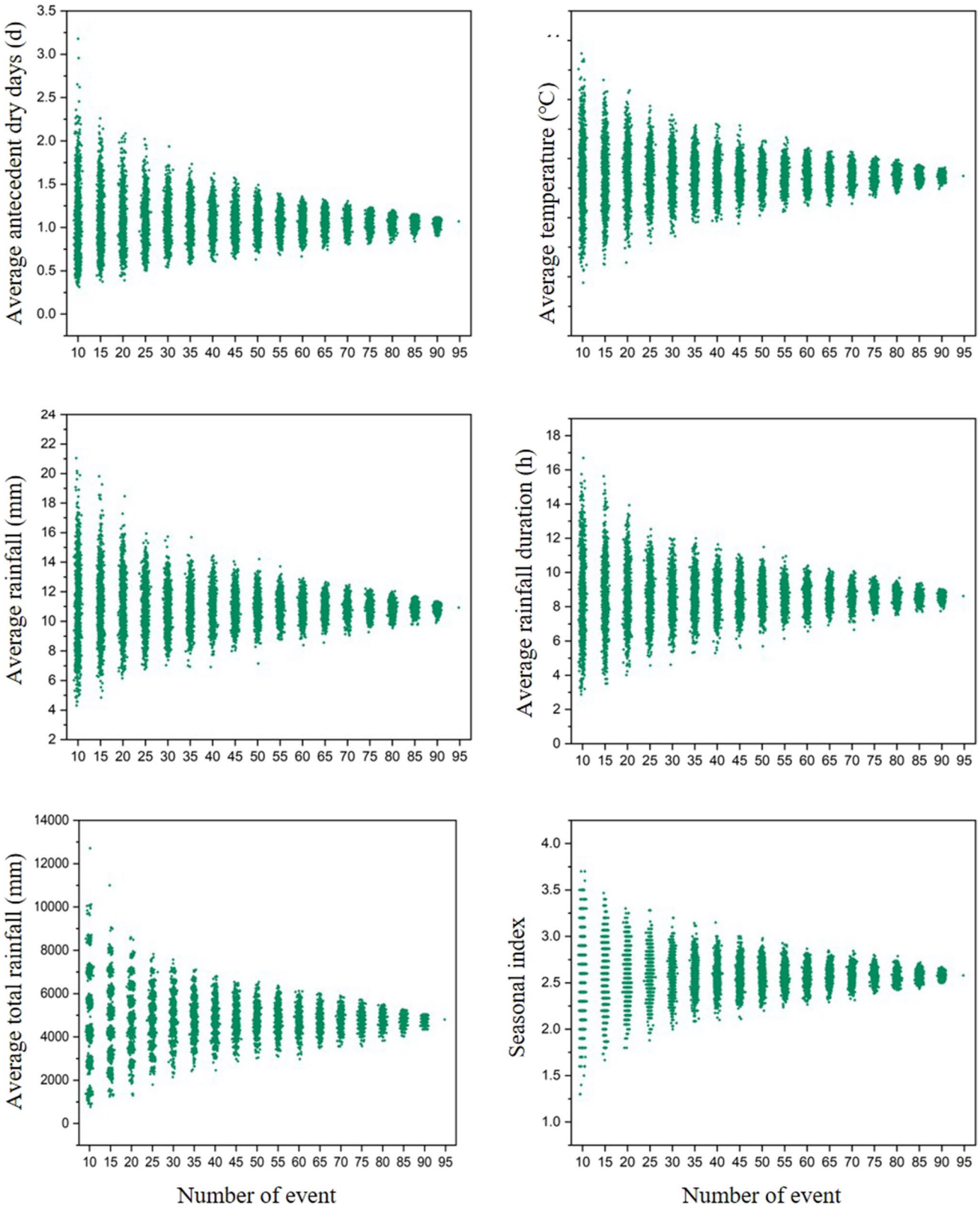

Fig. 2 Distribution of characteristic indicators under different event series length

Table 1 Classification of characteristic indicators of event series

Indicators	Category		
	Low	Medium	High
Average antecedent dry days (d)	0.3–1.3	1.3–2.3	2.3–3.3
Average temperature (℃)	5–8	8–11	11–14
Average rainfall (mm)	4–9.8	9.8–15.6	15.6–21.4
Average rainfall duration (h)	2.5–.5	7.5–12.5	12.5–17.5
Average total rainfall (mm)	600–4700	4700–8800	8800–12,900
Seasonal index	1–2	2–3	3–4

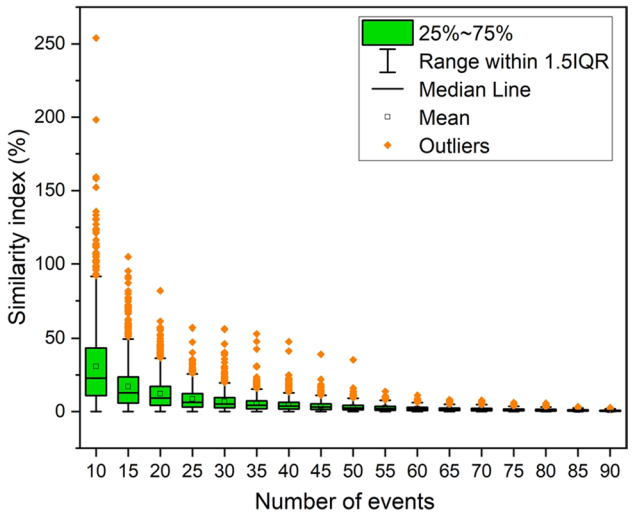

Fig. 3 Comparison of similarity in fitting results between short event series and reference series

reference series. The similarity index and 95% confidence interval were decreased substantially with the increase in the number of events. The average similarity index dropped below 10% when the event number exceeded above 20. When the series length exceeded 40 events, the average similarity index was below 5%. In addition, after the series length lower 20 events, the time span was more than a single year. Therefore, one possible explanation could be that the event series within 1 year were different in terms of event characteristic indicators, while more similar information was obtained after 1 year.

The similarity index calculated for different short event series were classified according to < 5%, 5 to 10%, 10 to 15%, 15 to 20%, and > 20%, as shown in Fig. 3. As referred in the "Classification of representative event series" section, when the similarity index was less than 5%, it was considered a representative short event series. Figure 4 shows that the proportion of representative events corresponding to each series length were 11.5%, 22.6%, 29.7%, 39%, 49.5%, 57.9%, 63.1%, 72.8%, 80.8%, 87.9%, 94.2%, 97.7%, 99.1%,

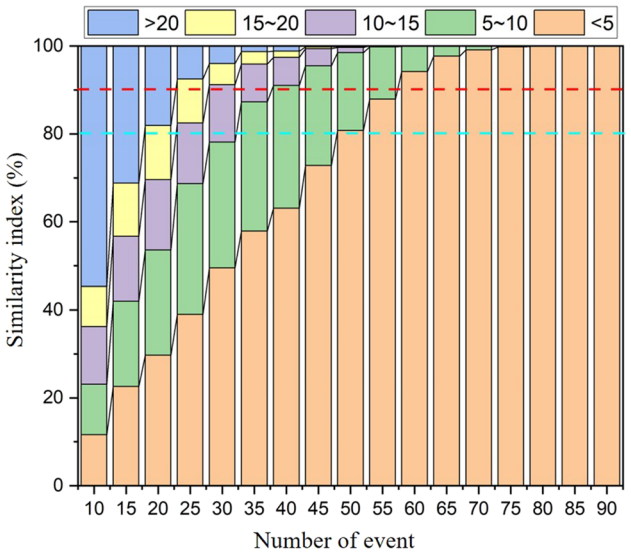

Fig. 4 Proportional distribution of similarity index for different short event series

99.8%, 99.9%, 100%, and 100%, respectively. Therefore, the minimum representative series length was 60 events. When the number of event events exceeded 70, the proportion of representative events exceeded 99%. If the proportion of non-representative events was lower than 20%, the minimum representative series length reduced to 50 events. Similarly, decision-makers could also choose the evaluation criteria for representative events based on specific needs. If the similarity index requirement was lower than 10%, the minimum representative series length corresponding to a non-representative proportion of 10% was 40 events.

The original dataset was resampled to generate a new dataset with a series length of 60 events in 500 sets, which was different from the previous 1000 sets. The fitting results of these 500-event series with the reference series were compared. Results showed that 470 series had a lower than 5% similarity index, while the similarity index of the other 30 series ranged between 5 and 10%. The proportion of non-representative event series was 6%, which was less than the threshold of 10%, proved that 60 was the representative short event series length. It is also verified that this classification method was practical and effective.

Statistical analysis of characteristic indicators

When the value of characteristic indicators in a representative short event series (predicted value) was closer to the reference series (expected value), the indicator was more representative. The RMSE and MARE of these two values were calculated, as shown in Table 2. When only compared the RMSE value, the seasonal index was the lowest, followed by the average antecedent dry days. When only compared

Table 2 Sensitivity analysis of characteristic indicators

Indicators	RMSE	MARE
Average antecedent dry days	0.146	9.425
Average temperature	0.434	3.148
Average rainfall	0.914	5.741
Average rainfall duration	0.789	6.245
Average total rainfall	612.921	8.872
Seasonal index	0.127	3.371

the MARE value, the average temperature was the lowest, followed by the seasonal index. Overall, the seasonal index performed well under both statistical methods, indicated that it was a representative short event series indicator that was more similar to the reference series. This may be due to the obvious seasonal change characteristic of grassed swale saturated hydraulic conductivity. The trend in changes within the year was basically consistent while representative event series were generally longer than 1 year. Therefore, this indicator was closer to the reference series.

Previous studies revealed that plants had obvious seasonal growth characteristics, and the rapid growth in summer could improve the infiltration capacity of fillers (Dagenais et al. 2018; Fort et al. 2012). The cold climate and low temperature also had a significant impact on infiltration but the degree of impacts varied. Soil freezing could hinder runoff infiltration but also generated preferential flow to enhance infiltration (Lefevre et al. 2009; Muthanna et al. 2007; Paus et al. 2016). The influence of antecedent dry days may be related to the evaporation process which helped to restore the infiltration capacity during drought periods (Boogaard 2022; Deletic 2000). Evaporation was also closely related to temperature and plants (Ebrahimian et al. 2019). Therefore, evaporation needs to be monitored in the future research which could be a significant representative factor.

Figures 5 shows the frequency distribution of representative series indicators under different category. For the average antecedent dry days, events in the "low" category had a higher probability of becoming representative, followed by the "medium" category, and the "high" category was the smallest. For average temperature, events in the "medium" category had a higher probability of becoming representative, followed by events in the "low" and "high" category. Average rainfall, average rainfall duration, and average total rainfall were three rainfall characteristic indicators; their frequency distribution also showed similar characteristics. Events in the "medium" category had a higher probability of becoming representative, followed by the "low" and the "high" category. For seasonal index, representative events were mainly distributed in the "medium" category, followed by the "high" and the "low" category. There was no trend of representative frequency increasing or decreasing with

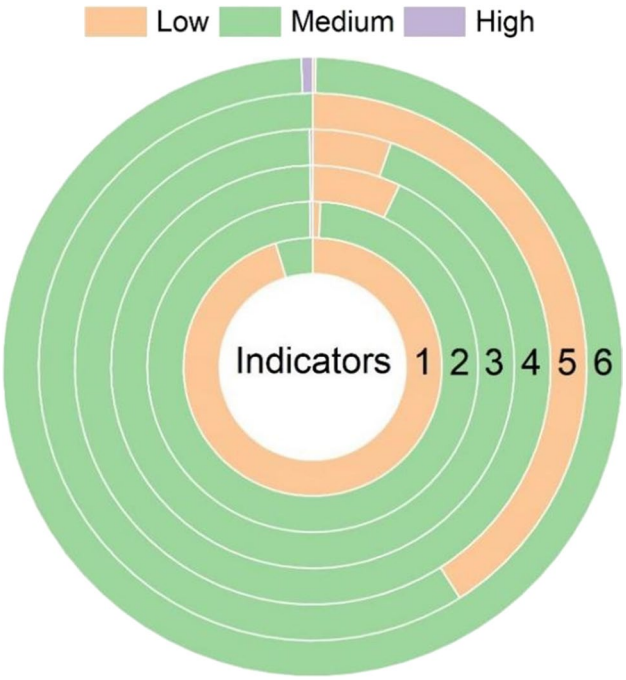

Fig. 5 Proportional distribution of representative events under different category of indicators (where 1 ~ 6 are average antecedent dry days, average temperature, average rainfall, average rainfall duration, average total rainfall, and seasonal index)

the change of category. This might be explained by the distribution of representative events, and the RMSE value corresponded to the category of each indicator presented in Table 3. The RMSE (0.119) of the "low" category of average antecedent dry days was smaller than that of the "medium" category (0.152), and the proportion of representative events corresponding to the "low" category was higher than that of the "medium" category. The RMSE (0.387) of the average temperature in the "medium" category was smaller than that in the "low" category (0.437) and the "high" category (0.427). The RMSE (410.686) of the "medium" category of average total rainfall was smaller than that of the "low" category, and the proportion of representative events corresponded to the "medium" category was higher than that of the "low" category. This trend indicated that the higher

Table 3 RMSE value of indicators' category

Indicators	Low	Medium	High
Average antecedent dry days	0.119	0.152	—
Average temperature	0.437	0.387	0.427
Average rainfall	0.576	0.741	1.136
Average rainfall duration	0.503	0.649	1.066
Average total rainfall	470.742	410.686	—
Seasonal index	0.203	0.116	0.109

the representative frequency occurred when the distribution of category values was uniform. However, for average rainfall, average rainfall duration, and seasonal index indicators, the frequency of representative events in the corresponding category was the low in case of larger RMSE. Conversely, frequency of the representative events may not necessarily be the highest in case of the smaller RMSE. This trend indicated that the uniformity of the distribution of category values was an essential but not enough condition for the frequency of event representativeness. It might be due to insufficient data and not included all features, which requires further research.

Establishment and validation of Bayesian network

Based on the statistical analysis, a Bayesian network was established to evaluate the representativeness of a given short event series. This network could serve as an auxiliary decision-making tool for researchers or engineers to evaluate whether short event series could provide similar results compared to long event series. The probabilities used to build this network are presented in Tables S3 and S4.

Bayesian network uses a white box concept, which allows the repeated experiments and its application to other scenarios. The probability used in this study was related to the dataset, but it could be improved by adding more event samples with new features. Figure 6 shows the Bayesian network structure and prior probability of each node. These categories were obtained from the discretization of the basic dataset, so that each event series indicators could be divided into three categories: high, medium, and low. These probabilities were represented as the "arc force" between a given node and its conditional node, indicating how much a category affects the output. In this network, no category has a greater influence than others, and the output result was a combination of probabilities of all categories, rather than a result determined by a single category.

Bayesian network could be used by adding new short event series. The characteristic indicator values of each event series were classified into three categories: high, medium,

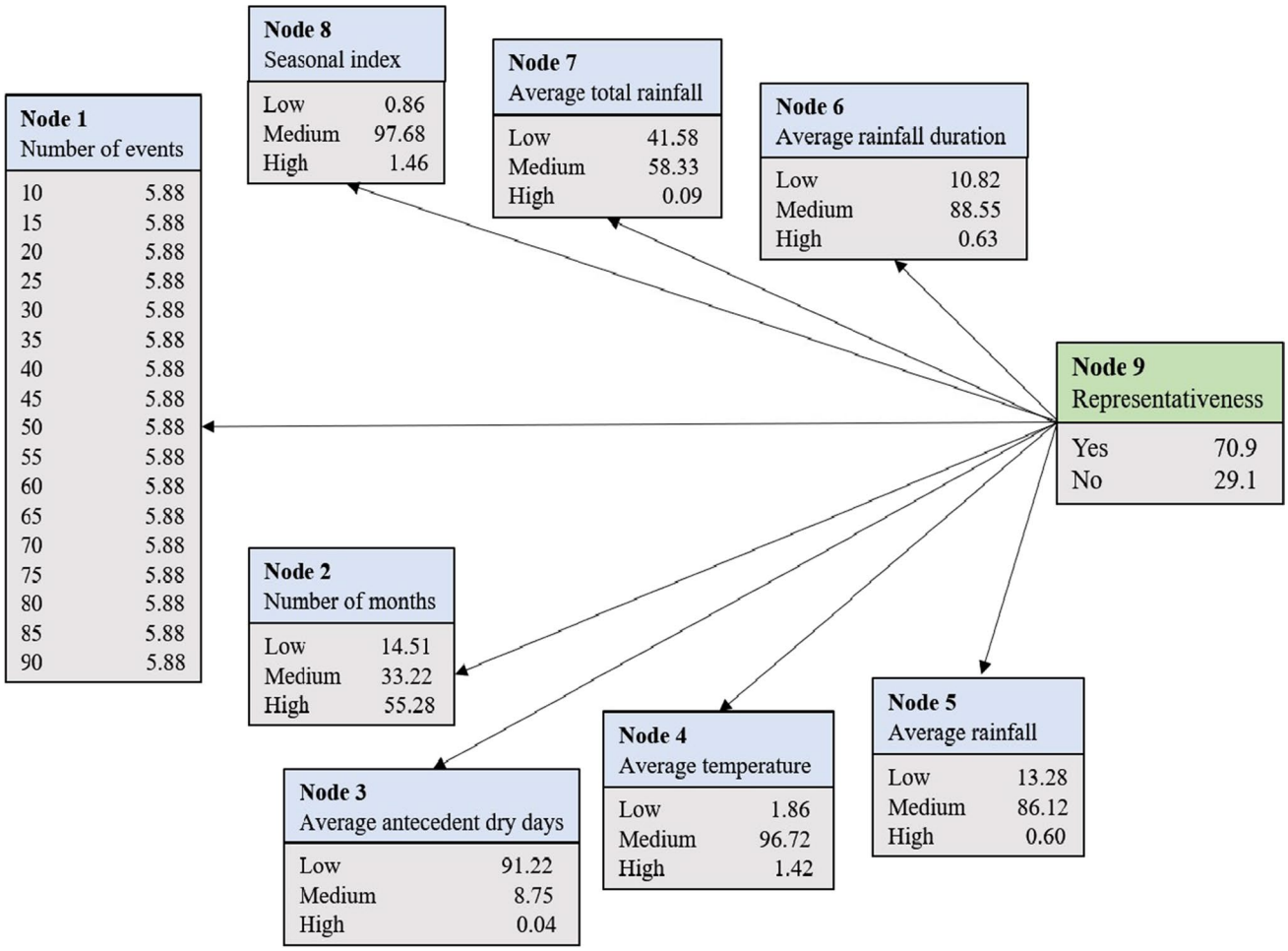

Fig. 6 Bayesian network for evaluation of long-term performance of grassed swales

Table 4 Matrix of confusion—Bayesian network performance

Experimental results	Predicted results		Performance indicators
	Yes	No	
Yes	2528 (TP)	499 (FN)	Accuracy—80.9%
			Precision—89.0%
No	313 (FP)	910 (TN)	Sensitivity—83.5%
			Specificity—74.4%

and low. Then, the nodes 1–8 were set based on the categories of indicators. The output node (Node 9) would predict the representativeness of the estimated event series, given by the percentage of "yes" (0–1).

For example, if there was a short series with 50 events (node 1), and node 2–8 were classified as "medium," "low," "medium," "medium," "medium," "low," and "medium," respectively. Then, the result returned by the Bayesian network for this series was "yes," with the probability of representativeness being 81.8%, and the probability of non-representativeness being 18.2%.

Performance analysis was also conducted by comparing the predicted results (output of Bayesian network) with the results obtained in experiments (representative and non-representative short event series). In this study, the data used for training was resampled to obtain 4250 short events series as the validation dataset. Table 4 shows the calculation results of performance indicators of Bayesian network. The accuracy of Bayesian network was 80.9% and the precision was 89%, the sensitivity was 83.5% and the specificity was 74.4%. There was no threshold value in previous research that determines the significance of the Bayesian network, but it was suggested that values greater than 80% were desirable for any performance measures (Geraldi and Ghisi 2019). Based on the results, this network showed strong robustness (high sensitivity) in predicting the representativeness event series when they were indeed representative and performed no robust (low specificity) in determining non-representative event series when they were indeed non-representative.

The final validation results are presented in Fig. 7 that illustrated the ROC curve of TPR (Sensitivity) versus FPR (1–Specificity). The blue dotted line was a random curve, which means the network had no ability to distinguish positive and negative cases. The red line is the ROC curve, which is above the blue line, indicating that the prediction results of Bayesian network did not follow the random distribution, guided by the model constructed with sample data and not by fortuity. In addition, the area under curve (AUC) value was 0.882, which is the area under the ROC curve. More reliable, the predicted results could be achieved if the AUC value is close to 1. Therefore, Bayesian network has ability to predict reliable results.

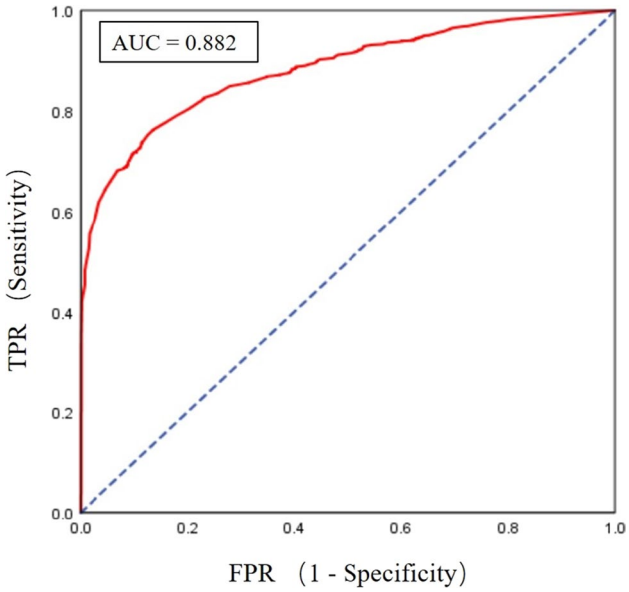

Fig. 7 ROC curve for Bayesian network performance analysis

Bayesian network has been applied in several areas, such as hydrology (Liu 2020) and Sponge City / LID (Hinesf and Landis 2014; Li 2020). However, the Bayesian network has limited applications in the specific theme of monitoring and simulating long-term hydrological performance of grassed swales and other similar facilities. Geraldi and Ghisi (2019) used Bayesian network to identify representative short-term time series in rainwater harvesting simulation, but the limitation of this work was to use the same data for training and validating which needed to be proved by new cases. Therefore, in this study, a new case of Utrecht swale was added to prove the network. Moreover, some studies proposed an efficient alternative to the use of short series by use of synthetic series (Brommundt and Bardossy 2022; Oliveira et al. 2015), which is beneficial due the incorporation of randomness in the data, generated a series extension from observed data. It also provide a new direction for adding man-made rainfall events to monitoring process in the future research.

Field applications and future research prospective

Field applications

Utrecht grassed swale was used as an example to verify the field application of Bayesian network. The verification process included two aspects: first one is to determine the representativeness of existing hydrological monitoring data for grassed swales, and another one is to guide monitoring

work by predicting the representativeness of upcoming rainfall events.

The first verification process is similar to the validation process in the "Establishment and validation of Bayesian network" section where 58 events (Table S2) were assumed as a reference event series. The reference dataset was randomly sampled to generate a set of short event series with event numbers of 10, 15, 20, 25, 30, 35, 40, 45, 50, and 55, each containing 1000 event series. Due to regional differences in rainfall and other characteristics, the new datasets were classified according to the "low," "medium," and "high" categories of their own characteristic indicator values.

The characteristic indicator values of reference event series corresponds to the categories of nodes 1–8, which were 60, "high," "low," "medium," "medium," "medium," "medium," "medium," and "medium," respectively, and brought them into the Bayesian network, and the output result was representative with a probability of 99.9%. Thus, it can be used as a reference series. The method presented in the "Classification of representative event series" section was used to classify representative and non-representative event series, and a validation dataset was generated. The prediction results are shown in Table 5 and Fig. 8 (ROC curve).

Table 5 shows that the accuracy, precision, and sensitivity of Bayesian network prediction were all above 80%, indicating a high probability of prediction for representative event series. The lower specificity indicated that the ability of the network in predicting non-representative event series was not good. In Fig. 8, the ROC curve was above the random curve and with a high value of AUC (0.824). Thus, the Bayesian network could be applicable to select and determine representative short event series of other grassed swales.

In the second verification process, it was assumed that the monitoring work was carried once a month (first event of each month) without the assistance of Bayesian network. Then, a short event series was obtained and the categories of node 1–8 in the network were set as 15, "high," "low," "medium," "medium," "medium," "medium," and "medium," respectively, and the output was non-representative (probability of representativeness was 0.41%). This

Table 5 Matrix of confusion—Bayesian network performance of field application

Experimental results	Predicted results		Performance indicators
	Yes	No	
Yes	6149 (TP)	834 (FN)	Accuracy—81.6%
			Precision—85.7%
No	1002 (FP)	2015 (TN)	Sensitivity—88.0%
			Specificity—66.8%

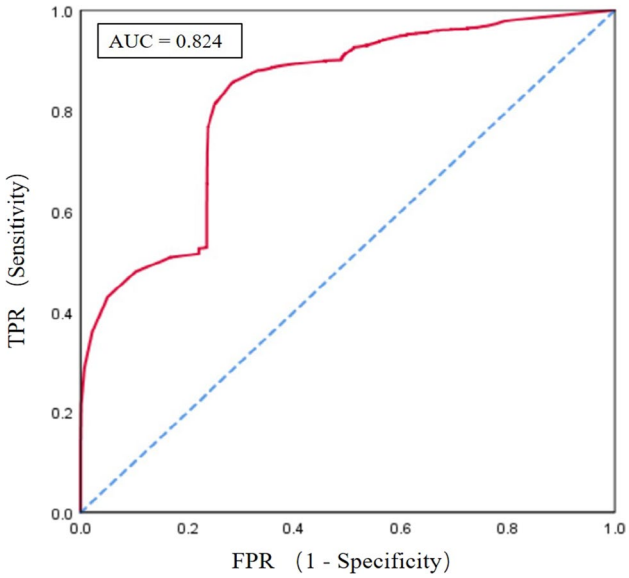

Fig. 8 ROC curve for Bayesian network application analysis

result was used as a comparison after Bayesian network was added in monitoring. If no event occurred, monitoring work would begin in November 2020 with the first rainfall event of each month being a necessary. Other rainfall events would be determined to be monitored or not through Bayesian network (the characteristics of rainfall would be obtained through weather forecasting in advance). For example, the first rainfall event in November was monitored, and the value of node 1–8 was calculated and classified, and then probability was obtained from the network. Before the second rainfall event, the approximate range of indicator values for the coming rainfall could be obtained based on weather forecasts. These two rainfall events were merged into a new dataset, and a new value of node 1–8 was calculated and classified, then the second representative probability value was obtained. If the value was less than the first value, the event would not be monitored. The specific results on representative event series generated by the Bayesian network are presented in Table S5. The fitting result showed that the similarity index between this representative series and the reference series was 0.03%, which is less than 5%. Therefore, the long-term hydrological monitoring method for grassed swales based on Bayesian network was effective, which not only ensured the simulation accuracy but also reduced the monitoring frequency.

Future research prospective

In this work, a Bayesian network was established based on the long-term monitoring data of Enschede grassed swale, to determine whether the monitoring data was representative in predicting the long-term hydrological effect

of grassed swale. However, this network is limited to areas with similar rainfall and climate conditions to Enschede. It is recommended to conduct long-term monitoring of grass swales under different climatic conditions, and to collect and added more characteristic information with regard to inflow/outflow, pollutants accumulation, plant growth, and human activities (maintenance, etc.) to improve the network. In addition, these characteristic factors are also need to be quantified and monitored to increase the number of nodes in Bayesian network, thus, to improve the accuracy of predictions by the network.

Conclusions

This study aimed to investigate the relation between the saturated hydraulic conductivity simulation results of grassed swales' filter medium and the characteristic indicators of various event series lengths used in simulation. The analysis showed that 60 events were the shortest representative series length according to the classifying rules. And by comparing distribution characteristic of representative events indicators, the seasonal index was a representative short event series indicator which was less affected by the change of the series length. Based on the findings, a decision-making tool was proposed by means of a Bayesian network to help stakeholders to determine whether a short hydrological monitoring data series was representative or not. From the validation results, Bayesian network revealed its significance in predicting the representativeness of a short event series of grassed swales. The network had robustness in predicting the representativeness event series but performed poorly in predicting non-representative series. The application of Bayesian network to Utrecht case also showed high prediction accuracy, and the optimized monitoring method based on Bayesian network greatly reduced the monitoring frequency while ensuring the accuracy of hydrological simulation. However, more cases and indicators are needed to improve the network prediction.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11356-024-32355-5.

Acknowledgements Authors would like to thank Southeast University, China, and TU-Delft, the Netherlands, for providing facilities required for current study.

Author contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Feikai Yang, Dafang Fu, Chris Zevenbergen, Floris C. Boogaard, and Rajendra Prasad Singh. The first draft of the manuscript was written by Feikai Yang and Rajendra Prasad Singh, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was finally supported by the China Scholarship Council (CSC No. 201906090081). This research was also supported by Jiangsu Provincial Key R&D Programme (Social Development) (Grant No. BE2022820) and SIA, grant number SVB/RAAK. PUB07.015 project "Groenblauwe oplossingen, kansen en risico's" (Green Infrastructure: Changes and challenges).

Data availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request

Declarations

Ethics approval The authors declare that current research fully abides by both local and international guidelines of ethical research regulations.

Consent to participate Not applicable

Consent for publication Not applicable

Competing interests The authors declare no competing interests.

References

Abualfaraj N, Cataldo J, Elborolosy Y, Fagan D, Woerdeman S, Carson T et al (2018) Monitoring and modeling the long-term rainfall-runoff response of the Jacob K. Javits Center Green Roof. Water 10(11):1494

Al-Rubaei AM, Viklander M, Blecken GT (2015) Long-term hydraulic performance of stormwater infiltration systems. Urban Water J 12(8):660–671

Besir AB, Cuce E (2018) Green roofs and facades: a comprehensive review. Renew Sust Energ Rev 82:915–939

Board T (2009) Urban stormwater management in the United States. Washington, DC: The National Academies Press

Boogaard FC (2022) Spatial and time variable long term infiltration rates of green infrastructure under extreme climate conditions, drought and highly intensive rainfall. Water 14(6):840

Borsuk ME, Stow CA, Reckhow KH (2004) A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. Ecol Model 173(2–3):219–239

Brommundt J, Bardossy A (2022) Stochastic generation of synthetic precipitation time series with high temporal and spatial resolution for engineering practice. University of Stuttgart, Germany

Brown RA, Skaggs RW, Hunt WF (2013) Calibration and validation of DRAINMOD to model bioretention hydrology. J Hydrol 486:430–442

Dagenais D, Brisson J, Fletcher TD (2018) The role of plants in bioretention systems: does the science underpin current guidance? Ecol Eng 120:532–545

Davis AP, Stagge JH, Jamil E, Kim H (2012) Hydraulic performance of grass swales for managing highway runoff. Water Res 46(20):6775–6786

Deletic A (2000) Sediment behaviour in overland flow over grassed areas. University of Aberdeen, England

Deletic A, Fletcher TD (2006) Performance of grass filters used for stormwater treatment-a field and modelling study. J Hydrol 317(3–4):261–275

Dierkes C, Geiger WF (1999) Pollution retention capabilities of roadside soils. Water Sci Technol 39(2):201–208

Ebrahimian A, Wadzuk B, Traver R (2019) Evapotranspiration in green stormwater infrastructure systems. Sci Total Environ 688:797–810

Fardel A, Peyneau PE, Béchet B, Lakel A, Rodriguez F (2019) Analysis of swale factors implicated in pollutant removal efficiency using a swale database. Environ Sci Pollut Res 26(2):1287–1302

Fort F, Jouany C, Cruz P (2012) Root and leaf functional trait relations in Poaceae species: implications of differing resource-acquisition strategies. J Plant Ecol 6(3):211–219

Fouli Y, Cade-Menun BJ, Cutforth HW (2013) Freeze-thaw cycles and soil water content effects on infiltration rate of three Saskatchewan soils. Canadian J Soil Sci 93(4):485–496

Garvelmann J, Pohl S, Weiler M (2015) Spatio-temporal controls of snowmelt and runoff generation during rain-on-snow events in a mid-latitude mountain catchment. Hydrol Process 29(17):3649–3664

Geraldi MS, Ghisi E (2017) Influence of the length of rainfall time series on rainwater harvesting systems: a case study in Berlin. Resour Conser Recycl 125:169–180

Geraldi MS, Ghisi E (2018) Assessment of the length of rainfall time series for rainwater harvesting in buildings. Resour Conserv Recycl 133:231–241

Geraldi MS, Ghisi E (2019) Short-term instead of long-term rainfall time series in rainwater harvesting simulation in houses: an assessment using Bayesian network. Resour Conser Recycl 144:1–12

Gong Y, Yin D, Liu C, Li J, Fang X (2019) The influence of external conditions on runoff quality control of grass swale in Beijing and Shenzhen, China. Water Pract Technol 14(2):482–494

Gregory JH, Dukes MD, Jones PH, Miller GL (2006) Effect of urban soil compaction on infiltration rate. J Soil Water Conserv 61(3):117–124

Haile TM, Hobiger G, Kammerer G, Allabashi R, Schaerfinger B, Fuerhacker M (2016) Hydraulic performance and pollutant concentration profile in a stormwater runoff filtration systems. Water Air Soil Pollut 227(1):34

Hakimdavar R, Culligan PJ, Guido A, Mcgillis WR (2016) The Soil Water Apportioning Method (SWAM): an approach for long-term, low-cost monitoring of green roof hydrologic performance. Ecol Eng 93:207–220

He L, Li S, Cui CH, Yang SS, Ding J, Wang GY et al (2022) Runoff control simulation and comprehensive benefit evaluation of low-impact development strategies in a typical cold climate area. Environ Res 206:112630

Hinesf EE, Landis WG (2014) Regional risk assessment of the Puyallup river watershed and the evaluation of low impact development in meeting management goals. Integr Environ Assess Manag 10(2):269–278

Hunt WF, Davis AP, Traver RG (2012) Meeting hydrologic and water quality goals through targeted bioretention design. J Environ Eng 138(6):698–707

Istchuk RN, Ghisi E (2022) Influence of rainfall time series indicators on the performance of residential rainwater harvesting systems. J Environ Manag 323:116163

Jennings SA, Lambert MF, Kuczera G (2010) Generating synthetic high resolution rainfall time series at sites with only daily rainfall using a master-target scaling approach. J Hydrol 393(3–4):163–173

Kandra HS, Deletic A, Mccarthy D (2014) Assessment of impact of filter design variables on clogging in stormwater filters. Water Resour Manag 28(7):1873–1885

Le Coustumer S, Fletcher TD, Deletic A, Barraud S, Lewis JF (2009) Hydraulic performance of biofilter systems for stormwater management: influences of design and operation. J Hydrol 376(1–2):16–23

Lefevre NJ, Davidson JD, Oberts GL (2009) Bioretention of simulated snowmelt: cold climate performance and design criteria. J Cold Reg Eng 145–154

Li DJ (2020) A research on risk factors of sponge city construction based on Bayesian network: a case study of the first batch of pilot cities in China. Zhejiang University of Technology, China

Liu Y (2020) Analysis of flow regime and ecological regulation of urban rivers. Shandong University, China

Lu LW, Chan FKS, Johnson M, Zhu FF, Xu YY (2023) The development of roadside green swales in the Chinese Sponge City Program: challenges and opportunities. Front Eng Manag 10:566–581

Marcot BG, Penman TD (2019) Advances in Bayesian network modelling: integration of modelling technologies. Environ Modell Softw 111:386–393

Mitchell VG (2010) How important is the selection of computational analysis method to the accuracy of rainwater tank behaviour modelling? Hydrol Process 21(21):2850–2861

Moghadas S, Leonhardt G, Marsalek J, Viklander M (2018) Modeling urban runoff from rain-on-snow events with the U.S. EPA SWMM model for current and future climate scenarios. J Cold Reg Eng 32(1):59–72

Muerdter CP, Wong CK, Lefevre GH (2018) Emerging investigator series: the role of vegetation in bioretention for stormwater treatment in the built environment: pollutant removal, hydrologic function, and ancillary benefits. Environ Sci Water Res Technol 4(5):592–612

Muthanna TM, Viklander M, Thorolfsson ST (2007) Seasonal climatic effects on the hydrology of a rain garden. Hydrol Process 22(11):1640–1649

Oliveira JPBD, Cecílio RA, Pruski FF, Zanetti SS (2015) Spatial distribution of rainfall erosivity in Brazil from synthetic precipitation series. Revista Brasileira De Ciências Agrárias 10(4):558–563

Paus KH, Morgan J, Gulliver JS, Leiknes T, Hozalski RM (2014) Assessment of the hydraulic and toxic metal removal capacities of bioretention cells after 2 to 8 years of service. Water Air Soil Pollut 225(1):1803

Paus KH, Muthanna TM, Braskerud BC (2016) The hydrological performance of bioretention cells in regions with cold climates: seasonal variation and implications for design. Hydrol Res 47(2):291–304

Pitt R, Chen SE, Clark SE, Swenson J, Ong CK (2008) Compaction's impacts on urban storm-water infiltration. J Irrig Drainage Eng 134(5):652–658

Rujner H, Leonhardt G, Marsalek J, Perttu AM, Viklander M (2018) The effects of initial soil moisture conditions on swale flow hydrographs. Hydrol Process 32(5):644–654

Rujner H, Leonhardt G, Perttu AM, Marsalek J, Viklander M (2016) Advancing green infrastructure design: field evaluation of grassed urban drainage swales. Novatech 2016: 9th International Conference on planning and technologies for sustainable management of Water in the City, Lyon, France

Saracoglu KE, Kazezyilmaz-Alhan CM (2023) Determination of grass swale hydrological performance with rainfall-watershed-swale experimental setup. J Hydrol Eng 28(3):04022043

Saudo-Fontaneda LA, Roces-García J, Coupe SJ, Barrios-Crespo E, Rey-Mahía C, Lvarez-Rabanal FP et al (2020) Descriptive analysis of the performance of a vegetated swale through long-term hydrological monitoring: a case study from Coventry. UK Water 12(10):2781

Shafique M, Kim R, Rafiq M (2018) Green roof benefits, opportunities and challenges–a review. Renew Sust Energ Rev 90:757–773

Wang M, Sun CH, Zhang DQ (2023) Opportunities and challenges in green stormwater infrastructure (GSI): a comprehensive and bibliometric review of ecosystem services from 2000 to 2021. Environ Res 236:116701

Willard L, Wynn-Thompson T, Krometis L, Neher T, Badgley B (2017) Does it pay to be mature? Evaluation of bioretention cell performance seven years postconstruction. J Environ Eng 143(9):04017041

Winston RJ, Powell JT, Hunt WF (2018) Retrofitting a grass swale with rock check dams: hydrologic impacts. Urban Water J 16(6):404–411

Xiao P, Wang T, Tian Y, Xie XM, You JJ, Tan XR et al (2023) A Bayesian network-based inhibition model of the rainstorm-landslide-debris flow disaster chain in mountainous areas: the case of the Greater Bay area, China. Water 15(17):3124

Xie YX, Wang Y, Huo MX, Geng Z, Fan W (2020) Risk of physical clogging induced by low-density suspended particles during managed aquifer recharge with reclaimed water: Evidences from laboratory experiments and numerical modeling. Environ Res 186:109527

Yang F, Fu D, Zevenbergen C, Boogaard FC, Singh RP (2023) Time-varying characteristics of saturated hydraulic conductivity in grassed swales based on the ensemble Kalman filter algorithm -a case study of two long-running swales in Netherlands. J Environ Manag 351:119760

Yousef YA, Hvitved-Jacobsen T, Wanielista MP, Harper HH (1987) Removal of contaminants in highway runoff flowing through swales. Sci Total Environ 59:391–399

Zaqout T, Andradóttir H (2021) Hydrologic performance of grass swales in cold maritime climates: impacts of frost, rain-on-snow and snow cover on flow and volume reduction. J Hydrol 597(11):126159

Zhao JH, Zhao YQ, Zhao XL, Jiang C (2016) Agricultural runoff pollution control by a grassed swales coupled with wetland detention ponds system: a case study in Taihu Basin, China. Environ Sci Pollut Res 23(9):9093–9104

Authors and Affiliations

Feikai Yang[1,2,3] · Dafang Fu[1,2] · Chris Zevenbergen[3] · Floris C. Boogaard[4,5] · Rajendra Prasad Singh[1,2]

✉ Rajendra Prasad Singh
rajupsc@seu.edu.cn

[1] School of Civil Engineering, Southeast University, Nanjing 210096, China

[2] Southeast University-Monash University Joint Research Centre for Future Cities, Nanjing 210096, China

[3] Department of Civil Engineering, Delft University of Technology (TU Delft), Gebouw 23, Stevinweg 1, 2628CN Delft, the Netherlands

[4] Research Centre for Built Environment NoorderRuimte, Hanze University of Applied Sciences, 9747 AS Groningen, the Netherlands

[5] Deltares, Daltonlaan 600, 3584 BK Utrecht, the Netherlands