

Modelling a water purification process for quality monitoring

Citation for published version (APA):

Meulen, van der, F. H., Luca, S., Overal, G., Dubbeldam, J. L. A., Di Bucchianico, A., & Jongbloed, G. (2014). Modelling a water purification process for quality monitoring. In J. Dubbeldam, W. Groenevelt, A. W. Heemink, D. Lahaye, C. Meerman, & F. Meulen, van der (Eds.), *Proceedings of the Study Group Mathematics with Industry (Delft, The Netherlands, January 27-31, 2014)* (pp. 36-52). Technische Universiteit Delft.

Document status and date:

Published: 01/01/2014

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Proceedings of the SWI 2014 Held in Delft

Modelling a water purification process for quality monitoring

Frank van der Meulen¹, Stijn Luca^{2*}, Gosse Overal³,
Johan Dubbeldam¹, Alessandro Di Bucchianico⁴
and Geurt Jongbloed¹

¹ TU Delft

² KU Leuven, Belgium

³ VU Amsterdam

⁴ TU Eindhoven

Abstract

This paper deals with a quality engineering problem introduced by ‘Waterlaboratorium Noord’ (WLN) situated at the Netherlands. Interest lies in determining an optimal sampling frequency that provides sufficient information on the water quality in a drinking water purification plant. The water purification plant that is studied consists of two aeration and filtration processes and a clear water reservoir where water is saved until distribution to households. One of the main processes during these filtration processes is iron removal.

A stochastic model is proposed that describes the decreasing effects on iron concentration after the filtration processes by multiplicative effects. This model is combined with an ordinary differential equation to model the amount of iron in the clear water reservoir that fluctuates due to the quality of the incoming filtrated water and the varying water demand.

In this way the iron concentration levels in the different compartments of a water purification plant can be simulated. Range and fluctuations approximate those of the observed data. Hence a realistic benchmark for detecting anomalies is obtained.

KEYWORDS: quality engineering, sampling frequency, stochastic model, differential equation, water quality

*Corresponding author

1 Introduction

1.1 WLN

The quality of water has an important impact on public health. Whether it is used for drinking, for food production or just for recreational purposes, contaminated water can lead to severe health problems (Karaniş et al., 2007). Therefore monitoring and keeping the quality of drinking water at a safe level is of crucial importance for our society.

The company ‘Waterlaboratorium Noord’ (WLN) takes care for the quality control of drinking water in Groningen and Drenthe. It monitors and improves the quality of drinking water provided by the water supply companies ‘Waterbedrijf Groningen’ and ‘Waterleidingsmaatschappij Overijssel’.

Generally water is retrieved as surface water or groundwater. This water needs to be treated in a purification plant for several reasons, to remove harmful substances, to ensure it looks clear and to remove pathogenic microorganisms, to name a few. This purified clear water is then pumped through a distribution network and finally flows clearly and steadily into the households. Figure 1 depicts a typical treatment scheme for groundwater purification.

1.2 Outline of the problem

The main question posed by WLN is the following:

Which monitoring frequency at the various stages in the water purification process is required for obtaining sufficient information on the water quality to prevent contaminated water to be delivered to households?

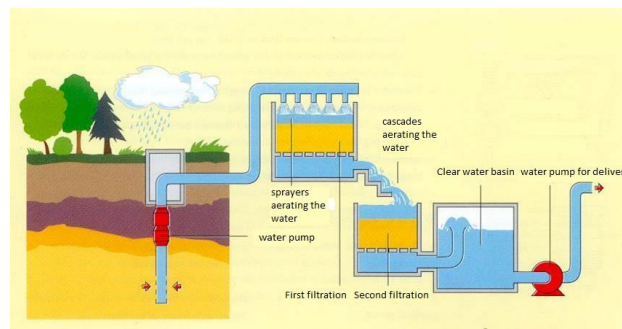


Figure 1: Illustration of a water purification process when groundwater is used as source.

One of the reasons this question is posed can be found in the Dutch legislation concerning the water quality. This legislation prescribes very strictly the quality control in the drinking water sources, in the clear water reservoirs and in the distribution networks. However, for the water inside the water purification plant there is only a rough guideline saying that the water company has to monitor the purification process for an ‘adequate process control’. Hence the only demand is that it is being frequently monitored on several locations spread along the plant such that a sufficient overview of quality can be obtained. At this moment only expert judgement is used to determine this monitoring frequency.

1.3 Approach

In this paper we restrict attention to a single water treatment plant with groundwater as a source and two filtration steps. One of the main goals is to detect malfunctions of the purification process which can result in unacceptable water quality. In an attempt to detect such anomalies a model is introduced for predicting iron concentrations at each place in the purification process given the iron concentrations in the groundwater. A next step would be to monitor the differences between these predictions and the observations to reassure the quality of Dutch water.

Figure 2 shows a schematic overview of the proposed model of the purification process. The four main locations of the purification process are interpreted as basins where measurements are taken from. The transitions between the first three basins are modeled using a stochastic model with a multiplicative effect while the fluctuations of the iron in the clear water before distribution is modeled using a differential equation. The implementation is performed using the statistical software package R (R Core Team, 2012).

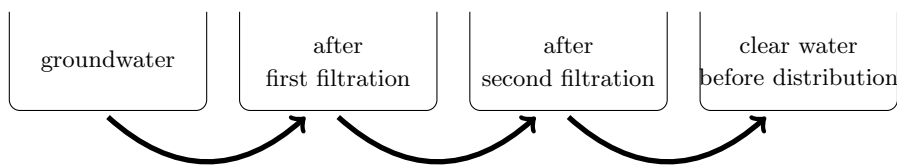


Figure 2: Schematic illustration of the water purification process

1.4 Outline of this article

In the next section we detail the data made available by WLN. In section 3 we present a model that relates the concentration of iron in groundwater

Proceedings of the SWI 2014 Held in Delft

to the concentration in the clear water basin. This model uses both ideas from stochastic processes and differential equations for modelling conservation laws. We then move on to the results obtained. The final section contains conclusions, discussion and recommendations.

2 Available data

The data concern a water treatment plant that uses groundwater and basically removes substances through aeration and filtration. There are no pathogenic microbes in the water, because it is pumped up from deep beneath Earth’s surface. The dataset consists of measurements of concentrations of several ions like e.g. Fe, Mg, NH_4 . These measurements have been taken at four places during the purification process:

1. in the groundwater,
2. after the first filtration,
3. after a second filtration,
4. before the distribution (clear water).

The data include measurements over a period of 5 years starting from January 2009. The measurement frequency is about 1-2 times a week. Figure 3(a) shows the observed iron concentrations on each of the four locations mentioned above. As one can see the iron concentration decreases after each filtration.

Flushes of the filters at regular times cause fluctuations in the iron concentration after each filtration as can be seen in figure 3(b). Immediately after a flush the remaining fraction of iron in the water reaches a peak and subsequently decreases rapidly back to its original value. The peaks after first and second filtration in figure 3(b) seem to occur periodically, however an exact flush period is not made available. Figure 4 shows some typical flush curves describing the remaining fraction of iron in the water as a function of time since the filter has been cleaned. These samples are taken every 10 minutes during a period of roughly 12 hours. The curve can be modeled using non-linear regression analysis as will be discussed in the next section.

Fluctuations in the clear water are expected to be lower as the large volume of the clear water reservoir averages out most of the fluctuations. Fluctuations in this basin can, among other things, be caused by variations in the water demand.

In what follows t denotes the inspection time of the measurements, expressed in hours, since the first available measurement. Note that the time scale t is different than the one used to describe the flush curves. There, the

Proceedings of the SWI 2014 Held in Delft

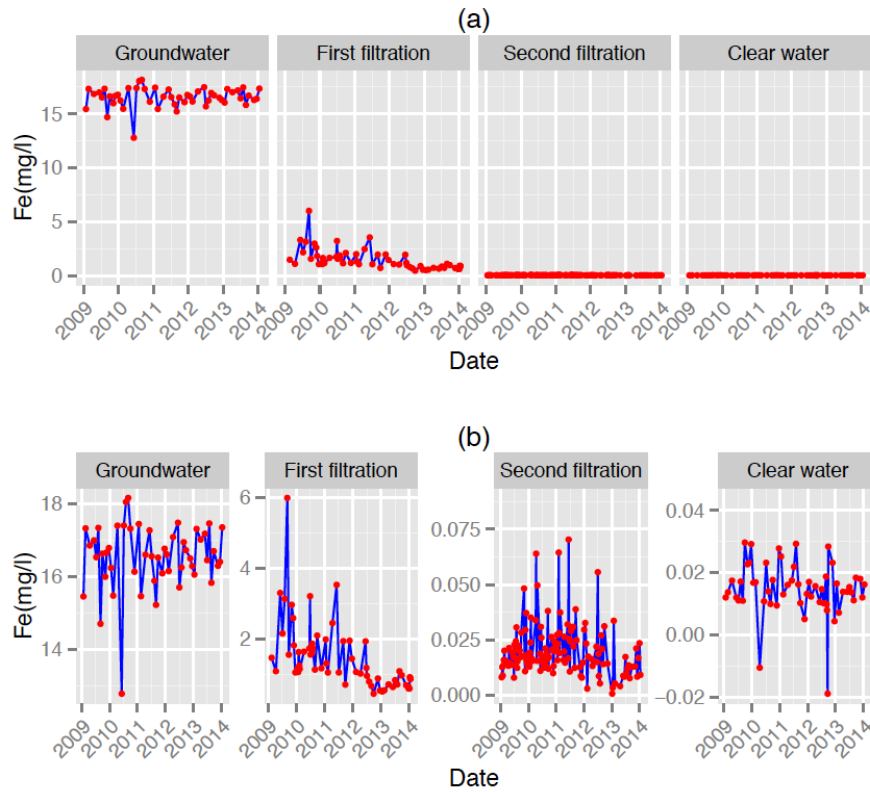


Figure 3: (a) Plots of observed iron concentrations at each location. (b) Plots with free axes scales to illustrate the fluctuations.

inspection time is expressed in hours after the last flushing. No link between these time scales is available, in the sense that for measurements y_t and z_t taken after first and second filtration respectively the time that is passed since the last flushing is not known.

3 Modelling the water purification process

3.1 A stochastic model for the first filtration process

The iron concentration in the water just before it enters the first filter is denoted by X_t and can be modeled using a time series model. However the observations in the dataset are obtained with 2-4 days in between. Therefore

Proceedings of the SWI 2014 Held in Delft

the ‘incoming iron concentrations’ can assumed to be independent realizations of a random variable X having density function f_X . As a function of time s after flushing the filter is assumed to have a multiplicative effect on the iron concentration in the water. This means that the iron concentration in the water measured at time s after flushing is given by

$$Y_s = X \cdot \alpha(s) \tag{1}$$

where $s \in [0, 12]$ and the cleaning of the filter is assumed to happen every twelve hours. As noted before, the function $\alpha(s)$ indicates the temporal dependence of suspension in the water after the first filtration and is measured as the remaining fraction of iron in the water as a function of time s since the last filter cleaning. Based on the analysis of this *flush curve* $s \mapsto \alpha(s)$ we assume the following model:

$$\alpha(s) = a \left(1 + \frac{s}{b} \right)^{-n} + c, \quad s \geq 0. \tag{2}$$

The function $s \mapsto \alpha(s)$ has a horizontal asymptote at $y = c$. The parameter n determines the shape (and steepness) of the curve whereas the parameter a determines the intercept at $(0, a + c)$. The parameter b on the other hand can be viewed as a scale parameter. Non-linear regression analyzes were performed to show the appropriateness of the chosen model in 2, see figure 4.

An optimal fit of the parameters $\theta = (a, b, c, n)$ is found using maximum likelihood estimation based on the measurements of the iron concentrations in the groundwater and after first filtration. This estimation is based on the following assumptions:

1. The ‘incoming iron concentrations’ are assumed to be independent and identically distributed according to a normal distribution $N(\mu_X, \sigma_X)$. Figure 5 indicates that the normality assumption is indeed not violated severely.
2. As noted before, the inspection time s after filtration at which the concentration Y_s is measured, is not known. For now this time is modelled by an uniformly distributed random variable on $[0, 12]$.

Assuming that the inspection times s for Y_s are drawn from a uniform distribution implies

$$Y = X \cdot \alpha(12U), \tag{3}$$

where $X \sim f_X$ and $U \sim \text{Unif}(0, 1)$ are independent. Based on the normality assumption it is clear that the distribution of Y will depend on the parameters $(\mu_X, \sigma_X, \theta)$.

Note that no paired observations are available, in the sense that for a particular input with concentration X , the corresponding concentration Y

Proceedings of the SWI 2014 Held in Delft

is measured. Such information could improve the model and would be recommended in the future. For the moment it is assumed that a number of independent realizations of Y is observed, originating from (3). We now derive the density of Y . First recall the formula for the probability density of the product Y of two independent continuous non-negative random variables X and Z :

$$f_Y(y) = \int_{x=0}^{\infty} \frac{1}{x} f_Z\left(\frac{y}{x}\right) f_X(x) dx. \quad (4)$$

In order to use this relation to obtain the density of Y in (3), we need the density of the random variable

$$Z = \alpha(12U) = a \left(1 + \frac{12U}{b}\right)^{-n} + c,$$

which is given by:

$$f_Z(z) = \frac{b}{12na} \left(\frac{a}{z-c}\right)^{\frac{1}{n}+1}, \quad z \in \left[c + \frac{a}{(1+12/b)^n}, c+a\right].$$

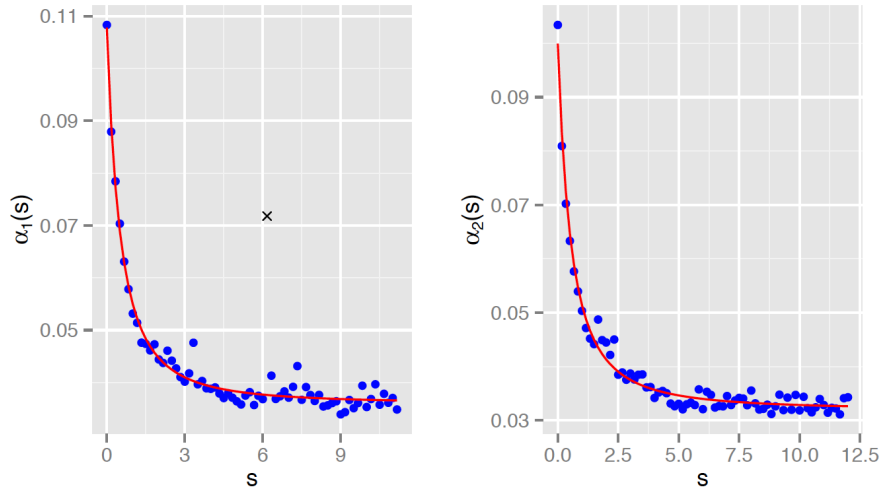


Figure 4: Flush curves of the first filtration process obtained after flushing the filter. The time s is expressed in hours after the moment of flushing. The fitted curves are respectively given by $\alpha_1(s) = 0.04 + \frac{0.07}{(1+s)^{1.92}}$ and $\alpha_2(s) = 0.03 + \frac{0.07}{(1+s)^{1.80}}$, where $b = 1$. The cross in the left plot is considered to be an outlier and is not incorporated in the regression analysis.

Proceedings of the SWI 2014 Held in Delft

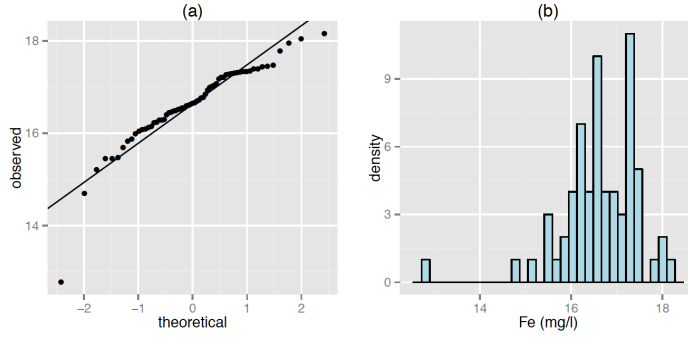


Figure 5: A quantile-quantile plot of the iron concentrations in groundwater with respect to the normal distribution (left) and a histogram of the iron concentrations (right).

Combined with (4), this leads to the following two-parameter model for the observed concentrations Y_1, \dots, Y_N of iron after the first filter:

$$f_Y(y; \mu_X, \sigma_X, \theta) = \frac{ba^{1/n}}{12n} \int_{x_\ell}^{x_r} x^{1/n} f_X(x, \mu_X, \sigma_X) (y - cx)^{-1-1/n} dx$$

where we make the dependence on the parameters $(\mu_X, \sigma_X, \theta)$ explicit and where the integration bounds are given by

$$x_\ell = x_\ell(y, \theta) = \frac{y}{c+a} \quad \text{and} \quad x_r = x_r(y, \theta) = \frac{y}{c+a(1+12/b)^{-n}}.$$

Given the observed values y_1, \dots, y_N of Y , the log likelihood function is given by

$$\ell(\mu_X, \sigma_X, \theta) = \sum_{i=1}^N \log f_Y(y_i; \mu_X, \sigma_X, \theta). \quad (5)$$

This function can be maximised numerically using the so-called limited memory BFGS algorithm (Byrd et al., 1995). This optimization algorithm allows constraints on the parameters and is implemented in R under the ‘stats’ package.

3.2 A stochastic model for the second filtration process

We feel the second filtration step can be dealt with in a similar fashion as the first filtration. Due to time limitations, this step has not been worked out yet. During the simulation, we simply divide the measurements after the second

Proceedings of the SWI 2014 Held in Delft

filtration step by a fixed number (chosen to match the data on average). Of course, this is a big simplification that needs to be handled more accurately in future work.

3.3 Modeling the concentration of iron in the clear water basin

To examine the propagation of the compounds in the water when the water is transferred between different basins, we developed an ordinary differential equation model. This model describes how the amount of a contaminant in the water, in the case under consideration this is iron, changes in the different basins. To illustrate the method, we derive the equations for the amount of iron in the clear water basin, but a similar procedure could be followed for the other basins as well.

We assume that the volume of the clear water basin $V(t)$ changes in time according to the following equation:

$$\frac{d}{dt}V(t) = c_0(t) - f(t), \quad (6)$$

where $c_0(t)$ denotes the volume influx to the clear water basin and $f(t)$ is the water outward flux that typically consists of a constant part and a fluctuating part, as the water demand is a fluctuating quantity. For example, during day time more water is used than in the night:

$$f(t) = 60(83 + 5 \sin(2\pi t/24)).$$

The numbers in this expression designate that 83 l/s is the typical rate at which water leaves the clear water reservoir. Furthermore the value of the amplitude of the sine function is chosen to correspond to the size of the fluctuations of iron concentrations, which we will see in the next section.

Next, we model the amount of iron, denoted by $u(t)$ in the clear water reservoir. The governing equation for $u(t)$ is

$$\frac{d}{dt}u(t) = c_0(t)g_{in}(t) - u(t)f(t)/V(t). \quad (7)$$

Equation (7) describes the influx of iron with rate $c_0(t)g_{in}(t)$ and outward flux with rate $u(t)f(t)/V(t)$. The function $g_{in}(t)$ which models the concentration of iron after the second filter is not known as a function of time. However from measurements that were performed at random times a typical size of the fluctuations can be inferred. Equations (6) and (7) can be solved in time using a simple Euler forward method which works as the system considered consists of linear equations (Xie, 2010).

Proceedings of the SWI 2014 Held in Delft

3.4 Assessing the fit of the model

To assess the fit of the model a simulation was performed. Results of this simulation can then be compared with the given observations of the clear water.

We propose to model the iron concentration in the groundwater by a stationary autoregressive time series model of order 1:

$$X_t = \mu_X + r(X_{t-1} - \mu_X) + \varepsilon_t, \quad (8)$$

where $\{\varepsilon_t\}_t$ is a sequence of independent $N(0, \sigma_\varepsilon^2)$ distributed random variable. Assuming $|r| < 1$ ensures that the (causal) stationary distribution of (8) exists uniquely and is normally distributed. As no data at this time are available to fit this model, we somewhat arbitrarily chose $r = 0.8$. This choice implies positive dependence of iron concentrations in the groundwater over different lags. We simulated 10^6 points with a time step of $dt = 0.5$. This corresponds to a simulation of 5.7 years of data which approximates the period over which the observed data of WLN is taken. Using our model one is able to predict the outcome when this data is considered to be the iron concentration of the groundwater.

The corresponding concentrations after first filtration are obtained by applying formula (1) using ratios $\alpha(s)$ evaluated at 10^6 equidistant times with a time leap of dt starting from $s = 0$. The concentrations after the second filtration were obtained in a similar way applying the same multiplicative procedure on the iron concentration after the first filtration. This resulted in a discrete time series consisting of 10^6 points that can be used as input for $g_{in}(t)$ of the differential equation in (7). The initial conditions to solve the differential equations were chosen to correspond to the stationary state values of $u(t)$ and $V(t)$:

$$\begin{cases} u(0) &= 0.015 V(0) \text{ mg} \\ V(0) &= 7500000 \text{ l} \end{cases} \quad (9)$$

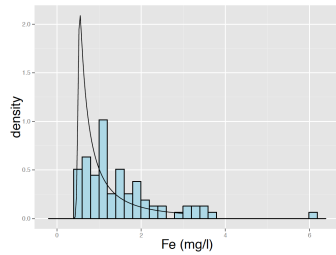
Finally a number of 70 points are randomly chosen from the simulated 10^6 points and compared with the observations of WLN.

4 Results

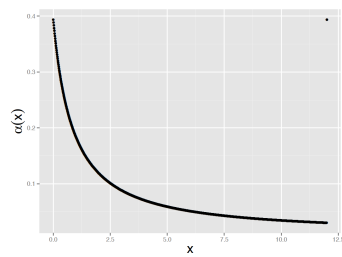
In the non-linear regression analysis of the flush curves, the parameter b in (2) could be set to 1 while keeping a satisfactory curve fitting. The parameters $(\mu_X, \sigma_X, a, c, n)$ were estimated by maximizing the log likelihood function (5):

$$\hat{\mu}_X = 16.62, \quad \hat{\sigma}_X = 0.88, \quad \hat{a} = 0.38, \quad \hat{c} = 0.01, \quad \hat{n} = 1.1496307$$

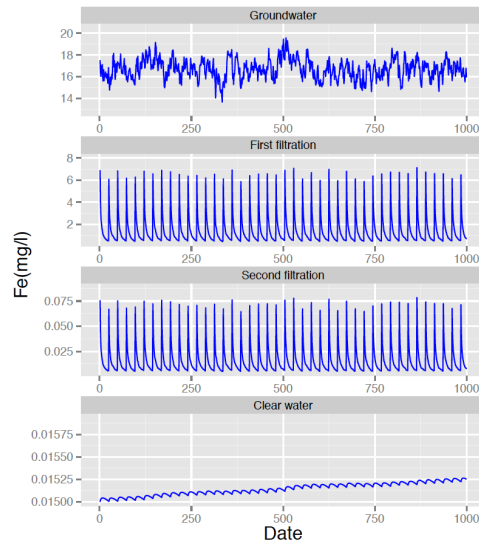
Proceedings of the SWI 2014 Held in Delft



(a) The fitted distribution f_Y on the observed data of Y .



(b) The flush curve after first filtration obtained using our MLE approach.



(c) Plots of simulated time series data in each basin for the first 1000 points simulated.

Figure 6(a) shows the fit of the density distribution f_Y to the observed iron concentrations after the first filtration. A slight underestimation of the variance is noted. Figure 6(b) shows the flush curve that is obtained using the estimated parameters $(\hat{a}, \hat{c}, \hat{n})$.

Figure 6(c) shows the simulated concentrations of iron in each basin using the time series data generated from (8). The x -axis shows the indices of the simulated data corresponding to some moment in time where the origin $t = 0$ has to be interpreted as the first inspection time available in the dataset of WLN.

As one can see the ranges of the simulated data in the first three basins of our model approximate those of the observed data in figure 3. Furthermore the decrease of the iron concentrations after first and second filtration is clear. The clear water contains an iron concentration of 0.015 at $t = 0$ as induced by (9). The figure shows an initial increase of the iron concentrations.

Figure 6 shows the complete simulated time series of iron concentrations in the clear water basin. After some time period the iron concentrations fluctuate around a stationary value of approximately 0.02. This corresponds with the observed data in the clear water basin shown in figure 3(b). This is, of course, in accordance with expectations as the large volume of the reservoir

Proceedings of the SWI 2014 Held in Delft

averages out most of the fluctuations.

The observations in figure 3 are sampled at irregular moments in time and with an irregular frequency, typically given by 1-2 times a week at different weekdays. Because of the flush activities the sampled observations reach a peak when taken at a moment short after flushing. This happens typically when a sample is taken within an hour after flushing, see figure 6(b). In a later time period, when the last flush occurred more than 3 hours ago, more steady concentrations are observed. Hence the observed fluctuations depicted in figure 3 are inherent to the sampling method.

This sampling method can be mimicked by randomly choosing a number of N observations from the time series obtained for the clear water basin (figure 7). To make a consistent comparison with the observed data, N is chosen to equal the number of observations available from the clear water basin, i.e. $N = 70$. In this way we were able to simulate fluctuations in the iron concentrations that are approximately of the same magnitude as the measured fluctuations.

5 Discussion

In this section we first summarize the conclusions that we may draw from our results. We then present recommendations to WLN for further actions to address the problem at hand. Finally, we suggest some further lines of research.

5.1 Conclusions

At this stage of research, the model as presented in section 3 may be too simple to realistically model the water purification process. This is partly due to

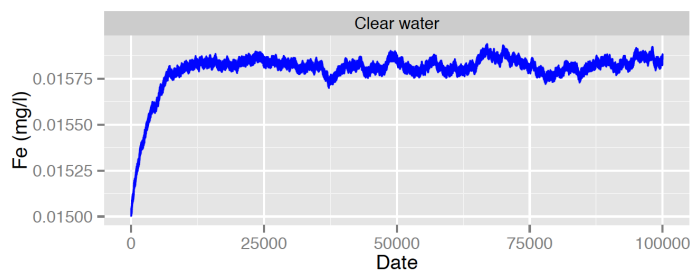


Figure 6: (a) Complete simulated time series of iron concentrations in the clear water basin.

Proceedings of the SWI 2014 Held in Delft

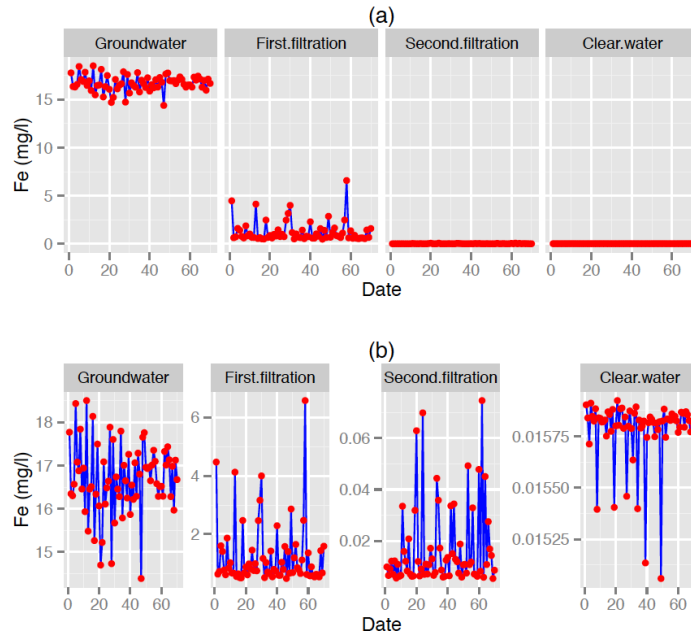


Figure 7: (a) Plots of simulated observed data in each basin. (b) Plots with free axes scales to illustrate the fluctuations.

time constraints for building the model, but also due to the nature of the available data right now. Presently, the data provided by WLN only allows for a rough calibration of the model parameters. More accurate estimation may be accomplished in the future by setting up a specific measurement scheme with the purpose of model calibration in mind. Once such measurements have been obtained and the model is sufficiently refined (i.e. realistically modeling the characteristics of the water purification process), process inherent variation can be estimated. The latter directly gives information on the required monitoring frequency: wildly fluctuating iron concentrations will require a relative high monitoring frequency, whereas a process with low variability may allow for less measurements taken over time.

To develop an effective and efficient monitoring procedure it is necessarily to properly define the process to be monitored and the desired performance of the monitoring procedure. These descriptions come from operational and legal constraints, but must be translated into statistical descriptions in order to be able to provide a sound basis for the assessment of a monitoring procedure.

A proper definition of the monitored process includes a description of

Proceedings of the SWI 2014 Held in Delft

the available data, a list of variables to be monitored, a description of an acceptable or “in-control” situation and the sampling strategy (Hawkins and Olwell, 1998; Kenett and Zacks, 1998). In this application the variables that are monitored are the concentrations of ions (e.g. iron). In this case an “in-control” situation for each variable is defined by so-called control limits. These control limits are driven by the natural variability of the concentration in the drinking water. For instance, in the case at hand one would desire that the iron concentration in the clear water remains at a constant (probably low) level while fluctuations are limited.

However it seems that the current approach performed by WLN is based on the monitoring of univariate variables evaluated relative to specification limits. These specification limits describe the maximal allowable deviation from a desired value of the variable, called the target value. In contrast to the control limits these specifications are determined externally and are not related to the natural variability of the variables. Therefore these specifications mostly allow a variability that is larger than the variability naturally induced by an in-control system. For instance a malfunctioning filter or a slow but persistent increasing value of a certain concentration does not have to lead to exceedances of the specification limits. Hence information about inherent changes of the water quality and the filtration processes can be lost using this approach.

The importance of the distinction between specification and control limits is widely known in the literature of statistical process control (Montgomery, 2013). To illustrate this difference further we can say that a purification process performs within specifications when the water quality is acceptable from a public health point of view. However at the same time the water company may suffer from excessive costs due to an out-of-control process caused by some malfunctioning. When one is aware of such malfunctioning preventive maintenance actions or increasing surveillance could be performed without interrupting purification. In this way unnecessarily costs can be avoided.

Furthermore, the performance of the monitoring procedure is mainly determined by the time that is needed to detect a malfunctioning or contamination. Using proper inherent control limits of a process that is in a steady state one can objectively determine appropriate sampling frequencies that assure the desired performance (Montgomery, 2013). However, due to flushing in the normal state after the first filtration, the process is not in a steady state. On the contrary it rather shows cyclic behaviour. Therefore, the effect of flushing for a single location was modeled with a simple stochastic model obtaining a reasonable agreement with actual data from iron concentrations. Monitoring using this model is then in fact monitoring a profile (Noorossana et al., 2012). In order to connect flows from one filter to another, we set up a system of coupled differential equations obtained from simple conservation laws.

Proceedings of the SWI 2014 Held in Delft

Finally, remark that one should also take into account that the lab needs time to perform analyses. Thus demanding detection of abnormal levels of certain ion concentrations within e.g. 2 hours is not feasible when the water quality analysis in the lab takes 1 day. Another important aspect is to determine optimal measurement locations in the purification plant. It may seem optimal to monitor as much as possible upstream in the purification plant to assure on-time detection. However, in this way one may fail to detect anomalies downstream. Thus, it is sensible to monitor at several locations.

To summarize, the main conclusions are:

- (i) monitoring is now performed by checking individual specification limits, and thus fails to take into account deviations from normal process deviations
- (ii) the cyclic concentration levels due to flushing may be modelled adequately by a simple model so that we have a realistic benchmark for detecting anomalies
- (iii) the flows from one filter to another may be modeled by a system of coupled differential equations obtained from simple conservation laws

5.2 Recommendations

We start with some recommendations on monitoring in general.

1. Change the monitoring procedure from checking specification limits to checking inherent concentration fluctuations in order to obtain a more responsive monitoring system (and thus have a well-grounded justification for the associated sampling frequency). These statistical limits will be more strict than the chemical/health limits, thus there is no risk for exceeding chemical/health limits.
2. Specify detection performance for all concentrations, taking into account processing time of analytical analyses in the lab and out-of-control scenarios like trends (see e.g., Friséen (2003) for a discussion of different performance metrics).
3. Study correlations between concentrations of different ions so that one may use the available data more efficiently.
4. Synchronize timing of measurements between compartments in a purification plant in order to track the flow water drops and thus improve detection performance.
5. Reduce variance in the beginning by blending the right wells.

Proceedings of the SWI 2014 Held in Delft

6. Perform a pilot study with in-line measurements on all measurement locations in one purification plant in order to improve the models of this paper.

The following recommendations concern modeling the effects of flushing.

1. Measure the incoming iron concentration together with corresponding (time-aligned) iron concentration after flushing ; this allows for much more accurate modelling.
2. Keep track of the actual mixture of incoming water sources since this has an impact on the distribution of the incoming iron concentrations.

5.3 Future Research

The study in this paper was restricted to the monitoring of iron concentrations due to time limitations. In future research other parameters could be studied as well. Furthermore, the model has to be completed to describe the second filtration process. As a next step simulations could be performed where anomalies are artificially implemented. Such simulations would be useful in finding an optimal monitoring frequency that enables us to detect a small (to be determined) shift (with a certain probability). Finally, it would be interesting to adapt our model to include correlations between different ions. In this way an alarm system can be build to detect dangers of combined high levels.

Acknowledgements

The authors would like to express their appreciation to Peter van der Maas from WLN who has introduced the problem treated in this paper that was fascinating and challenging to work on. Also a warm gratitude goes out to the organising committee of the Study Group Mathematics with Industry held at TU Delft to give us the possibility to work on this problem in a collaborative atmosphere.

References

- Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.
- M. Frisé. Statistical surveillance. Optimality and methods. *International Statistical Review*, 71(2):403–434, 2003.

Proceedings of the SWI 2014 Held in Delft

- D.M. Hawkins and D.H. Olwell. *Cumulative Sum Charts and Charting for Quality Improvement*. Springer, New York, 1998.
- P. Karanis, C. Kourenti, and H. Smith. Waterborne transmission of protozoan parasites: A worldwide review of outbreaks and lessons learnt. *Journal of Water and Health*, 5(1):1–38, 2007.
- R.S. Kenett and S. Zacks. *Modern Industrial Statistics*. Duxbury Press, 1998.
- D.C. Montgomery. *Statistical Quality Control*. John Wiley & Sons, 2013.
- R. Noorossana, A. Saghaei, and A. Amiri. *Statistical Analysis of Profile Monitoring*, volume 865. Wiley, New York, 2012.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.
- Wei-Chau Xie. *Differential Equations for Engineers*. Cambridge University Press, 2010.