Online Water-Quality Monitoring based on Pattern Analysis

Rebecca M. Page[1]1, Daniel Waldmann[1], Achim Gahr[2]
[1,2] Endress+Hauser (Schweiz) AG, Kägenstrasse 2, 4153 Reinach, Switzerland
[3] Endress+Hauser Conducta GmbH+Co. KG, Dieselstrasse 24, 70839 Gerlingen, Germany
[1]*rebecca.page@ch.endress.com*

ABSTRACT

To date, drinking water quality monitoring frequently relies on a threshold-based approach coupled with occasional manual sampling for reference analysis and as evidence for legal requirements concerning the water quality. However, the increased availability of online measurements provides a good basis for an adaptive approach to high-resolution monitoring of water quality. In this case study, patterns in water quality of limestone springs were identified using multivariate analysis and artificial neural networks. Self-organizing maps were used to calculate system states based on six online parameters (spring discharge, turbidity, pH, el. conductivity and spectral absorption at 254 nm). A non-linear Sammon projection highlighted the relationship between the different system states, rendering a basis for the quantification of change occurring during the observation period in December 2015 - January 2016. The multivariate approach highlighted different phases during an event based on the relative location in a scatter plot and on the xy distance between two system states based on consecutive measurements. As this approach does not require the definition of thresholds and considers actual changes in system state, it is applicable to complex systems and adaptive management strategies.

Keywords: Drinking water quality; online early warning system; multivariate pattern analysis

1 INTRODUCTION

In Switzerland, as in many other countries, drinking water is mainly extracted from porous aquifers or collected from karst springs. In both cases, the local and regional geological settings strongly influence both water composition and quality, including chemical, physical and biological components. Furthermore, activities and industries in the catchment area, such as agriculture and construction sites, are often drivers for variability in water quality and quantity. In combination, the geological settings, activities, seasonal practices and event-driven dynamics, lead to spatial-temporally heterogeneous patterns in water quality [1]. The nature of the combined effect of these influences on water quality is difficult to predict, even more so when several influences are superimposed and conflicts of usage arise. Under such circumstances, an effective management of drinking water supply systems requires knowledge on the processes influencing the water quality (and quantity) both over time and space. One set of approaches is based on an offline analysis of groundwater providing a basis for long-term management of the drinking water resource. For example, modelling approaches have been able to identify karst springs with elevated vulnerability in terms of microbial contamination [2] or define protective areas around groundwater extraction wells to reduce the likelihood of contamination [3]. Short-term management of drinking water resources aims to avoid contamination using online measurements and rapid decisions leading to changes in the operating system of the water supply, e.g. changing the flow field of the groundwater by reducing groundwater extraction or increasing the level of disinfection (drinking water treatment).

Only 30% of the drinking water in Switzerland is treated, a further 30% only runs through a single treatment step (UV-disinfection). 10% is filtered then disinfected and the remaining 30%, including all of the water originating from the lakes (approx. 20% of total water production of 0.94 billion m^3 in 2010) is treated in a multi-step process. There is a general consensus that additional treatment steps should be avoided where possible. Approximately 35% of the total drinking water consumed in 2010 was delivered by five large suppliers [4]. Most of the water suppliers, however, delivered water to communities with less than 10'000 inhabitants, accounting for 20% of the total drinking water consumption in 2010 [4]. Many of these smaller water suppliers are run by the local community and operated by part-time employees, usually without training in data analysis.

To date, water quality monitoring frequently relies on a threshold-based approach coupled with occasional manual sampling for reference analysis and evidence for legal requirements concerning the water quality. However, the increased availability of online measurements as well as additional and aggregated data, e.g. integrated from different sources, provides a good basis for an adaptive approach to high-resolution monitoring of water quality and with this, a significant support for the operators of the drinking water producers. Considering the natural heterogeneity of untreated water quality, structural organization and the desire to maintain a minimal level of treatment, the water suppliers are often confronted with the difficult task of making rapid decisions to avoid contamination. The following case study presents an adaptive approach to monitoring changes in water quality based on multivariate pattern analysis of online measurements. The aim was to design a method to capture variability and change in heterogeneous systems to assess potential contamination without the use of thresholds. Furthermore, the method needed to have the potential to be automated and integrated into online management systems and to be operational without any training in data analysis.

2 STUDY SITE

The data in this case study was recorded in a drinking water supply system located in the limestone hills in NW Switzerland. These Jurassic Karst systems provide ample high-quality drinking water, but are vulnerable to contamination [1, 2]. The springs in this case study, as many other karst springs, were characterized by a high degree of dynamics covering different scales including seasonal and operational-related fluctuations. The discharge from seven springs was converged and treated before being pumped into a reservoir, which fed into the local distribution network. When turbidity exceeded 1 FNU, the water was discharged into the stream; otherwise, the water was treated by UV-disinfection. The water supplier monitored spring discharge (combined discharge only) and turbidity. In addition to the measurements already in place, a bypass panel was placed in the pump station to monitor the following additional parameters: electrical conductivity (EC), spectral absorption coefficient at 254 nm (SAC254), pH and temperature. All measurements were made in the combined discharge before treatment. The time-series shown in Figure 1 were recorded between 30.12.2015 and 11.01.2016. The temporal resolution of the data used in the analysis and shown in Figures 1 and 2 was 10 minutes. All sensors were manufactured by Endress+Hauser. The measured data were transmitted to the operating system of the water supplier via a data logger (Liquiline, Endress+Hauser). The data were managed in an SQL database and visualized as part of the water suppliers Supervisory Control And Data Acquisition (SCADA) system.

3 RESULTS

3.1 Univariate Temporal Dynamics

The seven springs connected to the water supply had a common catchment area with relatively homogenous land use. The parameters recorded in the springs were thus subject to very similar precipitation events and land surface pressures and processes. The data shown in Figure 1 have not been treated; the outliers on the 10th January were be assigned to a data transmission irregularity, as both SAC254 and temperature dropped to zero. Spring discharge showed three clear increases during the observation period, with peak discharges reaching 400 l/m, which was a four-fold increase over initial conditions. Turbidity and SAC254 also showed three increases. El. conductivity showed three characteristic decreases as surface water with a lower ionic strength reached the springs [1].

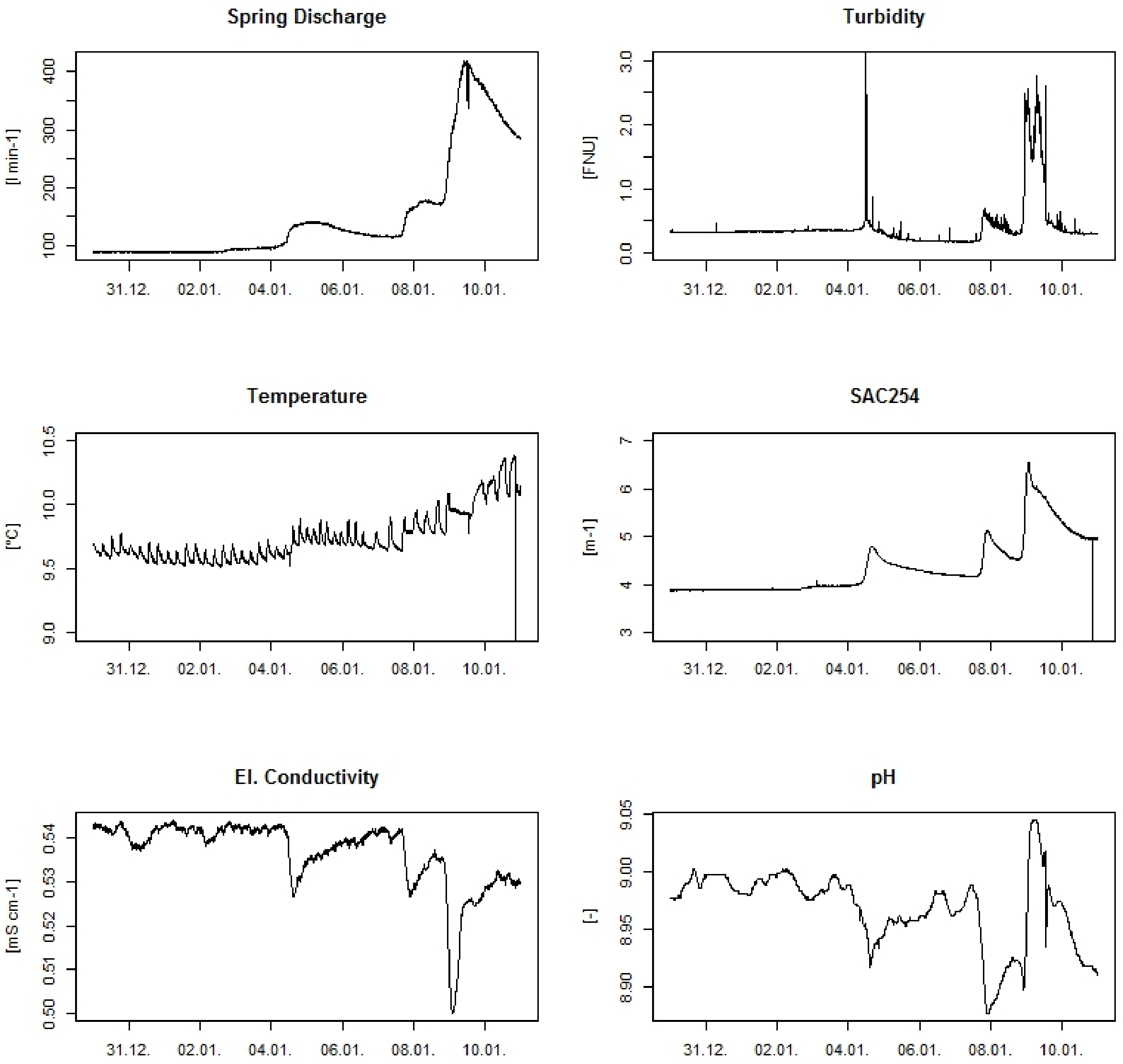

Figure 1 Time-series of untreated measurements of six parameters used to monitor water quality of a karst spring. SAC254 is the spectral absorption coefficient at 254 nm. The measurement resolution was 10 minutes.

The fluctuations visible in the temperature time-series showed a daily pattern related to the operation of the pumps in the room where the panel was located. The hot air from the pumps heated the water in the thin tubes of the panel by approximately 0.2- 0.4 °C. As pumping ceased during the event on the 8th/9th January, the daily temperature fluctuations also ceased. There was an indication of further operational influence around 1 pm on the 9th January, whereby all parameters but el. conductivity appeared briefly affected (Figure 1). pH showed two decreases during the observation period. There was a third decrease around midnight on the 8th/9th January. However the initial drop was followed by a rapid increase in pH.

Following infiltration, surface water travels along different paths through limestone karst systems before reaching the springs. The routes the infiltrated water runs can through vary between fast propagation in large underground rivers to the slow movement through the limestone matrix [2]. In this case study, besides the route through the karst system, the location where the discharged water converged with water from other springs also had an effect on the shape of the time-series, i.e. how long it took for the water to reach the monitoring point after exiting the karst system at the spring. The different travel times through a karst system caused the time-series to show 'piggybacking'. This phrase refers to a secondary peak reflecting the arrival time of infiltrated surface water with a longer travel time. 'Piggybacking' was most pronounced in the turbidity measurements, whereby two peaks in turbidity levels were visible (event 8th and 9th January, Figure 1). SAC254 also showed a small indication and the sharp increase in pH coincided with the turbidity 'piggyback'.

3.2 Multivariate System State

The multivariate data analysis in this case study was based on a proxy approach to monitoring variability in water quality. Self-organizing maps are a form of artificial neural networks, which can be used to identify and analyze patterns in complex system data [5]. In this application they were used to: a) reduce the dimension of the data set, b) calculate actual system states based on online time-series and c) lay emphasis on the similarities and dissimilarities between system states. The output of the self-organizing maps is generally a two dimensional grid with the system states assigned to an area in the map [5]. To increase the visibility of the similarities and the dynamics of the system as it moves from one state to another, the 'best-matching units' (output of the self-organizing map) were projected into a two-dimensional space. A nonlinear Sammon projection [6] was used to project the system states into a space defined by the variability in the data set. This method minimizes the difference between the distributions of the time-related 'best-matching units' and the original input.

Based on time-series such as the ones shown in Figure 1, the multivariate approach identified and evaluated changes in system state as part of a pattern of water quality. The six parameters (spring discharge, turbidity, temperature, SAC254, el. conductivity and pH) were simultaneously considered for each observation time (corresponding to one point in Figure 2).

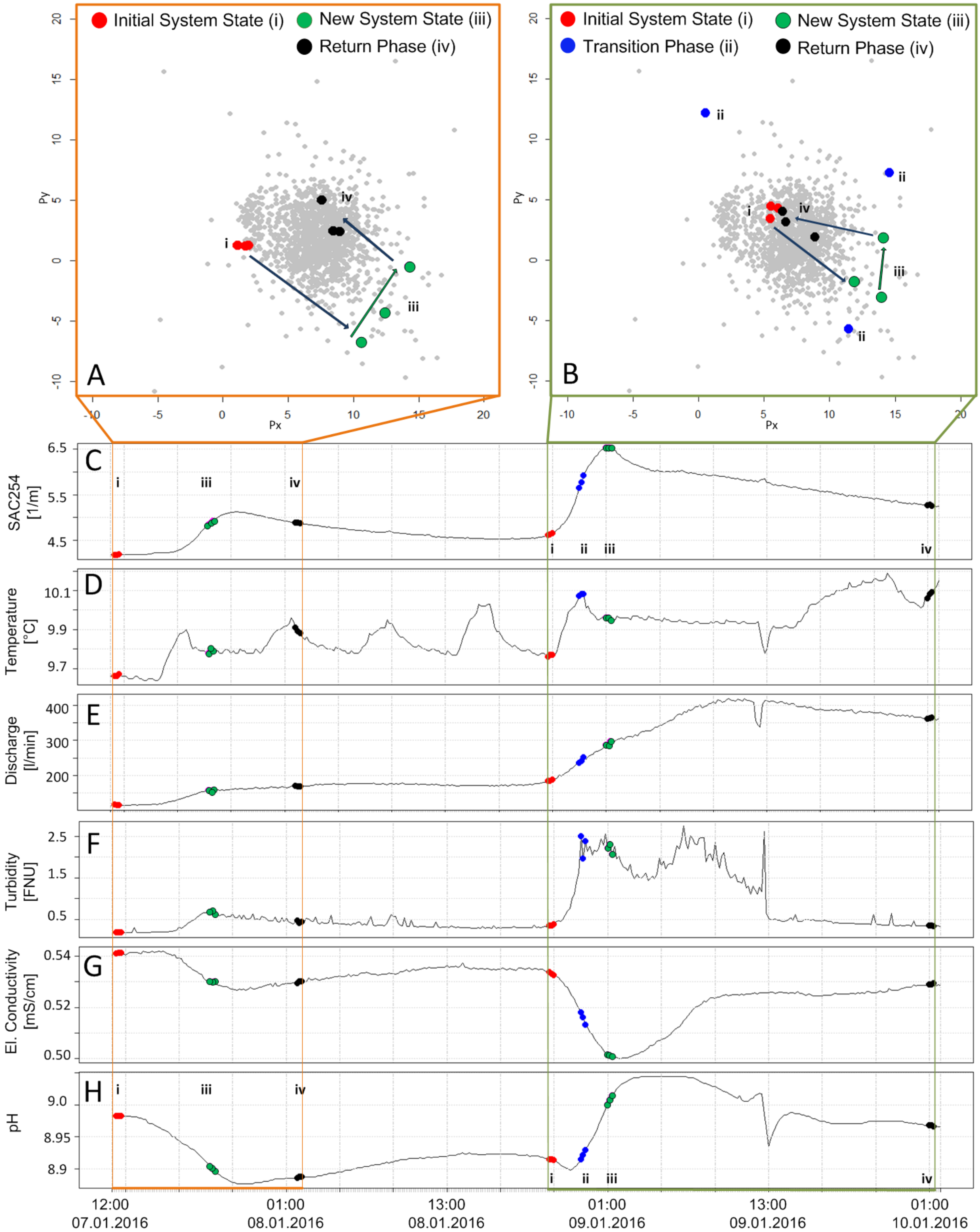

Figure 2 Subplots A and B show the projected results of the multivariate self-organizing map. The projection axes refer to system state space defined by the distribution of the underlying data. Two events between the 7th and 10th January 2016 were highlighted to show different phases through which the system state moved during an event. The subplots C to H show the input data used for the calculation. The different phases are highlighted by three consecutive system states each (three phases during event one and four phases during event two).

The relative values of the input data define the state of the system at the time of the measurement in relation to the other system states, which were calculated from the other measurement times in the observation period. Figure 2 shows the distribution of the system states in relation to each other. Each point in subplots A and B represents one observation time and all six parameters. The coloured points in subplots A and B correspond to different phases the system passed through during two events in January 2016 (7th/8th and 8th/9th January). The axes represent the projection limits and are dependent on the overall heterogeneity of the system states occurring in the data set. For each phase of each event, three points were highlighted in Figure 2. Each cluster of three was calculated from consecutive measurements with a 10-minute resolution between the measurements. The initial phase (i, red points) was characterized by little variability: points are close to each other in the XY-scatter plot. Subplot B (event on the 8th and 9th January) includes a transition phase (ii, blue points), whereby considerable variability occurred in the system: the blue points are spread over a large area of the plot. The new system state (iii, green points) was characterized by more sparse distribution (larger distances between the points). This suggests fewer similar states in the data set (overall few points in the area) and greater variability in the underlying data. The last phase (iv, black points) was characterized by less variability as the points were closer to each other, indicating more inert system states as there was less change over the 30 minutes during which the data were recorded. Following the first event (7th / 8th January, subplot A), the system did not return to the previous state: the red and the black points are not in the same location in the plot. However, as shown in subplot B, the system returned to a similar, not identical state following the second event: the black points (iv) are in a similar location to the red points (i).

4 DISCUSSION

The data set recorded in the water discharged from seven karst springs in January 2016 in NW Switzerland clearly showed two characteristics of a complex system: a) operational influences: daily temperature fluctuations due to exhaust air from the pumps, b) superposition: 'piggybacking' visible in the turbidity data linked to different flow paths and travel times from surface infiltration to arrival at the observation point in the pump house. Each individual parameter bears the potential as an indicator for surface water infiltration and thus as an indicator for potential contamination. For example, SAC254 often represents organic matter from the surface carried in with infiltrating water. However, the reactions of individual parameters may vary over time and between water supply systems. Beside parameter-specific variations, the historical sequence of events, both hydrological and operational, can influence the histograms of the measured parameters, as illustrated by the 'piggybacking'. The interpretation of the individual time-series requires process knowledge of both the hydrogeological and the operational aspects of each water supply.

The multivariate approach consisted of combining all six parameters to calculate the system state for each observation time. The projection of the results into a 2D plot highlighted the overall pattern of system states representing water quality. The location of each individual system state in the scatter plot could then be assessed based on the overall distribution of system states from an observation period, as well as the system states from immediately prior and posterior measurements. The multivariate approach highlighted new system states located in sparsely covered areas of the two-dimensional plot (Figure 2). Gradual shifts in system state appeared as trends (e.g.

iii, green points in Figure 2), whereas rapid shifts moved from one side of the plot to another (e.g. during the transition phase in subplot B, Figure 2). Furthermore, the dynamics of a system could be followed throughout the event, i.e. each phase of an event is captured (onset, peak and 'return to normal').

The multivariate analysis method based on self-organizing maps and projection thus summarized the dynamics in the original data set and reduced the amount of information to be interpreted. When viewed as a scatter plot, the interpretation of the results of the multivariate method also requires familiarity with the analysis method. However, the quantification of the variability and dynamics can be simplified by looking at the sparseness of the points and their temporal succession (distance in the xy projection between two measurements). This quantification could be automated and integrated into a SCADA system, e.g. as an alarm index indicating the current system state.

This data-based approach to pattern analysis does not define parameter thresholds, but assesses the system state based on the multivariate data set and its behavior over time. In this form, the neural network required little input information, as the training was based on data from within each observation period. The motivation for this solution was threefold: Firstly, differentiating between critical and non-critical change in the system over time was a main identifier of potentially hazardous system states. Secondly, the relationship between individual system states with other system states could be captured by the heterogeneity of the results and visualized by the xy extent of the scatter plots (Figure 2, subplots A and B). Thirdly, the simplified quantification bears a high potential to transfer this approach between different drinking water suppliers and to integrate it into SCADA systems. This approach has been tested in several drinking water supply systems with different settings and using different combinations of parameters [1,7]. Based on these characteristics, the approach allows for fluctuations inherent to the karst spring characteristics, such as seasonal or regular operational influences and can be used as an online early-warning system to detect quality-relevant changes. Furthermore, the user, human or an automated system, can rapidly identify similar system states and observe changes as they occur, i.e. the beginning and end of contamination events or baseline shifts. As the extent of an event and its effects on any of the parameters recorded may shift and change over time it was important to select a method that does not rely on fixed thresholds to manage a dynamic system at the interface between a heterogeneous natural system and a regularized technical process.

5 OUTLOOK

The proxy-based approach applied in this case study makes use of a methodology that allows for change and thus provides an adaptive approach to monitoring the dynamics in water quality. To ensure the system under observation is depicted correctly, the multivariate analysis depends on the selection of the monitored parameters, i.e. in combination they need to be sensitive to changes in water composition that are likely to affect water quality. Future work will include the automated quantification of the changes occurring in the pattern of system states. Furthermore, the visualization and communication of the results to the end-user are an essential part of any early warning system.

By adopting an online data-driven monitoring system, further adaptive management strategies can be developed and implemented. This field-based element is a key part of adaptive management strategy for drinking water suppliers. With further technological developments, such as sensor data availability, the reliability of a measurement can be tracked, i.e. if sensor maintenance is overdue, the reliability of a measurement may be impaired. In combination with the measured values, such additional information can provide the decision maker with sufficient evidence to efficiently manage the process. It may also help to deal with outlier data and identify potential causes for new system states. However, a large part of this transparency in system analysis relies on the availability of the information, e.g. by installing OPC UA (standardized software interfaces with unified architecture) and ensuring a (near) constant availability of the sensors.

Adaptive management strategies include groundwater extraction regulated on a well-to-well basis depending on time-variant pressures, as well as sampling based on event-driven dynamics of a spring. Solutions that combine the characteristics of the process under scrutiny with online measurements, e.g. by pattern analysis, can yield a greater efficiency and prove more robust under change. In this sense, smart systems encompass reactive approaches to managing water resources, whereby they identify changes in the system early on and thus provide ample time to select and implement suitable operational actions.

References

[1] R. M. Page, M. D. Besmer, J. Epting, J. A. Sigrist, F. Hammes, P. Huggenberger, "Online analysis: Deeper insights into water quality dynamics in spring water", Science of the Total Environment, 599-600: 227-236, 2017

[2] C. Butscher and P. Huggenberger, "Modeling the Temporal Variability of Karst Groundwater Vulnerability, with Implications for Climate Change" *Environmental Science & Technology*, vol. 43, p.1665-1669, 2009

[3] A. Affolter, P. Huggenberger, S. Scheidler, J. Epting, "Adaptives Grundwassermanagement in urbanen Gebieten", Grundwasser, vol.15: 147-161, 2010

[4] M. Freiburghaus, "Statistische Übersicht über die Wasserversorgung in der Schweiz 2010", Aqua&Gas, vol. 3|2012: 54:59, 2012

[5] T. Kohohen, "Self-organizing maps of massive databases", Eng. Intell. Syste. Electrical Eng. Commun, vol. 9 (4): 179-185, 2001

[6] J. W. Sammon, "A nonlinear mapping for data structure analysis", IEEE Trans. Comput. C, vol. 18(5): 401-409, 1969

[7] R. M. Page, P. Huggenberger, G. Lischeid, "Multivariate Analysis of Groundwater-Quality Time-Series Using Self-Organizing Maps and Sammon's Mapping", *Water Resources Management*, vol. 29:3957-3970, 2015