Rapid gravity filtration operational performance assessment and diagnosis for preventative maintenance from on-line data

A. Upton [a], B. Jefferson [a], G. Moore [b], P. Jarvis [a,*]

[a] Cranfield Water Science Institute, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK
[b] Scottish Water, Castle House, 6 Castle Dr, Dunfermline, UK

HIGHLIGHTS

- Filtration performance was assessed using different data driven diagnostic methods.
- Filtration was effectively modelled using operationally relevant predictor variables.
- The CART algorithm was an effective tool for the diagnosis of operational filtration performance.

ARTICLE INFO

ABSTRACT

Rapid gravity filters, the final particulate barrier in many water treatment systems, are typically monitored using on-line turbidity, flow and head loss instrumentation. Current metrics for assessing filtration performance from on-line turbidity data were critically assessed and observed not to effectively and consistently summarise the important properties of a turbidity distribution and the associated water quality risk. In the absence of a consistent risk function for turbidity in treated water, using on-line turbidity as an indicative rather than a quantitative variable appears to be more practical. Best practice suggests that filtered water turbidity should be maintained below 0.1 NTU, at higher turbidity we can be less confident of an effective particle and pathogen barrier. Based on this simple distinction filtration performance has been described in terms of reliability and resilience by characterising the likelihood, frequency and duration of turbidity spikes greater than 0.1 NTU. This view of filtration performance is then used to frame operational diagnosis of unsatisfactory performance in terms of a machine learning classification problem. Through calculation of operationally relevant predictor variables and application of the Classification and Regression Tree (CART) algorithm the conditions associated with the greatest risk of poor filtration performance can be effectively modelled and communicated in operational terms. This provides a method for an evidence based decision support which can be used to efficiently manage individual pathogen barriers in a multi-barrier system.

1. Introduction

Rapid gravity filters provide the final barrier to particulates in most large municipal water supply systems. Public health risk arising from large water supply systems is primarily associated with short duration contamination or challenge events, such as those associated with extreme weather events, which are poorly captured by regulatory sampling programmes [1]. Breakthrough and transient periods of high turbidity have been associated with increased concentrations of oocysts, suspended solids and spore forming bacteria in distribution systems [2,3]. In addition, turbidity is widely interpreted and assumed to indicate removal performance of water-borne environmental pathogens [4,5]. Therefore, on-line turbidity meters provide a record to evidence filtration performance with a degree of granularity far greater than can be achieved by regulatory sampling.

Turbidity quantifies the extent to which suspended particles scatter light subject to their concentration, size and colour [6]. As an optical property, turbidity is not a direct health risk but has been associated with the presence of bacteria, the shielding of microorganisms from disinfection, causing additional chlorine demand, increasing disinfection-by-product (DBP) formation and promoting biological growth in distribution [7,8]. In the UK, the prescribed value for turbidity is 4 Nephelometric Turbidity Units (NTU) at the customer's tap with an indicative limit of 1 NTU for water leav-

ing the treatment works [9]. The World Health Organisation suggests that prior to chlorination large municipal supplies should average below 0.2 NTU and not exceed 0.5 NTU [10]. A best practice target of 0.1 NTU has been proposed to limit the risk of pathogen passage [11]. Water utilities aim to maintain low filtered water turbidity in order to minimise risk of bacteriological failure, reduce the cost of additional chemical dosing and lower DBP formation [12]. Though limitations to the sensitivity of turbidity have been observed in comparison to particle count monitoring its simplicity, reliability and economy ensure that turbidity remains the most widely used parameter for monitoring filter performance [12,13].

Visualisation of turbidity records is routinely used to assess and diagnose performance [14]. Typically, efforts by operators and scientists to monitor and improve the performance of filtration processes in terms of turbidity have used averages, percentile statistics, or compliance with a target value over various periods [12,15–17]. To aid consistent and objective management, investigators and practitioners have developed turbidity robustness indices (TRIs) to improve understanding of performance [11,18,19]. However, these metrics have not routinely been applied in practice. One of the aims of the research was to assess the suitability and reliability of these indices. Alternative approaches to performance assessment can be based around the best practice target of 0.1 NTU [11]. Performance can then be described in terms of the likelihood, frequency, and duration of quality target breaches. Such an approach, allows the application of basic reliability engineering metrics such as the mean time between failures (MTBF) and the mean time to recovery (MTTR) which can be applied to indicate reliability and resilience.

Once detected, a process fault is typically diagnosed by one of the following: from reference to prior information in quantitative or qualitative models of the process; by using historical data; or by combining more than one approach [20]. Purely quantitative modelling approaches to diagnosis of filtration performance are impractical because the underlying complex non-linear particle separation process is not accurately described by theory and measurement. Phenomenological and theoretical filtration models often rely on measurements which are not routinely collected in full scale water treatment [21,22]. Turbidity, for example, is not a quantitative measurement. The formalisation of qualitative knowledge into models is challenged by behavioural complexity of the process, inflexibility to new conditions and the generation of spurious diagnosis [23]. Process history based methods have been broadly categorised as quantitative or qualitative depending upon the method by which historical data is transformed and applied within the diagnostic system[24]. Current guidance suggests a form of manual qualitative trend analysis for rapid gravity filter fault diagnosis. This requires the time-consuming manual inspection and interpretation of filter profiles in order to identify potential issues and confirmation with further physical inspections and process investigation [15]. This investigation proposes a quantitative method to identify key operational issues associated with elevated filtrate turbidity, applicable over extended periods to provide easily interpretable diagnostic models for rapid gravity filtration operation and maintenance decisions. Such models can guide investigations reducing the time and financial and environmental cost incurred.

Treatment operators and managers need efficient, effective, robust and justifiable tools and methods for the aggregation and interpretation of large volumes of filter monitoring data into useful information from which evidence based decisions can be made. Using a turbidity target, such as the best-practice level of 0.1 NTU, we can frame the analysis of control system data as a classification problem whereby we identify the conditions associated with greater likelihood of high filtrate turbidity. Classification is a common task in machine learning and can be achieved by numer-

ous methods which can broadly be categorised into; linear, Bayesian, tree-based, clustering, neural-network and ensemble approaches. Broadly, linear methods such as logistic regression, identify and optimise a linear function to classify between one or more categories and Bayesian methods apply Bayes' rule. Tree based methods use recursive binary splitting of the feature space to fit a stepwise function. Clustering methods typically classify based on the Cartesian distance. Neural networks mimic the function of the human brain by optimising a collection of weightings and transfer functions (neurons) to return the most effective classifying function from a given architecture. Ensemble methods combine the results of many simple classification models to improve overall performance. Classification trees have been chosen for this application based on their primary virtue which is interpretability. This is key for the efficient and successful retrospective implementation of such a decision support tool for operators and managers facilitating more effective management of individual pathogen barriers. Further advantages of the classification tree methods are that they work effectively using discrete and continuous variables of any distribution and are insensitive to outliers [25].

The primary criticisms of classification trees are comparatively poor accuracy, a tendency to over-fit, instability and poor capture of additive structure. The objective of this investigation was to develop workable methods which can be applied to improve operational and preventative maintenance decision making on water treatment assets and for this reason interpretability trumps accuracy in this application. Though it is likely that alternative classification methods may produce better classification accuracy, the generation of the classification tree models allows far more broadly accessible communication and sense checking of the diagnosis. The tendency of classification trees to over-fit the data can be mitigated by appropriately using k-fold cross validation to estimate the extent of model pruning required. The problem that a small change in the data can cause a large change in the model and that similar splits often appear on multiple branches is inherent in the binary splitting algorithm and are the trade-off for the simplicity of interpretation [25].

Though classification trees have been implemented by numerous algorithms the two most popular methods are the classification and regression tree approach (CART) and the C4.5 and C5.0 algorithms [26]. The main distinctions between the implementation of these methods are the use of different functions to inform the split location, alternative pruning procedures, the possibility for multiway splits on categorical predictors and the potential for conversion to rules. CART has been widely applied and popular due to its accessibility and ease of interpretation when applied to non-linear processes [25]. CART has been applied to understanding and managing water contamination events and mechanisms [27,28]. Through recursive partitioning of explanatory variables, the conditions associated with an outcome of interest can be simplified and presented in an interpretable tree format using the classification and regression tree (CART) algorithm [29]. The CART algorithm is used in this study to produce interpretable models describing the operational conditions associated with the occurrence of elevated filtrate turbidity.

The aims of this paper were therefore to develop intelligent, data-driven decision support systems by assessing and developing existing performance metrics for summarising the performance of filtration processes in terms of turbidity and utilising other typical sources of data to identify the likely causes.

2. Materials and methods

Data was extracted from the control system at a water treatment plant in Scotland treating a mix of two upland surface water