

Multivariate Analysis of Groundwater-Quality Time-Series Using Self-organizing Maps and Sammon's Mapping

Rebecca M. Page^{1,2} · Peter Huggenberger¹ · Gunnar Lischeid³

Received: 8 April 2014 / Accepted: 27 May 2015 /
Published online: 13 June 2015
© Springer Science+Business Media Dordrecht 2015

Abstract Groundwater extracted from alluvial aquifers close to rivers is vulnerable to contamination by infiltrating river water. Infiltration is often increased during high discharge events, when the levels of waterborne pathogens are also increased. Water suppliers with low-level treatment thus rely on alternative measures derived from information on system state to manage the resource and maintain drinking-water quality. In this study, a combination of Self-Organizing Maps and Sammon's Mapping (SOM-SM) was used as a proxy analysis of a multivariate time-series to detect critical system states whereby contamination of the drinking water extraction wells is imminent. Groundwater head, temperature and electrical conductivity time-series from groundwater observation wells were analysed using the SOM-SM method. Independent measurements (spectral absorption coefficient, turbidity, particle density and river stage) were used. This approach can identify critical system states and can be integrated into an adaptive, online, automated groundwater-management process.

Keywords Groundwater · Time-series analysis · Self-organizing map · Sammon's mapping · Drinking water quality

✉ Rebecca M. Page
rebecca.page@unibas.ch

¹ Applied and Environmental Geology, Department of Environmental Sciences, University of Basel, Bernoullistrasse 32, 4056 Basel, Switzerland

² Endress+Hauser Metso AG, Kägenstrasse 2, 4153 Reinach, Switzerland

³ Institute of Landscape Hydrology, Leibniz Centre for Agricultural Landscape Research, Eberswalder Strasse 84, 15374 Müncheberg, Germany

1 Introduction

Alluvial groundwater is an important source of drinking water in many countries, providing approximately 50 % of the global drinking-water supply, as well as a significant part of water required for industrial and agricultural purposes (Zektser and Everett 2004). Many drinking water extraction wells are located near rivers, drawing water from high-permeability alluvial deposits in the river valley. The filter capacity of the riverbank and alluvial aquifer is not constant and varies with different hydrological conditions. Not all conditions enable the filter capacity to remove river-borne microbial contamination before it reaches groundwater extraction wells (Taylor et al. 2004; Dash et al. 2010). Increased infiltration of river water into an aquifer during high discharge events can result in microbial contamination of drinking water extraction wells (Regli et al. 2003). The contamination potential is strongly dependent on the nature of high discharge events, the location of the extraction wells relative to the river and the composition of the aquifer material and river bed (Regli et al. 2003; Page et al. 2012). Alongside these spatial considerations, the duration and level of contamination, the filter capacity and the distance travelled by the freshly infiltrated water are influenced by temporal factors including (a) the history of high discharge events, (b) the relative change in river stage over time, (c) seasonality, (d) the varying load of river water contaminants (Regli et al. 2003). The resulting interaction of spatial and temporal factors influencing groundwater quality fluctuations can lead to problems in securing the supply of drinking water. Instead of investing in elaborate water treatment equipment, it may often suffice to manage drinking water production according to actual groundwater quality. Three possible management options for water suppliers with a strong temporal variability of water quality, are to (a) continue extraction and invest in additional treatment steps (e.g., filters, UV, disinfection), (b) stop extracting and purchase water from other suppliers, or (c) reduce extraction from wells at risk and monitor the water quality and its temporal development ('adaptive management').

However, determining actual groundwater quality is strongly based on the selection of representative measurement parameters and the timely interpretation of patterns observed in the measurements. Early warning systems function by detecting changes in the system under observation, e.g., extracted water quality, as they occur. Low temporal resolution or time delays in such systems reduce the breadth of management options available. Technological advancements (e.g., remote data collection) and availability of measurement equipment allow a temporally high-resolution collection of data (online and real-time) (O'Flynn et al. 2010). However, the greater amount of data available also makes the interpretation a greater task, requiring an efficient mechanism to recognize the system states that are of interest, i.e., hazardous or suboptimal system states, especially when current and past situations need to be considered jointly. Hazardous system states need to be rapidly identified, a task which requires the operations manager to be able to view and assess the current system state and decide on the best course of action within a short period of time before any damage has been done. The complexity is also increased, as these states can be induced by one or more measured parameters indicating a deviation from normal, or desired situations (Camplani et al. 2009; Fuertes et al. 2010). The identification and interpretation of significant fluctuations in time series requires time and knowledge, a complex task ideally automated so that the operator of the process under surveillance can select the steps necessary to steer the system back to the desired state. This calls for not only a structured data management system, but also for a methodology for interpreting the recorded data (Stefanovic et al. 2015).

One method used to identify and track system states are Self-Organizing Maps (SOM), which are a form of artificial neural networks (ANN) (Kohonen 2001; Mustonen et al. 2008; Fuertes et al. 2010). These ANNs are able to extract the inherent structure, i.e., the underlying patterns, directly from a data set without an explicit physical model by resolving nonlinear input–output relationships in complex systems. They have a broad spectrum of applications, ranging from speech recognition over image analysis to anomaly detection (Kohonen 2001; Mustonen et al. 2008; Camplani et al. 2009). In this study, the methodology has been applied to identify hazardous events based on time series analysis. The first step is to find patterns in a system, e.g., groundwater time series, by considering changes over time and associating them with system states using independent parameters. During the next step, the different system states can be classified as e.g., hazardous or non-critical.

The overall aim of this study is to identify critical system states of groundwater quality as they develop, potentially enabling an early initiation of counter measures, e.g., reduced extraction, to avoid contamination of a drinking water supply plant. The methodology is based on a multivariate data set of groundwater quality proxy indicators. Furthermore, specific objectives are to (1) identify patterns in groundwater quality, which can be associated with hazardous system states, and (2) assess the degree to which this methodology represents the inherent patterns in groundwater quality over time and space. This study tests the methodology as a novel approach to managing riverine groundwater extraction by combining online measurements with automated multivariate analysis.

2 Methodology

For the development and testing of the methodology and conceptual approach, we selected a water supply system using several wells, some of which are located close to a river and affected by freshly infiltrating river water during high discharge events.

2.1 Experimental Field Site

The experimental field site is located in the lower Birs Valley in North-west Switzerland (Fig. 1). It is an alluvial system with a shallow, coarse, gravelly aquifer. The Birs runs 75 km through the Swiss Jura and joins the River Rhine in Basel with a catchment area of 911 km². The mean annual flow near the confluence of the tributary and the Rhine is 15.4 m³ s⁻¹ with storm flows reaching up to 383 m³ s⁻¹.

The aquifer material mainly consists of carbonate gravel, often well-rounded, with variable sediment sorting, and few intercalations of clay or silt layers, resulting in several orders of magnitude variance in hydraulic properties (conductivities between 3×10^{-3} and 10^{-2} ms⁻¹). The carbonate gravel components of the aquifer are of Triassic to Jurassic origin. The lower aquifer boundary (aquitard) consists of Tertiary deposits of the *Elsässer Molasse*, generally with very low hydraulic conductivities (10^{-8} to 10^{-5} ms⁻¹). During average hydrological situations, the thickness of the saturated zone varies between 0.6 and 10 m (smallest next to the river, largest further away).

Most of the water is supplied to the aquifer by river water infiltration, lateral inflow from the local catchment area, and intermittent artificial recharge. Affolter et al. (2010) give a detailed description of the water budget in the study area. Eight extraction wells supply water to approximately 51'000 people, amounting to 5×10^6 m³ drinking water per year.

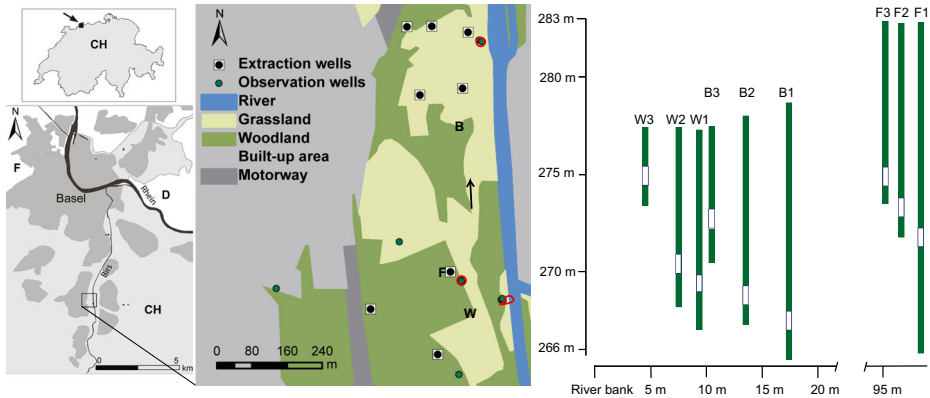


Fig. 1 Map of the study area showing the locations of the wells as well as the depth of the filter sections in the aquifer. There are three clusters: W1, W2 and W3, F1, F2 and F3 and B1, B2 and B3

The time-series used in the analysis were recorded in nine groundwater observation wells close to extraction wells susceptible to contamination due to river water infiltration and the river bank (Fig. 1).

Two measurement set-ups were used: 1) multiparameter instruments recording pressure, temperature, hydraulic electrical conductivity and turbidity (YSI 600OMS V2), 2) three instruments recording pressure, hydraulic electrical conductivity (and temperature) and turbidity (Endress+Hauser Waterpilot FMX167, Indumax CLS50 and Turbimax W CUS41). Additional time-series of spectral absorption coefficient (SAC, measured at 254 nm, Endress+Hauser Viomax CAS51D) and particle density (2–10 μm , Endress+Hauser Lasersens CYM790) were available only for three observation wells (B1, B2 and B3). The time-series were quality checked, e.g., to detect instrument maintenance. The data resolution used in the analysis was 0.5 h.

The criteria for selecting the events were based on the completeness of the time series (no missing values) and the magnitude of the event (river stage). The event had to be detectable in the time series, however not a very large event, as the small to intermediate events were more of a challenge to detect and to test the sensitivity of the methodology.

2.2 Data Analysis

The Self-Organizing Map (SOM, Kohonen (2001)) is an ANN method based on competitive, unsupervised learning. Kohonen (2001) provides explanatory figures of the following section. SOMs have found use in many engineering applications, for example, for monitoring industrial process states or drinking-water quality in distribution networks (Dominguez et al. 2007; Mustonen et al. 2008; Corona et al. 2010). A SOM consists of a set of nodes in a regular 2-D or 3-D grid where every node represents a so-called reference vector. An input data matrix with three vectors each will give a reference vector with three values. Adjacent nodes are very similar to each other. Every row-vector of the data matrix is assigned to the most similar reference vector (node). Once this is accomplished, each observation time (row-vector of the input data matrix) is mapped to the grid based on similarity rather than temporal sequence. Thus, the SOM allows tracking of the system state by visualizing trajectories through time (Dominguez et al. 2007).

A first step in creating a SOM is to determine the size and shape of the basic grid, which will later be used to build the map. The number n of units, or nodes in the grid, was defined according to Vesanto et al. (2000) as a function of the size of the input data set:

$$n = 5 \cdot x^{0.50321} \tag{1}$$

where x is the number of rows in the input matrix (in this case times series of temperature, electrical conductivity and groundwater head, Fig. 2). The dimension of the grid is based on the ratio between the two largest eigenvalues of the correlation matrix of the input matrix. By using the ratio of the two largest eigenvalues as an indication of the size of grid necessary to represent the patterns in the input data, the lattice spacing in the grid becomes approximately uniform (Kohonen 2001).

The reference vectors were initialized based on the two eigenvectors with the largest eigenvalues. The eigenvectors were normalized and multiplied with the square roots of the corresponding eigenvalues to give a set of structural vectors with the same dimension as the number of eigenvalues (number of columns in the input matrix). This resulted in a 2-D grid with 21×12 nodes.

Then the SOM is adapted to the data in an iterative procedure (“training”). The nodes of the grid are initially associated with a set of reference vectors and the grid is gradually reshaped by varying the values of the reference vectors to fit the patterns of the input data the best. Stepwise, the best-matching units (BMUs) are identified for each row vector of the input data matrix, that is, the node with the greatest similarity with the input data row vector, determined by calculating the Euclidean distance between the reference vector associated with each node (m_i , where i =number of nodes) and the row vector of the input data set (Eq. 2) (Vesanto and Alhoniemi 2000).

$$BMU = \operatorname{argmin}_i(\|x - m_i\|) \tag{2}$$

During subsequent training, the nodes surrounding the BMU are updated and modified using a Gaussian neighbourhood function (h_{BMU_i}) (Eq. 3). Components of this smoothing

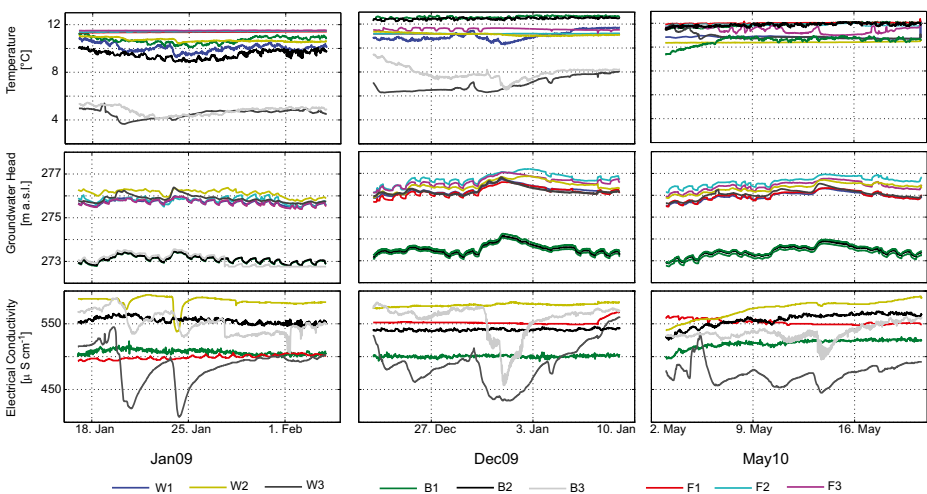


Fig. 2 Time series of electrical conductivity, groundwater head, temperature used in the analysis. The location of the observation wells is shown in Fig. 1

kernel are: the learning rate α , distance from each node (i) to the BMU of each time step (r_{BMUi}), and the width of the kernel σ (radius of the topological neighbourhood). α and σ both decrease with number of training steps (Kohonen 2001).

$$h_{BMUi} = \alpha \cdot \exp\left(-\frac{r_{BMUi}^2}{2\sigma^2}\right) \tag{3}$$

The influence of a node identified as a BMU therefore decreases with increasing distance between each BMU and any other node i .

The updating and modification of the reference vectors ($m_i(t+1)$) in the grid is a function of the previous state $m_i(t)$, the neighbourhood function $h_{BMUi}(t)$ and the input data vector $x(t)$ (Eq. 4). t denotes the training step.

$$m_i(t + 1) = m_i(t) + h_{BMUi}(t)[x(t) - m_i(t)] \tag{4}$$

In the end of the training procedure, any row vector of the input data can be ascribed to a best-matching reference vector or node of the SOM. However, the distance between system state representations in the visualization is not proportional to the dissimilarity in the input data set, e.g., system states that are only vaguely similar may be close to each other in the map giving a false sense of similarity. In addition, the BMUs often represent more than one row vector of input data, i.e., multiple system states. This is difficult to visualize and small differences between system states are disregarded. Some authors combine SOMs with nonlinear projection methods to capture the temporal aspect of system state development (Bernataviciene et al. 2006; Mustonen et al. 2008; Lischeid 2009). To emphasize the temporal resolution and visualization of variation in the original data set, especially over time, the output of the SOMs were further subjected to the Sammon’s mapping algorithm (Sammon 1969). The BMUs derived from the SOM analysis were used to calculate coordinates that formed the basis for the projection space in the Sammon’s Mapping (SM) algorithm.

Sammon’s Mapping is a nonlinear mapping algorithm aiming at preserving the distances in the measurement vector in a 2-D projection (Sammon 1969). SM is useful in determining the shape and density of clusters and the relative differences between these clusters (Kolehmainen et al. 2003). In SM each row vector of the input data matrix is represented by a single point. The points are located so that the distance between them is approximately proportional to the dissimilarities between the respective row vectors. Multivariate time series can thus be visualized by trajectories that follow the succession of BMUs through time (Corona et al. 2010).

The methodology is a form of multidimensional scaling (MDS), whereby the relative distances in the data are preserved to the greatest extent possible (Sammon 1969). The objective of the optimization problem (Eq. 5) is to minimize the Sammon’s Error (E) by using a gradient descent approach based on interpoint distances in the input space (d_{ij}^*) and in the projection space (d_{ij}), whereby i and j represent any two system states. The error (E , Eq. 5) expresses how well the present configuration of the K system state representations in the projection (output) space fits the K points in the input space.

$$E = \frac{1}{\sum_{i < j} d_{ij}^*} \sum_{i < j}^K \frac{[d_{ij}^* - d_{ij}]^2}{d_{ij}^*} \tag{5}$$

To avoid the problematic of local minima, a priori knowledge can be used to initialize the map. The BMUs derived from the SOM analysis provide coordinates for the Sammon's projection. Henceforth, the combination of using the SOM codebook vectors to initiate the Sammon's projection is referred to as SOM-SM.

Measures of fit are given by the topographic error (TE) and by the quantisation error (QE) of the SOM (Postolache et al. 2005). They describe the percentage data vectors for which the 1st and 2nd BMUs are not adjacent cells in the SOM grid (QE), and the average distance between each data vector and its BMU (TE). The squared Pearson correlation coefficient R^2 (coefficient of determination) gives an indication of the truthfulness of the SOM-SM projection by measuring the correlation between the Euclidean distances in the input matrix and those of the projection matrix. The correlation coefficient thus becomes a measure of representative strength of the projection in portraying the input data. These three measures of fit were used to quantify the ability of the methodology to represent the inherent patterns in the input data set.

The time-series were normalized separately for each time period and variable by subtracting its mean and dividing by the standard deviation. The separate normalization prevents variables with large fluctuations (e.g., electrical conductivity) dominating the analysis. The normalized time-series were combined to form one data set, which was used for the SOM-SM. The SOMs were calculated using the SOM Toolbox (Vesanto et al. 2000), developed at the Laboratory of Information and Computer Science in the Helsinki University of Technology. Both steps of the analysis (SOM and SM) were carried out with MATLAB (Mathworks).

The SOM-SM analysis was carried out with 24 time-series: groundwater head and temperature measurements from nine observation wells and electrical conductivity time series from six observation wells (Figs. 1 and 2). Time series of three electrical conductivity sensors could not be used due to measurement errors (W1, F2, F3). The time-series were normalized for each time period to avoid differences based on e.g., seasonality, which were not the focus of this study. Three time periods were used: 17. January - 03. February 2009 (time period *Jan. 09, 864 measurement time steps*), 23. December 2009-08. January 2010 (time period *Dec. 09, 816 measurement time steps*) and 03.-20. May 2010 (time period *May 10, 847 measurement time steps*). They were analysed as one joint data set: the dimension of the resulting matrix was 2527×24 and the map size 21×12 .

3 Results

3.1 Time-Series

Groundwater head, temperature and electrical conductivity time-series from nine observation wells were used in the SOM-SM analysis (Fig. 2). Each time period (Jan 09, Dec 09 and May 10) included at least one high discharge event.

Temperature fluctuated around 11 °C in seven of the nine observation wells in all three time periods. The two exceptions are B3 and W3, both located close to the river in the upper region of the aquifer. The temperature recorded in these observation wells differed between the three time periods: the lowest temperatures for both observation wells were recorded in Jan. 09 (around 4 °C), followed by Dec. 09 (around 7 °C), the temperatures recorded in May 10 are not distinguishable from the other groundwater temperature time series. Nevertheless, small fluctuations (0.75–1.5 °C) were observed during high discharge events.

High discharge events were more easily recognizable as increases in groundwater head (Fig. 2). The smaller, daily fluctuating signal observed in groundwater head can be attributed to the pumping regime in the eight groundwater extraction wells in the study area (Fig. 1). This signal is not detectable during the high discharge events, suggesting it was either masked or extraction was stopped. Electrical conductivity showed little variation between time periods, fluctuating between 400 and 600 $\mu\text{S cm}^{-1}$. Characteristically, el. conductivity decreased following a high discharge event (in relation to increase in groundwater head). In some cases, a small increase was observed first (e.g., W3).

Of the three parameters, el. conductivity showed the strongest response to the high discharge event in the observation wells located close to the river, while those further away (F1, B1 and B2) showed little or no response.

3.2 SOM-SM Analysis

The SOM-SM methodology aims to reduce the amount of time required to obtain useful and relevant information for the management of groundwater extraction wells. It is designed to rapidly identify any potentially hazardous changes of the system state.

The topographic error TE and quantization error QE of the self-organizing map were 0.035 and 1.775, respectively. The correlation coefficient for Sammon's projection R^2 was 0.836.

The visualization of the SOM-SM analysis result is shown in Figs. 3, 4 and 5. There is one point per system state, i.e., 2527 points representing 2527 observations, as each point summarises one row and all 24 variables (columns) of the data matrix. The axes (Sammon's **MAP 1** and Sammon's **MAP 2**) define a two-dimensional Cartesian coordinate system, which is used to show the proximity and distance between individual observations (represented by points in Figs. 3, 4 and 5) based on the SOM-SM analysis. The lines indicate subsequent observations. The shading is used to show the measured values of different parameters in relation to the multivariate results of the SOM-SM analysis.

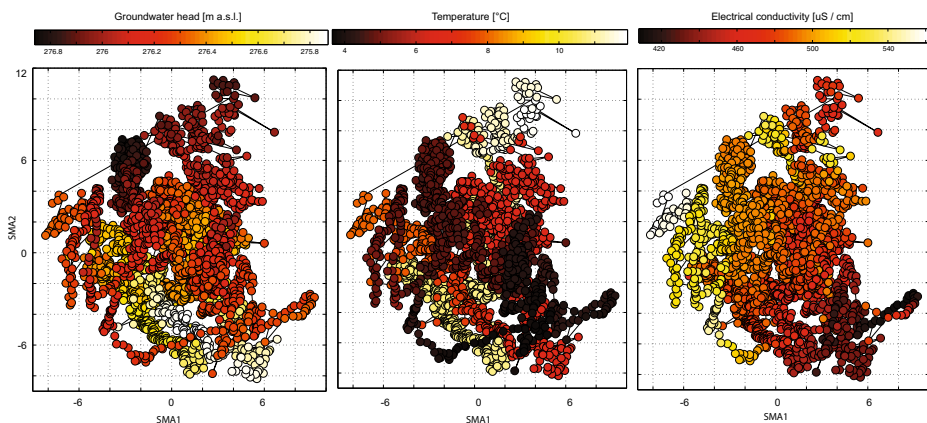


Fig. 3 SOM-SM projection results (system states), whereby the shading of the points reflects groundwater head, electrical conductivity and temperature data from well W3 highlighting different situations. Each point symbolises the similarity of a system state relative to the rest of the system states encountered during each time period evaluated

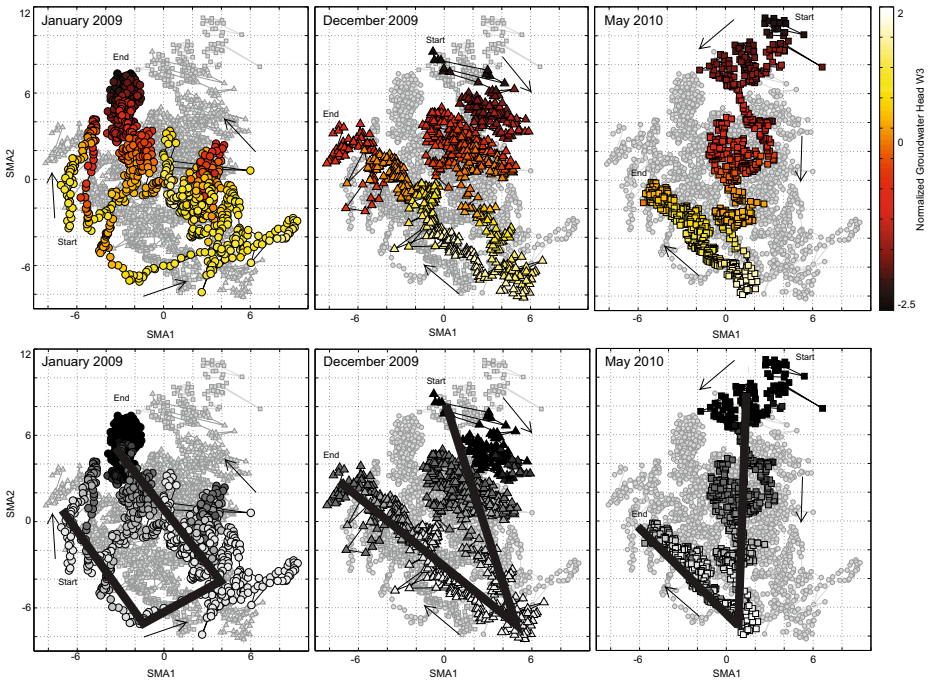


Fig. 4 SOM-SM projection results for each time period, the shading (groundwater head in W3) of the points is adapted to the maximal range of each time period. The lower three subfigures indicate the u- and v-shaped trajectories representing system change over time. The arrows show the direction of the trajectory of system state for each time period

3.3 General Pattern and Overview

Figure 3 highlights the overall shape of the projection, which is roughly triangular. The majority of the points were located in the upper half of the projection space, with fewer points in the lower region. The shading in Fig. 3 shows the values of three different parameters used for the calculation (subset of groundwater head, temperature and electrical conductivity from observation well W3). Observation well W3 is located closest to the river and shows a strong response to high discharge events (Fig. 4). Groundwater head and electrical conductivity show a trend in the shading from the top left to the bottom right. High groundwater head levels and low electrical conductivity measures were located in the lower region of the projection. The temperature shading did not show such a pattern, but reflected the different time periods used in the analysis as distinguished by differing temperatures in W3 (Fig. 2).

3.4 Temporal Development During Events

Figure 4 shows the results for each of the three time periods separately (non-active time periods are indicated in grey in the background). The point shading is given by normalized groundwater head in W3. The lighter shades are generally in the lower region of the projection, indicating higher groundwater head during these system states.

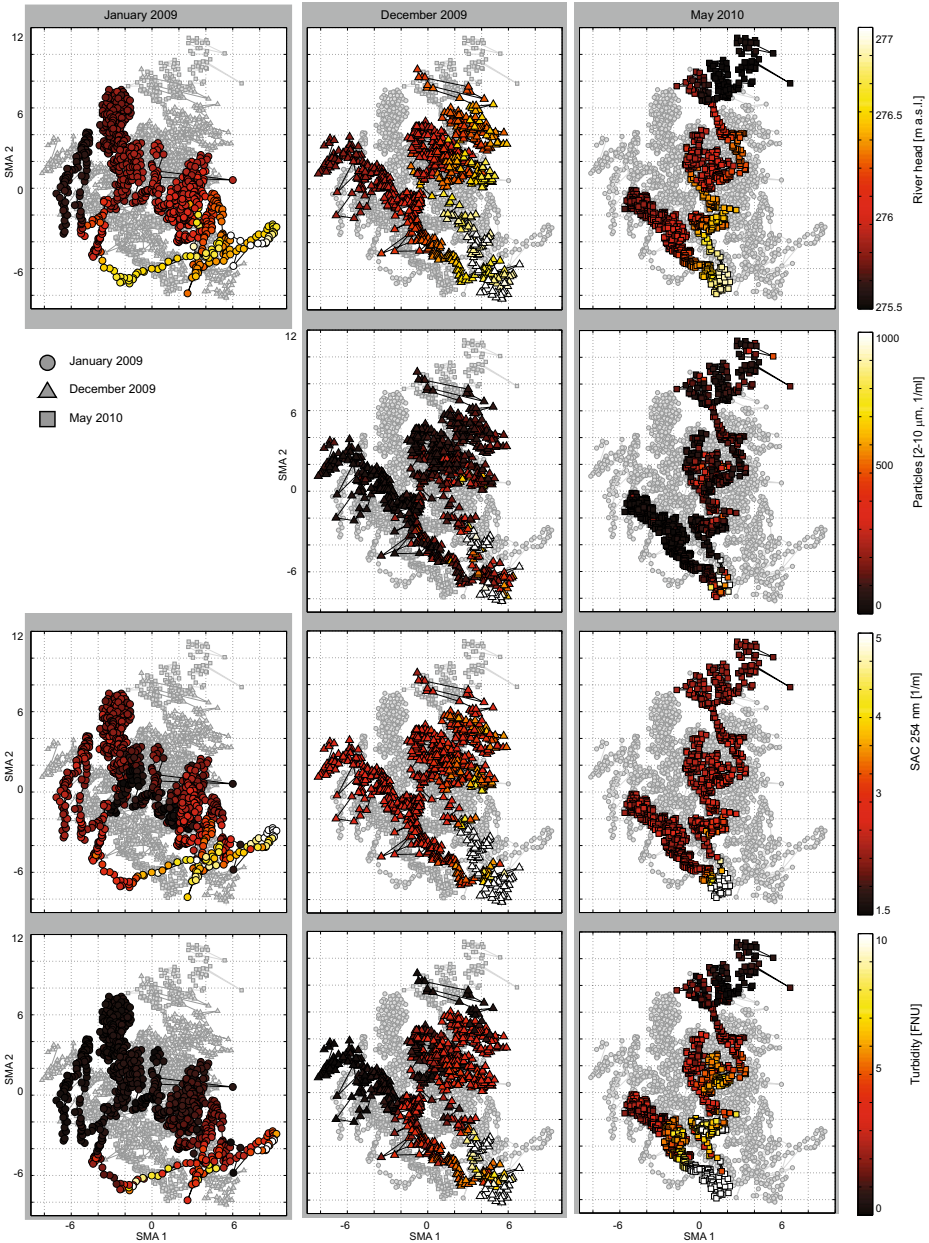


Fig. 5 Projection of SOM-SM analysis with independent parameters (river stage, turbidity, spectral absorption coefficient (SAC) and particle counts). The latter three were measured in B3 (Fig. 1). Lighter shades indicate higher parameter values associated with high discharge events. Each subplot shows the relation between system state (SOM-SM analysis) and an independent variable during one time period

The development of system state over time is similar for all three events: each time period starts in the upper half of the plot, moves towards the lower right quadrant and back up to the upper half, creating a u-, or v-shaped trajectory (Fig. 4).

Not only are the system states of the three time periods similar, as highlighted in Fig. 3, but the dynamics during high discharge events (trajectories) are also similar, as highlighted in Fig. 4, suggesting a pattern that can be easily identified and interpreted.

3.5 Identifying Hazardous System States

Figure 5 shows the same analysis and projection, however the shading is taken from parameters not used in the SOM-SM analysis (river head, particle density, spectral absorption coefficient (SAC) and turbidity). They are used to highlight potentially hazardous situations. The shading of particle density, SAC and turbidity is based on the time series recorded in B3. The fourth parameter used to identify high discharge events in the SOM-SM analysis was river stage, which was included in this study as the drinking water suppliers currently use it as a trigger to manage the extraction regime. The shading in the projection shows a pattern similar to the ones seen in Figs. 3 and 4 (groundwater head and electrical conductivity): higher river stage, turbidity, SAC measurements and particle counts (lighter shades) are located in the lowest area of the projection space, while the majority of the points are darker and located in the upper areas of the projection space. Despite the similarity between time periods in terms of point location and system development during high discharge events, SAC and turbidity highlight a potential characteristic of each high discharge event: the change between darker and lighter shades during the event in January 2009 is drawn out considerably longer than in May 2010, where only the tip of the projection shows lighter shades. The lighter shades, and thus the higher values of river stage cover a larger extent of the dip in the projected points than the other indicator parameters do, suggesting an initial rise in river head and subsequent infiltration of river water with higher turbidity and SAC levels. The particle time series from Jan 09 were not included due to large measurement errors.

4 Discussion

In this study, time-series of three groundwater parameters were analysed to identify patterns reflecting changes in system state. The three high discharge events analysed differed from each other in terms of magnitude and velocity in which the measured groundwater parameters (groundwater head, electrical conductivity and temperature) responded. Groundwater head time-series showed a strong response to all three events and appeared strongly influenced by pressure wave propagation through the aquifer linked to drinking water extraction, river water infiltration and rapid river stage rises (Fig. 2). Methods like principal component analysis have shown that multiple influences, such as groundwater extraction and river-stage fluctuations, can be superimposed and may mask each other, depending on their magnitude and the distance of each observation well to the source of disturbance (Lewandowski et al. 2009; Page et al. 2012 also seen in Fig. 2). The observed decrease in electrical conductivity suggests dilution caused by infiltrating river water with lower ion concentrations than the groundwater. Some events are characterized by an initial increase in electrical conductivity, e.g., during the first event in Jan 09 or May 10 (Fig. 2). Both effects, the in- and the decrease of electrical conductivity, were not observed in the wells further away from the river (Fig. 2, e.g., B1). A number of reasons may have led to the dampening and loss of the signal over distance, e.g., advection, assimilation, or, potentially, the non-conservative behaviour of the parameter (Cirpka et al. 2007), or a varying influence of the regional groundwater flow component as

not all observation wells are equally strongly influenced by river-groundwater interactions. The combination of these effects makes a generalized interpretation of electrical conductivity values difficult and requires case-by-case consideration. Temperature showed the least fluctuations during each time period and only minor reactions to the high discharge events. However, seasonality and river temperature strongly influenced groundwater temperature close to the river (Fig. 2). The strength of temperature as an indicator for river water infiltration is limited in the experimental site when river and groundwater temperatures are similar, which usually occurs at least twice a year (in the spring and autumn).

Despite these limitations, the SOM-SM analysis and projection indicated a characteristic behaviour of the system throughout three different time periods and high discharge events (Fig. 4). The use of independent parameters (river head, turbidity, SAC and particle density) supported the indication of characteristic behaviour of the system under observation. High turbidity values are often associated with surface water infiltration and potential contamination and are used by drinking-water supply-plant operators for water quality control (Iglesias et al. 2014). Similarly, the elevated SAC values in the groundwater observation wells indicated an increase in organic matter content. As SAC has been used as an indicator for faecal contamination in other studies (Stadler et al. 2010), high SAC measurements are characteristic of infiltrating surface water and critical system states. Increased particle density is another parameter that has been used as an indicator for faecal bacteria contamination (Auckenthaler et al. 2002; Pronk et al. 2007). These additional groundwater quality parameters and river stage are strongly associated with river water infiltration and can be used as indices of critical situations, where drinking-water extraction wells are vulnerable to river-borne microbial contamination.

The indicators for potential contamination showed the highest values in the lower right area of the SOM-SM projection, thus the critical system states during each time period were located in the same area of the projection and indicated a similar behaviour during high discharge events. The consistent position of points representing critical system states in the SOM-SM projection, as validated by elevated indicator parameter values (Fig. 5), suggests that the combination of groundwater head, temperature and el. conductivity measurements can be used to identify potentially hazardous situations. The temporal dynamic of system state can be illustrated by considering the trajectories over time, as the lower density u- and v-shaped trajectories reflect the different system states during the course of the three high discharge events (Fig. 4). Furthermore, all three high discharge events showed a similar pattern (Figs. 3 and 5). These two pieces of information (length of event and pattern) enable a relatively early initial warning (before indicator parameters react) and an indication of the development stage of a high discharge event. In addition, the definition of the time before system state returned to a previous level is a further estimate that can be drawn from the SOM-SM analysis and that provides the drinking water supply manager with additional information on actual resource quality.

The effectiveness of the SOM-SM method increases with the magnitude of change in system state. Rapid infiltration and associated changes in measured parameters will evoke large changes in the SOM-SM projection, which can then be used to issue a warning to the water supplier and provide sufficient time for necessary management options to be taken, for example reduction of extraction or switch to other extraction wells. However, events with only small or gradual changes could be missed if the minimal change between measurement times required for detection is not adequately defined. The onset and duration of potentially hazardous situations require further validation using a range of events, including large and small, as well as rapid and gradual changes in system state. These two considerations (minimal change threshold as well as onset and duration) can be part of a fine-tuning process, dependent on the system under consideration

and the management options, e.g., the number of active wells or the pump stages. As demonstrated by Dominguez et al. (2007), the SOM-SM approach to process supervision can be an important part of a decentralized and distributed management system and can be included in a quality assurance HACCP-approach (Hazard Analysis and Critical Control Point). The approach can be fully automated: 1) *online instrumentation* reporting to a database, 2) *database* connected to all parameters recorded in the production process, including water extraction and treatment information, 3) *analysis software* integrated into the operational system of the water suppliers, 4) *warning system* adapted to the needs of the user.

5 Conclusions

Traditionally, in the study area and for water suppliers with similar situations, the approach to managing the water quality of the extracted groundwater has been based on a univariate threshold consideration relying on the expertise (and availability) of the operator of the water supply to detect potentially hazardous situations. The novel approach in this study combines online measurements, data management and automated pattern recognition to achieve an adaptive management solution that provides 24 h surveillance. The online identification of potential contamination of groundwater extraction wells is a step towards an adaptive approach to groundwater management, allowing a temporally-differentiated extraction plan. The SOM-SM analysis provides a means by which changes in system state can be quantified and integrated into the operational system of water suppliers.

Combining the SOM and Sammon's Mapping to a multivariate proxy analysis provided the basis for a timely identification of critical system states. The analysis using groundwater head, electrical conductivity and temperature was able to identify changes in system state indicating the arrival of rapidly infiltrating river water and with it, potential contamination. Independent time-series (turbidity, river stage, SAC and particle density) validate the patterns during high discharge events and identify situations, when the system is strongly influenced by infiltrating river water, thus representing potential contamination of the extraction wells. The resumption of extraction requires further validation.

As the approach can be automated after validation, it can be integrated into the operational management of riverine groundwater extraction well fields to provide a continuous online assessment of actual system state. The time of detection of infiltrating river water is based on changes occurring in the measured time-series and can be used to issue warnings to the water supplier. Besides issuing warnings concerning the onset of critical system states, information on when extraction can be resumed is an important element of the management process.

Such an online, automated system can supply additional degrees of freedom to the management of extraction wells near rivers, which have high-quality water during most of the year, but where water quality is severely compromised during high discharge events. The complex task of identifying hazardous situations for drinking-water quality in a heterogeneous environment can thus be simplified by using a SOM-SM proxy analysis as part of quality management and assurance.

Acknowledgments The authors thank Stefan Scheidler from the Applied and Environmental Geology Group, University of Basel, Endress+Hauser Metso AG and the Waterworks Reinach and Surroundings (WWRuU) for their support. This work was funded by the Swiss Innovation Promotion Agency CTI (projects number 8999.1 PFIW-IW and 12611.2 PFIW-IW) and the Freiwillige Akademische Gesellschaft Basel.

References

- Affolter A, Huggenberger P, Scheidler S, Epting J (2010) Adaptive groundwater management in urban areas: effect of surface water-groundwater interaction using the example of artificial groundwater recharge and in- and exfiltration of the river Birs (Switzerland). *Grundwasser* 15(3):147–161
- Auckenthaler A, Raso G, Huggenberger P (2002) Particle transport in a karst aquifer: natural and artificial tracer experiments with bacteria, bacteriophages and microspheres. *Water Sci Technol* 46(3):131–138
- Bernataviciene J, Dzemyda G, Kurasova O, Marcinkevicius V (2006) Optimal decisions in combining the SOM with nonlinear projection methods. *Eur J Oper Res* 173(3):729–745
- Camplani M, Cannas B, Fanni A, Pautasso G, Sias G, Sonato P, Asdex Upgrade Team (2009) Tracking of the plasma states in a nuclear fusion device using SOMs. In: *Engineering Applications of Neural Networks*. Brown DP, Draganova C, Pimenidis E, Mouratidis H (eds.) *Communications in Computer and Information Science* 43, 430–437
- Cirpka OA, Fienen MN, Hofer M, Hoehn E, Tessarini A, Kipfer R, Kitanidis PK (2007) Analyzing bank filtration by deconvoluting time series of electric conductivity. *Ground Water* 45(3):318–328
- Corona F, Mulas M, Baratti R, Romagnoli JA (2010) On the topological modeling and analysis of industrial process data using the SOM. *Comput Chem Eng* 34(12):2022–2032
- Dash RR, Prakash EVPB, Kumar P, Mehrotra I, Sandhu C, Grishek T (2010) River bank filtration in Hardiwar, India: removal of turbidity, organics and bacteria. *Hydrogeol J* 18(4):973–983
- Dominguez M, Fuertes JJ, Reguera P, Diaz I, Cuadrado AA (2007) Internet-based remote supervision of industrial processes using Self-Organizing maps. *Eng Appl Artif Intel* 20(6):757–765
- Fuertes JJ, Dominguez M, Reguera P, Prada MA, Diaz I, Cuadrado AA (2010) Visual dynamic model based on self-organizing maps for supervision and fault detection in industrial processes. *Eng Appl Artif Intel* 23:8–17
- Iglesias C, Martinez TJ, Garcia Nieto PJ, Alonso Fernandez JR, Diaz Muniz C, Pineiro JJ, Taboada J (2014) Turbidity prediction in a river basin by using artificial neural networks: a case study in Northern Spain. *Water Resour Manag* 28(2):319–331
- Kohonen T (2001) *Self-organizing maps*. Springer
- Kolehmainen M, Ronkko P, Raatikainen A (2003) Monitoring of yeast fermentation by ion mobility spectrometry measurement and data visualisation with self-organizing maps. *Anal Chim Acta* 484(1):93–100
- Lewandowski J, Lischeid G, Nützmann G (2009) Drivers of water level fluctuations and hydrological exchange between groundwater and surface water at the lowland River Spree (Germany): field study and statistical analysis. *Hydrol Process* 23(15):2117–2128
- Lischeid G (2009) Non-linear visualization and analysis of large water quality data sets: a model-free basis for efficient monitoring and risk assessment. *Stochastic Envir Res Risk Assess* 23(7):977–990
- Mustonen SM, Tissari S, Huikko M, Kolehmainen M, Lehtola MJ, Hirvonen A (2008) Evaluating online data of water quality changes in a pilot drinking water distribution system with multivariate data exploration methods. *Water Res* 42(10–11):2421–2430
- O'Flynn B, Regan F, Lawlor A, Wallace J, Torres J, O'Mathuna C (2010) Experiences and recommendations in deploying a real-time, water quality monitoring system. *Meas Sci Technol* 21(124004):10
- Page RM, Lischeid G, Epting J, Huggenberger P (2012) Principal component analysis of time series for identifying indicator variables for riverine groundwater extraction management. *J Hydrol* 432–433:137–144
- Postolache OA, Silva Girão PMB, Dias Pereira JM, Geirinhas Ramos HM (2005) Self-organizing maps application in a remote water quality monitoring system. *IEEE Trans Instrum Meas* 54(1):322–329
- Pronk M, Goldscheider N, Zopfi J (2007) Particle-size distribution as indicator for faecal bacteria contamination of drinking water from karst springs. *Environ Sci Tech* 42(24):8400–8405
- Regli C, Rauber M, Huggenberger P (2003) Analysis of aquifer heterogeneity within a well capture zone, comparison of model data with field experiments: a case study from the river Wiese, Switzerland. *Aquat Sci* 65(2):111–128
- Sammon J (1969) A nonlinear mapping for data structure analysis. *IEEE Trans Comput* C18(5):401–409
- Stadler H, Klock E, Skritek P, Mach RL, Zerobin W, Farnleitner AH (2010) The spectral absorption coefficient at 254 nm as a real-time early warning proxy for detecting faecal pollution events at alpine karst water resources. *Water Sci Technol* 62(8):1898–1906
- Stefanovic N, Radojevic I, Ostojic A, Comic L, Topuzovi M (2015) Composite Web information system for management of water resources. *Water Resour Manag* 29:2285–2301
- Taylor R, Cronin A, Pedley S, Barker J, Atkinson T (2004) The implications of groundwater velocity variations on microbial transport and wellhead protection—review of field evidence. *FEMS Microbiol Ecol* 49(1):17–26
- Vesanto J, Alhoniemi E (2000) Clustering of the self-organizing map. *IEEE Trans Neural Netw* 11(3):586–600
- Vesanto J, Himberg J, Alhoniemi E, Parhankangas J (2000) SOM toolbox for Matlab 5. Helsinki University of Technology, Finland
- Zektser IS, Everett LG (2004) Groundwater resources of the world and their use. UNESCO IHP-VI, Series on Groundwater No.6