Dalla Libera Nico (Orcid ID: 0000-0001-8994-6860)
Pedretti Daniele (Orcid ID: 0000-0001-7677-3948)
Tateo Fabio (Orcid ID: 0000-0002-2335-0199)
Piccinini Leonardo (Orcid ID: 0000-0001-5462-9961)
Fabbri Paolo (Orcid ID: 0000-0003-3821-471X)

Conceptual model of arsenic mobility in the shallow alluvial aquifers near Venice (Italy) elucidated through machine learning and geochemical modeling

Nico Dalla Libera[1], Daniele Pedretti[2], Fabio Tateo[3], Leonardo Mason[4], Leonardo Piccinini[1], Paolo Fabbri[1*]

[1] Department of Geosciences, Università degli Studi di Padova, 35131, Padova, Italy.
[2] Dipartimento di Scienze della Terra "A. Desio", Università degli Studi di Milano (UNIMI), 20133 Milan, Italy.
[3] Institute of Geosciences and Georesources – CNR – IGG, at Department of Geosciences, Università degli Studi di Padova, 35131, Padova, Italy.
[4] Veneto Region Environmental Protection Agency (ARPAV), 30174, Venice, Italy.
* Corresponding author: Paolo Fabbri (paolo.fabbri@unipd.it)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019WR026234

Abstract (num words: 244/300)

This work proposed a novel method to elucidate the controls of As mobility in complex aquifers based on an unsupervised machine learning algorithm, Self-Organizing Map (SOM) and process-based geochemical modeling. The approach is tested in the shallow aquifers of the Venetian Alluvial Plain (VAP) near Venice, Italy, where As concentrations seasonally and locally exceed recommended drinking water limits. SOM was fed using information from two geochemical surveys on eight VAP boreholes, and continuous reading of hourly groundwater head levels and weekly geochemical analyses from three VAP boreholes between mid-October 2017 and end of January 2018. The SOM analysis is consistent with redox-controlled dissolution-precipitation hydrous ferric oxides (HFOs) as a key control of As mobility in the aquifer. Dissolved As is positively correlated to Fe and NH_4^+ and negatively to the oxidizing-reducing potential (ORP). Negative correlation between As and groundwater head levels suggests a redox control by rainfall-driven recharge, which adds oxidants to the aquifer while progressively attenuating As. This mechanism is tested using process-based geochemical modeling, which simulates different transport modalities of oxidants entering the aquifer. Starting from reducing aquifer conditions, the model reproduces correctly the observed ORP and the trends in As and Fe, when the function describing the occurrence of oxidizing events scales according to the temporal occurrence of rainfall events. Heterogeneity can strongly control the local-scale effectiveness of recharge as a natural As attenuating factor, requiring a different model analysis to be properly assessed to be developed in a follow-up study.

1 Introduction

Geogenic (i.e. naturally occurring) arsenic (As) contamination is a well-known issue that affects unconsolidated and fractured aquifers worldwide (e.g. Desbarats *et al.*, 2017; Choudhury *et al.*, 2018; Erickson *et al.*, 2018; Huang *et al.*, 2018; Jakobsen *et al.*, 2018; Pedretti *et al.*, 2019), including Italian aquifers (e.g. Aiuppa *et al.*, 2003; Ungaro *et al.*, 2008; Carraro *et al.*, 2013; Molinari *et al.*, 2013; Rotiroti *et al.*, 2014). Identifying a valid site-specific conceptual model able to describe As mobility in groundwater is fundamental for many purposes related to As risk management. This includes the setup of process-based geochemical models (e.g. Appelo and Postma, 2005; Postma *et al.*, 2007; Rotiroti *et al.*, 2014; Michael and Khan, 2016; Rathi *et al.*, 2017; Jakobsen *et al.*, 2018; Gao *et al.*, 2020) predicting the spatial and temporal scales of occurrence of As concentrations in groundwater (e.g. predicting where As concentrations exceed the 10 µg/L recommended by World Health Organization, WHO).

While modern multicomponent reactive transport codes can resolve multiple nonlinear processes concurring in the control of As mobility (Steefel *et al.*, 2005), incorporating spatial heterogeneity of entry parameters in such models remains computationally very demanding. Simple and effective conceptual models of dominant geochemical processes are preferred for applications such as As risk assessment in heterogeneous alluvial systems (Michael and Khan, 2016; Jakobsen *et al.*, 2018). However, the nonlinear interplay of multiple physical and geochemical factors controlling As mobility in heterogeneous aquifers (e.g. Smedley and Kinniburgh, 2013) makes the decision of the selection of effective conceptual models uncertain.

In the recent years, machine learning (ML) has been increasingly used for exploratory data analysis (EDA) of environmental variables (e.g. Solomatine and Shrestha, 2009; Ejarque-Gonzalez and Butturini, 2014; Sahoo *et al.*, 2017). Among the several existing ML algorithms, Self-Organizing Map (SOM), also known as Kohonen map or network (Kohonen, 2013) is an efficient and powerful technique for EDA. SOM is based on unsupervised- or supervised-learning artificial neural networks (ANN), which can be used for multiple purposes, from clustering analysis to correlation hunting or novelty detection. SOM is popular for industrial, finance or linguistics applications (e.g. Kohonen, 1990, 2013; Vesanto and Ahola, 1999; Vesanto *et al.*, 1999; Himberg *et al.*,

2001; Kalteh *et al.*, 2008). Some studies evaluated its superiority compared to established methods for EDA, such as the widely-adopted principal component analysis (PCA), particularly when applied to large datasets (e.g. Reusch *et al.*, 2005; Astel *et al.*, 2007; Iseri *et al.*, 2009; Karafistan and Gemikonakli, 2020). For water related studies, SOM has been successfully used for water quality classification (Li *et al.*, 2018) or for interpreting biogeochemical processes by correlating multiple variables in complex environments (Melo *et al.*, 2019).

While SOM could be potentially suited to assess nonlinearly coupled geochemical processes controlling As mobility in heterogeneous aquifers, no documented applications of this kind have been presented so far, to our best knowledge. In general, applications of ML for the assessment of As-related geochemical mechanisms are still limited (e.g. Manca and Cervone, 2013; Park *et al.*, 2016; Singh *et al.*, 2018; Smith *et al.*, 2018). Smith *et al.* (2018) presented a ML analysis based on random forests to assess the nonlinear interdependency among subsidence, As concentrations and other variables. They concluded that As mobility was linked to clay strata after finding a positive relationship between aquifer clay content and As concentrations.

The overarching objective of this study was to showcase the combined use of SOM and geochemical modeling as a novel and efficient approach to support elucidating physical and geochemical processes controlling As mobility in aquifers. Specifically, we used SOM to support the development of a simple and effective conceptual geochemical model that could be used for process-based geochemical modeling able to quantify As dynamics in space and time and the associated risks.

We demonstrated the proposed approach to elucidate the main geochemical processes controlling As mobility in the Venetian Alluvial Plain (VAP) near Venice, Italy. The VAP shows geogenic As contamination affecting the local shallow aquifers (Carraro *et al.*, 2013, 2015), leading to complex spatio-temporal variability of concentrations that locally and temporally exceed 10 µg/L (Dalla Libera *et al.*, 2016, 2017, 2018). To date, no documented work has investigated with sufficient detail physical and geochemical processes linked to As mobilization in the shallow aquifers of the VAP, making this a novel contribution for the understanding of As risk in the studied area. In particular, while it is generally understood and accepted that As mobility in the VAP depends on the geochemical interaction between freshwater and

As-rich alluvial materials, the dominant time-dependent mechanisms explaining local excess of As have not been isolated so far. This issue urges clarification of the processes controlling As mobility, as shallow aquifers are used for agricultural activities (e.g., crop irrigation), yard irrigation and potentially direct human consumption unknown to authorities.

The adopted methodology is defined as follows. We identified a simple initial conceptual model that could generally explain As mobility in the VAP. We chose it among previously-presented conceptual models describing As mobility in other sites with similar geological and hydrogeochemical setups (e.g. Smedley and Kinniburgh (2013). Then, site-specific field data were collected to feed SOM analysis and the geochemical model, in order to refine and adapt the conceptual model to the VAP conditions. Exploratory data analysis (EDA) using SOM in *correlation hunting* mode (e.g. Vesanto and Ahola, 1999; Himberg *et al.*, 2001; Barreto and Pérez-Uribe, 2007) was performed to highlight correlated pairs of variables that could be connected to geochemical processes controlling As mobility in the studied area. Ultimately, we adopted the well-known geochemical code PHREEQC (Parkhurst and Appelo, 2013) to evaluate the physical and geochemical validity of key processes of the proposed conceptual model, in particularly to test redox controlling mechanisms related to ground surface – aquifer interactions.

2 Site description and initial conceptual model

2.1 Geological and hydrogeological background

The VAP (Figure 1A) is located in the proximity of the Venice lagoon (Northern Italy). The VAP is characterized by continental climate, with limited rainfall events during the winter and summer seasons, and higher rainfall amount in autumn and spring. Total annual rainfall ranges between 600 mm and 1110 mm. Climatic data used in this work, including daily rainfall data (presented later) were collected from a weather station managed by the Veneto Region Environmental Protection Agency (ARPAV) and located in the proximity of the site. A summary of local climatic information is provided in Table 1. Figure 1B shows the location of the Western Agricultural Areas (WAA), a subset (~3.8 km^2) of the VAP which served as experimental site for this work. The WAA includes all the important elements representative of a typical agricultural area

of the VAP (Ungaro *et al.*, 2008; Carraro *et al.*, 2015; Beretta and Terrenghi, 2017), such as similar land use, geological, hydrogeological and geographical setting. Moreover, the WAA is easily accessible for experimental activities, making it suitable to infer key mechanisms controlling As mobility in the shallow VAP aquifers.

Geologically, the VAP result from the morphological action of the Brenta River during the last 30000 years (Pleistocene – Holocene; Figure 1C). The alluvial plain is characterized by a marked heterogeneity in grain size distributions due to a past environment shaped by the braided river (e.g. Bondesan *et al.*, 2008). In the upper part of the plain, gravelly and sandy alluvial deposits are present, whereas in the distal region, silty-clayey deposits prevail in the alluvial system, according to the typical structure of the Brenta River's mega fan (e.g. Fontana *et al.*, 2008). Specifically, in the distal part of the plain, the subsurface shows sequences of fine-grained sediments (silt and clay) and coarser materials (fine and silty sand). Peat deposits, with a thickness up to several decimeters and a lateral extension up to several hundred meters, are also present in the floodplain environments.

The resulting hydrogeological configuration of the VAP is a sequence of hydrofacies typical of a distal plain, with overlapping sequences of silty-clayed aquitards and aquicludes and fine-grained sandy aquifers. Overall, this sedimentological setting results in a multilayered aquifer system with different degrees of confinement (Vorlicek *et al.*, 2004; Fabbri *et al.*, 2011, 2016; Fabbri and Piccinini, 2013; Piccinini *et al.*, 2016, 2017). Such multilayer systems formed by sequences of low- and high-permeable layers (i.e. aquitards and aquifers) are commonly found in other alluvial aquifers with elevated As concentrations, such as the lower Mekong basin (e.g. Richards *et al.*, 2019) or Red River Floodplain, Vietnam (e.g. Postma *et al.*, 2007).

A stratigraphic sketch of the VAP, highlighting the position of the WAA, is shown in Figure 1C. The shallowest aquifers of the WAA are composed of fine and silty sands embedded within a clayish-silty matrix. Slug tests performed in the WAA (not reported) suggested that, on average, the study area shows hydraulic conductivities close to 5-10 m/d. The regional-scale direction of groundwater is oriented NW-SE, with mean hydraulic gradient ranging from 0.005 to <0.001 while approaching the Venice Lagoon. Local-scale stratigraphic well logs obtained from the drilling of boreholes (used as groundwater piezometers) in the WAA suggest that the aquifer comprises, on average,

the first ten meters below the surface. The aquifer is topped by a low-permeable fine-grained silty and clayey layer. This stratigraphic configuration is observed throughout the entire WAA, leading local administrations to traditionally consider the shallow aquifer as hydraulically confined (i.e. continuously found under pressurized conditions) in all points of the WAA.

2.2 Initial conceptual model

We hypothesized redox-controlled dissolution-precipitation of hydro-ferric oxides (HFOs) as the main mechanism controlling As mobility in the VAP. According to this model, in iron-rich environments As tends to adsorb onto HFOs, which are precipitated during oxidizing conditions and to desorb from HFOs when these oxides dissolve during reducing conditions (Appelo and Postma, 2005). Previous works on the VAP had already postulated this mechanism (Carraro *et al.*, 2013, 2015; Dalla Libera *et al.*, 2017, 2018), which is commonly adopted when modeling As mobility in several other young aquifers of the world (e.g. Smedley and Kinniburgh, 2013), such as Bengal Basin (Kinniburgh and Smedley, 2001), the Mekong Valley and Red River aquifers in Vietnam (Berg *et al.*, 2001; Postma *et al.*, 2007) or in Cambodia (Richards *et al.*, 2019). Redox-controlled As mobility has been also reported in other Italian sites (e.g. Aiuppa *et al.*, 2003; Ungaro *et al.*, 2008; Molinari *et al.*, 2013; Rotiroti *et al.*, 2014).

Maps of arsenic concentrations reported by Dalla Libera *et al.* (2017, 2018) depicted “patchy” distributions of As concentration, suggesting that redox-controlled As variations may be linked to a heterogeneous distribution of aquifer properties controlling the local magnitude of redox variations in the aquifer. Previous studies suggested that reductive conditions in the VAP can be generally attributed to degradation of organic matter (e.g. peat-rich layers), which is sparsely found in the local aquifers (Carraro *et al.*, 2015). Redox variations could be also controlled by surface-driven processes. For instance, Ramesh Kumar and Riyazuddin (2012) reported that rainfall-driven recharge events altered the redox status of reduced shallow aquifers in an Indian site, modifying the relative ratios of As(III) and As(IV) as well as other redox couples. Other studies also identified the control of hydrological events on As mobility in young sedimentary aquifers in South-East Asia (e.g. Chowdhury *et al.*, 1999; Harvey *et al.*, 2006; Postma *et al.*, 2007; Michael and Khan, 2016; Richards *et al.*, 2019). In the VAP, the connection between the surface and the

aquifer can be controlled by “confining capacity” of the fine-grained layer topping the aquifer. If a sufficiently thick fine-grained layer is uniformly settled all over the studied area, poor connection between the aquifer and the surface can be expected. In turn, this may limit short-scale variations of the redox status caused by ingress of oxidants from the surface are inhibited. On the contrary, if the layer is locally more permeable, less thick or even discontinuous (stratigraphic windows in low-permeable aquitards are typically found in Northern Italy shallow aquifers, e.g. Pedretti *et al.*, 2013), a connection between the surface and the shallow aquifer can exist and be an important factor for the control of As mobility.

Considering the above-mentioned processes, the resulting initial conceptual incorporating both redox-controlled precipitation/dissolution of HFOs and the impact of surface-driven redox-altering processes is schematically illustrated in Figure 2. The aquifer is assumed to lie below an initial *vadose* zone, in which oxidizing conditions may fluctuate in space and time. The time dependency can be attributed to surface-related processes, and in particular to rainfall events. As the aquifer is recharged during rainfall events, oxidizing species may enter the aquifer, increasing the ORP and therefore progressively precipitating HFOs. Thus, rainfall-driven recharge can be a direct factor controlling the natural attenuation of As in the aquifer. The space dependency can be attributed to the local connection between the surface and the aquifer. If the aquifer is locally well isolated from the surface, recharge of oxidizing species is less efficient than in case of better connection between the surface and the aquifer.

In the VAP, existing boreholes are generally screened along the entire vertical section of the aquifer. The geochemical composition of the borehole waters results from a (flux-weighted) mixture of waters entering the borehole along the different depths of the screened interval (e.g. Pedretti and Fiori, 2013; Pedretti *et al.*, 2016). As the oxidation front progresses downwards during recharge occurs, more oxidized groundwater enters the boreholes and mixes with the resident water in reduced conditions. The resulting geochemical composition of the groundwater pumped out from a well screened along the entire aquifer vertical section during recharge events should therefore be a progressive attenuation of As on HFO surfaces, which increasingly form in the aquifer while the oxidation front progresses to deeper depths.

Verifying the validity of the postulated conceptual model presented in Figure 2 constitutes the main goal of the remaining steps of the analysis.

3 Materials and Methods

3.1 Field data collection

A site-specific dataset was created collecting information from routine monitoring activities performed by ARPAV and study-specific activities. The dataset is published in an online data repository (see Acknowledgments), which contains supplementary information (SI) supporting the analyses presented below.

Information obtained from ARPAV includes the geochemical sampling of eight HDPE boreholes (piezometers) in the WAA (Figure 1B). The piezometers are screened within the shallowest aquifer, in which As concentrations exceed WHO limits and which constitute the main target of this study. The piezometers are sampled twice a year by ARPAV as part of a larger quality monitoring network for the environmental purposes from the European Directive 2000/60/EC. Measurements of multiple parameters relevant to assess redox-sensitive As mobility, such as NH_4^+, SO_4^{2+}, Fe, Mn, pH, ORP, electrical conductivity (EC), groundwater temperature (T_{gw}) and dissolved oxygen concentration (DO), were performed in situ and at ISO-certified ARPAV laboratories. Since the boreholes are screened along the entire vertical section corresponding to the aquifer levels, the information collected through groundwater sampling is interpreted as “depth-integrated”. Unfortunately, there was no information from multilevel samplers available for this study.

In-situ parameters such as pH, ORP, EC, DO and T_{gw} were evaluated through a multi-parametric probe YSI Professional Plus, calibrated weekly to avoid drifting and other measurements errors. The ORP was estimated with a platinum electrode and successive automatically corrected to Eh with reference to the standard hydrogen electrode. The Eh values were later converted in pE for geochemical calculations. Water sampling and analyses followed the standard procedures suggested by the Italian Environmental Protection and Technical Services Agency (APAT, 2003, 2006) and other protocols (UNI-EN-ISO-17294-2:2016). Although ARPAV has been managing these piezometers for about 10 years, the data before the beginning of this

study were largely incomplete and deemed not suitable to provide supportive evidences to identify As-related geochemical processes (the main goal of this work).

Information from ARPAV was integrated through experimental activities on the WAA. Specifically, these activities aimed at obtaining a detailed assessment of physical and chemical variables that could affect its mobility at shorter scales than the average frequency of ARPAV monitoring activities. A key aspect of the postulated conceptual model is that rainfall-driven recharge can alter the redox status of the aquifer. As recharge is generally occurring in the winter season in the studied area (e.g. Beretta and Terrenghi, 2017), an experimental survey was started on 6th December 2017 and concluded on 26th March 2018, targeting three of the eight piezometers managed by ARPAV (PZP036, PZP040, PZP041) (Figure 1B). The 115 experimental days covered the transition from a dry season to a wet season, during which the local aquifer was expected to be progressively recharged by rainfall events, hydrogeochemical analyses were performed on a weekly basis in these three piezometers. The methodology and analytical procedures were identical to those adopted by ARPAV for their routine monitoring. In these same piezometers, groundwater head levels were continuously collected with an hourly frequency using HOBO data loggers.

The specific selection of the three piezometers PZP036, PZP040, PZP041 was based on the zonation of aqueous As concentrations in the WAA described by Dalla Libera *et al.*, (2016) and the spatially-variable thickness of the fine-grained deposits topping the aquifer. Piezometer PZP036 is located in the central part of the site, PZP040 in its northern part and PZP041 in its southern part. Close to these piezometers, three new shallow boreholes (Figure 1B) were drilled and stratigraphic data were obtained from the extracted soil cores. These cores allowed analyzing the stratigraphic sequence of the first 6m of soil below the ground, as well as the mineralogical composition of the soils. Bulk concentrations of solid phases were analyzed through X-ray Powder Diffraction (XRPD), X-ray Fluorescence (XRF) and Inductively Coupled Plasma Optical Emission Spectroscopy (ICP-OES) (technical details are provided in the SI). We did not perform specific mineralogical analyses to quantify the extent of amorphous and crystalline iron and manganese phases, which could be relevant to identify solid-phase sorption sites controlling As mobility. Such analyses are left open for future developments.

3.2 Self-Organizing Map (SOM)

The essential idea of SOM is to project objects onto a two-dimensional (2D) plane, such that similar objects are close together and dissimilar objects are far away from each other. Compared to other EDA approaches involving dimensionality reduction, such as PCA, SOM tends to conserve the original topology of the data, which is important to perform correlation among SOM maps, as described below. Objects are mapped to specific positions in 2D planes called *units.* Each unit is associated with an object called a *codebook vector*. During training, SOM calculates the distance of the new data points to these codebook vectors and assign each object to the unit with the most similar codebook vector (the *best-matching* or *winning unit*). Detailed information on SOM can be found in Himberg *et al.*, 2001; Kohonen, 2001, 2013; Kalteh *et al.*, 2008. In this work, SOM analyses were performed using the R code `kohonen 3.0` (Wehrens and Buydens, 2007; Wehrens and Kruisselbrink, 2018). The parameters adopted to run the simulations (including the selection of initialization methods and neighboring kernels) are based on the default values of the adopted code.

The input of our SOM analysis was the $N \times m_N$ matrix representing the time series of aqueous concentrations from the *N* geochemical variables collected in the analysis (i.e. the values reported in Table A, SI). Here, m_N is the number of temporal measurements collected for each variable. Groundwater head changes were also considered as SOM input variables, as they can show possible correlations with the geochemical variables (a key hypothesis of this study). Since our goal is to identify temporally dependent correlations among As and other dynamic variables, we did not use SOM for correlations with static solid phases (Table B, SI). Data preprocessing and preparation, required in any SOM analysis (as well as in any EDA), is described in the SI. This includes technical details regarding data selection and segmentation, including error filtering and data censoring. The SI also reports the adopted methodology and results of the model training, the estimation of fitting quality (i.e. quantization errors) and ML rate.

The outputs of the SOM analysis can be of different types. The most known is possibly the unified distance matrix (or *U-matrix*), which is used SOM for clustering analysis. The U-matrix is a single toroidal color map, where the results from the analysis of the multidimensional $N \times m_N$ matrix is plotted in a reduced 2D space. Hexagons are

typically used to highlight the boundaries of a gridded maps, on which the nodes are located, even though other cell geometries can be equally selected. The U-matrix shows represent the distances among neighboring input vectors, and they are used to identify the existence of clusters in the entire dataset. However, cluster analysis does not provide information on the pairwise correlation between As and other variables, which is the information of interest to test the validity of the proposed conceptual model.

To obtain pairwise correlation between As and the other variables, we used SOM for *correlation hunting* analysis (e.g. Vesanto and Ahola, 1999; Himberg *et al.*, 2001). Using this approach, the outputs from SOM analysis are the *component planes*, which can be seen as the sliced version of the U-matrix in the different individual variables (components) forming the map (e.g. Himberg et al. 2001). In our case, each component plane reports the distribution of the normalized values (e.g. concentrations, pH, ORP) which is mapped preserving the topology of the input vectors in the 2D toroidal maps generated by SOM. The normalization approach is described in the SI.

The comparison of component planes can be used to directly compute the pairwise correlations, according to the following rationale:

- if two maps show similar color patterns, positive correlation between the pairs of corresponding variables is likely;
- if the maps show opposite color patterns, negative correlation is likely;
- if the maps show dissimilar color patterns, correlation is unlikely.

To quantify the correlation between As and the variables, the Spearman's correlation coefficient ρ was calculated between the As SOM map and the other (N-1) maps through the R package `stats`. The Spearman's correlation coefficient is a nonparametric measure of rank correlation, which determines how well a monotonic function describes the relationship between two variables. The Spearman's correlation ($p < 0.05$) results in $\rho \rightarrow 0$ for uncorrelated variables and $\rho \rightarrow \pm 1$ for positively (+) or negatively (-) correlated variables.

3.3 Geochemical modeling

3.3.1 Preliminary considerations

PHREEQC has been widely adopted to simulate redox-controlled mechanisms associated with As mobility in alluvial aquifers (e.g. Aiuppa *et al.*, 2003; Sracek *et al.*, 2004; Appelo and Postma, 2005; Postma *et al.*, 2007; Wallis *et al.*, 2011). Since rainfall-driven recharge was postulated as a relevant factor controlling the fate of HFO-controlled As, we used PHREEQC to evaluate the effect of different transport modalities controlling the rate of oxidizing species entering the aquifer during recharge. Specifically, we tested if the ingress of oxidizing species, and consequently As mobilization, could be limited by the frequency of rainfall events controlling aquifer recharge and by the transport modalities of oxidizing species entering the aquifer bypassing the fine-grained layer topping the aquifer.

Accordingly, a model was developed with the main purpose of reproducing the time series of the three weekly monitored piezometers (PZP036, PZP040 and PZP041) by mimicking three possible transport mechanisms controlling the ingress of oxidizing compounds in the aquifer:

a) Transport of oxidizing species (e.g. $O_{2(aq)}$) dissolved in the infiltrating water during rainfall events.

b) Transport of gaseous oxidizing species (e.g. $O_{2(g)}$) in contact with the groundwater.

c) A mixed situation between Scenario (a) and (b).

Scenario (a) assumed a transient ingress of oxidizing species described by functions that scales the flux of oxidants to the aquifer according to the frequency of recharge events and the solubility of oxidation species in water. The *worst* case is when infiltrating water reaching the aquifer penetrates the vadose zone so quickly that it does not equilibrate with the soil mineralogical compositions, resulting in infiltrating water reflecting the composition of rainfall. Scenario (b) implied a continuous, steady-state source of oxidizing species. The *best* case occurs when the aquifer is perfectly connected to the atmosphere (an open or *untapped* system). In this case, the

concentration of oxidizing species is not limited by the effects of gas solubility in water and gaseous molecules are also directly added to the reactive system.

The mixed scenario (c) represented an intermediate scenario between (a) and (b) scenarios. Here, both gaseous (e.g. $O_{2(g)}$) and dissolved aqueous species (e.g. $O_{2(aq)}$) can both equilibrate with the soils and bypass the vadose zone, for instance being "pushed" downward by the propagation of the wetting front and transported to the aquifer, as proposed by Healy and Cook (2002). While the amount of total oxidizing species entering the aquifer can be still rate-limited by the occurrence of rainfall events, rainfall events could trigger a mixture of oxidizing conditions in the system (with larger fluxes of oxidizing species), creating a series of pulses that determine a progressive transition from reducing to oxidizing aquifer conditions, precipitation of HFO and in turn attenuation of As.

3.3.2 Model setup and input data

The aquifer was conceptualized as a zero-dimensional batch reactor. Starting from reduced aquifer conditions (i.e. low ORP values), the model replicated the increase in ORP and the associated attenuation of As by sorption on HFOs following the addition of oxidizing species to the reactive system. Speciation was performed using the thermodynamic database WATEQ4F (Ball and Nordstrom, 1991). Far-from-equilibrium (kinetic) reactions were described through Monod laws (Monod, 1949). Table C (SI) summarizes key equations adopted for this analysis, with corresponding equilibrium constants (k_{eq}) at 25°C.

In the first phase of the model setup, the goal was to find appropriate initial conditions with low ORP. To this end, we modelled a water geochemical composition (Table 2) similar to the one found at the first day of observations (6th December 2017, the model "time zero"). This water was equilibrated with initial concentrations of mineralogical phase estimated from the experimental characterization of the solid matrix (Table B, SI) as well as expert knowledge on the VAP The simulated minerals in the aquifer were calcite, organic matter (OM) and ferrihydrite. Arsenic-bearing sulfides were initially included in the simulations, yet the results were insensitive to the presence of these minerals, in agreement with the analysis by Carraro *et al.* (2015) and were therefore removed (this aspect is further discussed below). Calcite and ferrihydrite

were assumed to be in equilibrium in the system, while OM was simulated as a rate-limited degradation process described by a Monod kinetic law, defined as

$$\frac{dS}{dt} = -k_{max}\frac{S}{k' + S} \quad (1)$$

where *S* is the concentration of the organic substrate utilized by the microbes for the degradation process, *t* is the time, k_{max} is the maximal rate and *k'* is the half-saturation constant. Specifically, this reaction was applied as an additive Monod kinetics model of form

$$\frac{dS_{SOC}}{dt} = -c\, r_{TEA}\, S_{Soc} \quad (2)$$

which considers the concentration of the main terminal electron acceptors (TEAs) in the system (Appelo and Postma, 2005). In Eq.(2), S_{SOC} is the concentration of soil organic carbon (*SOC*), *c* is a first-order reaction constant and r_{TEA} represents the electron exchange rate between the oxidants and the reductant (i.e. *SOC*). We considered $O_2, Fe(OH)_3, NO_3^-$ and SO_4^{2-} as the main oxidants, such that

$$r_{TEA} = \left(\frac{[O_2]}{k'_{O_2} + [O_2]} + \frac{[Fe(OH)_3]}{k'_{Fe(OH)_3} + [Fe(OH)_3]} + \frac{[NO_3^-]}{k'_{NO_3^-} + [NO_3^-]} + \frac{[SO_4^{2-}]}{k'_{SO_4^{2-}} + [SO_4^{2-}]}\right) \quad (3)$$

Preliminary analyses (not reported) were performed to find suitable OM kinetic rates that enabled ORP reduction according to the observed results during the monitoring time (115 days).

Once suitable initial conditions were obtained, the system was perturbed through a 115-step function modulating the amount of oxidants entering the aquifer with time. This effective approach was adopted since transport through the unsaturated zone cannot be mechanistically simulated by PHREEQC. The function took three different general forms, according to each of the postulated transport modalities of oxidizing species.

- For Scenario (a), the function added oxidizing species exclusively during the recorded rainfall events. The composition of the infiltrating water reflected the composition of the rainfall water (Table 2). No contribution from gaseous

phases was simulated. Dissolved oxygen ($O_{2(aq)}$) was assumed to be in equilibrium with atmospheric gas composition at the ground level. A 94:6 ratio was assumed as the effective mix between native and infiltrating water in each step, respectively, which is similar to the estimate 6% of groundwater recharge found by Beretta and Terrenghi (2017) for the VAP.

- For Scenario (b), the function added a steady ingress of gaseous oxidizing species. $O_{2(g)}$ was simulated, mimicking the increase of groundwater heads that tends to better connect the aquifer with the atmosphere. In this scenario, ingress of oxidants as aqueous species from recharge events was not considered. Oxygen gas molecules at a partial pressure of 0.21 atm entered the aquifer.
- For Scenario (c), we considered both mechanisms accounted for in Scenarios (a) and (b). Rainfall-controlled transient ingress of oxidizing species in dissolved form was coupled with an extra contribution of $O_{2(g)}$. The total oxidants flux is rate-limited by a scaling function, which was manually calibrated to reproduce the observed ORP. The function actively added oxidants solely during recorded rainfall events, being zero otherwise.

The main redox-sensitive processes attenuating As and hypothesized in the study area were electrostatic processes occurring at the molecular scale and responsible for the sorption/desorption of As (e.g. Jessen *et al.*, 2012; Biswas *et al.*, 2014; Rawson *et al.*, 2016). We adopted a surface complex model (SCM) to provide a mechanistic quantification of As adsorption behavior as a function of the dissolved As concentration and speciation and the solution pH. The SCM considers the influence of competing ions such as phosphate or bicarbonate as well as the density of adsorption sites on the host minerals, mainly the HFOs (e.g. Rathi *et al.*, 2017). Sorption and desorption of As to reactive surfaces was therefore simulated through a set of surface complexation reactions for As(III) and As(V) based on the surface sorption sites defined for HFOs, the primary expected sorption sites for As. The WATEQ4F database provided the reference stability constants for thermodynamic equilibrium between HFOs and different compounds based on the widely-adopted Dzombak and Morel (1990) database (Table C, SI). We used the HFO properties reported in Appelo and Postma (2005, Table 7.5).

To calculate the goodness-of-fit of the different models, we adopted the RMSE-observations standard deviation ratio (RSR) (e.g. Moriasi *et al.*, 2007), which standardizes the root-mean-square error (RMSE) using the observations standard deviation (σ_{OBS}), such that for a general variable *Y*,

$$RSR = \frac{RMSE}{\sigma_{OBS}} = \frac{\sqrt{\sum_{i=1}^{n}(Y_i^{obs} - Y_i^{sim})^2}}{\sqrt{\sum_{i=1}^{n}(Y_i^{obs} - \bar{Y})^2}} \quad (4)$$

where $\bar{Y}$ is the mean. Moriasi *et al.* (2007) suggested that a "very good" performance rating is obtained when *RSR*<0.5, "good" when 0.5<*RSR*<0.6, "satisfactory" when 0.6<*RSR*<0.7 and "unsatisfactory" when *RSR*>0.7.

4 Results and analysis

4.1 Stratigraphic analysis

The analysis of the soil samples obtained from the manually-drilled boreholes confirmed the presence of a fine-grained layer at the top of the stratigraphic sequence in the three drilled points. However, the thickness of this fine-grained layer was observed to vary significantly among the three points. The thickness was about 1 m at the manually-drilled borehole located some 1.5 m away from one of the sites piezometers PZP036. In similar proximity of another piezometer, PZP040, the thickness of the fine-grained layer exceeded 3 m and in the proximity of PZP041 it exceeded 4 m.

Considering the strong variability in the fine-grained layer, we advocate that the assumption of "perfectly confined conditions" of the shallow aquifer can be disputed. In some points, the aquifer can be sufficiently thin to locally generate a better connection between the surface and the aquifer. Such connection could be boosted by agricultural activities, largely occurring in the VAP. Tillage is known to alter the natural oxygen flux into the soil (e.g. Clay *et al.*, 1992). In all boreholes, we observed a clear effect of agricultural tillage, which has altered the natural compaction of the sediments in the first 0.2 mm potentially increasing the amount of macropores. Although we did not estimate the change in permeability or porosity in tillage-altered soil profiles, it seems reasonable that the

conceptual model should also include, a postulated, the potential effects surface-driven mechanisms controlling the redox variations in the shallow aquifers of the studied area.

4.2 Exploratory data analysis

A summary of the results obtained from the combination of all temporal measurements from the eight piezometers is provided in Table 1. We found that mean As concentrations were close to 10 µg/L, i.e. the WHO recommended drinking water concentration threshold, with maximum As concentration of 45 µg/L. The ORP was found to vary between negative values with minimum -178 mV (corresponding to reduced aquifer conditions) and nearly zero values (corresponding to near-oxidized aquifer conditions). The groundwater pH was found at circumneutral conditions, consistent with the carbonate-rich nature of Northern Italy alluvial aquifers (e.g. Rotiroti *et al.*, 2014). Under reducing and circumneutral conditions, As is highly mobile Smedley and Kinniburgh, 2013). Small variations in pH should not substantially affect As mobility in the studied area compared to the more pronounced variations in redox conditions.

Figure 3 shows the time series of As, ORP groundwater heads (*H*) from the three boreholes with more frequently measurements. At PZP036, we observed a progressive increase in *H*, up to a final difference $\Delta H \approx 0.3$ m compared to the initial values, and an increase of ORP from negative values close to ORP ≈ -110 mV to ORP ≈ 0 m. During the same period, the data revealed a decrease of As (from 2µg/L to <1 µg/L) and Fe (from 1200 µg/L to <200 µg/L). At PZP040, the trends of these variables were different. Head levels still fluctuated with final $\Delta H \approx 0.4$ m, yet ORP remained negative, with an initial drop ORP ≈ -90 mV and a later recovering to ORP ≈ -70 mV. Similarly, As and Fe showed an initial increase in concentrations (opposite to ORP drop), followed by a reduction in concentration (opposite to ORP recovery). PZP041 showed a third different behavior. Again, head level changes remain comparable to those observed in the previous piezometers ($\Delta H \approx 0.4$ m). However, a more marked inverse relationship between ORP and As, with maximum concentrations of As close to 45 µg/L when ORP ≈ -140 mV. The difference in As concentrations and redox variations observed in the three experimental piezometers may be due to the heterogeneous abundance of the local As source (e.g., variability in the abundance of

previously existing As-bearing HFOs prior to dissolution) and to the physical heterogeneity of the soils, which could limit the efficiency of rainfall-driven recharge as a redox controller for the aquifer, according to the postulated conceptual model.

The application of SOM helped quantifying the correlation between As and the multiple variables that could be geochemically linked to its mobility. The results are shown at the top of Figure 4 in the form of black and white SOM component planes. From a visual perspective, the As toroidal map (green frame) showed maximum values in the top part of the map, with a peak in the middle-right part, and minimum value in the top-left corner of the map. This peculiar distribution of values is a random outcome of the SOM analysis, which depends on the random initialization (i.e. the *seed*) of the artificial neural network. We highlight that these patterns does not correspond to a specific spatial organization of As concentrations in real-life space. The same applies to the other maps of the variables. The key point is that the component planes, by pairs, maintain the same relative topological distribution of value, and thus their pairwise correlation, independently from the adopted seed.

To find the correlations among variables, the Spearman's correlation coefficient ρ was computed. The results (bottom of Figure 4) support the hypothesis of redox-controlled dissolution-precipitation of HFOs controlling As mobility. We found direct correlations between As and Fe (ρ = 0.66) and NH_4^+ (ρ = 0.58) and negative correlations with ORP (ρ = -0.42), which is consistent with the enhanced mobility of As when the aquifer tends to reduced conditions - and contrarywise. The inverse correlation between As and *H* (ρ = -0.43) suggests that potentiometric head changes and redox variations may be mainly driven by rainfall-driven recharge in the studied area, as postulated. Recharge typically occurs from October to May and occasionally from June to September in the VAP. Recharge could drive the ingress of oxidants into the aquifer, which would explain the observed progressive increase in ORP within the aquifer as the experimental time elapses (Figure 3). The positive correlation between As and Cl (ρ = 0.54), which is a conservative element that can also be introduced into the system during recharge events (e.g., salts accumulating on the surface and quickly leached into the aquifer during rainfall events), seems also consistent with these hypotheses.

The fact that short-term rainfall event controls the redox variation means that a connection between the surface and the aquifer exists in the VAP, further

corroborating the analysis at the end of Section 4.1. The fine-grained layer topping the aquifer could be considered as an aquitard, allowing connection between the surface and the sandy aquifer. However, it should be recalled that the SOM analysis lumps together the results from all boreholes. Soil heterogeneities may determine a spatial variability in the control of recharge on the change in redox status, thus generate local departures from this generalize behavior. As such, the fine-grained layer may be differently impervious in the different portion of the studied area.

Initially low ORP values were probably due to the degradation of organic matter concentrated within peat horizons and dispersed in other sediments, a process that has been already hypothesized to be associated with dissolved As concentrations in the VAP (Carraro *et al.*, 2013, 2015), in other Italian aquifers (e.g. Rotiroti *et al.*, 2014) and in other sites in the world (e.g. McArthur *et al.*, 2001). The correlation of As with NH_4^+ (*ρ=0.58*) and, more weakly, with TOC (*ρ=0.3*) supports these hypotheses. The degradation of organic components may thus be involved in the control of the redox conditions of the subsurface, which favors the dissolution of HFOs.

Another important result from the SOM analysis was the inverse correlation between As and SO_4^{-2} (*ρ* = -0.71), which suggests a decoupled geochemistry of As and sulphate under variable redox conditions. Sulfide-bearing sediments are sometimes assumed to be a source/sink of As in alluvial aquifers (Saunders *et al.*, 2008). While pyrite and other sulfides were identified by bulk mineralogical analyses on solid sample, As uptake or release by sulfides is not expected to be a predominant in the studied aquifer. Maximum concentrations of SO_4^{-2} during high ORP stages are far below the expected values when pyrite oxidation occurs, suggesting that the oxidation of As-bearing sediment may not be a primary actor explaining the occurrence of As in the aquifer. These results agree, for instance, with the observations reported by Harvey *et al.* (2002). It is noted that much lower ORP values than those characterizing the shallow aquifers were observed at deeper depths of the VAP. Carraro *et al.* (2015) conducted a study in the central VAP, reporting ORP as low as -285 mV at average depths of 31 m below the ground surface. At redox conditions with ORP below -250 mV, peculiar relationships between As and S compounds were observed, and the solubility of As could be modelled considering the possible occurrence of pararealgar, which was detected in the aquifer (Artioli *et al.*, 2019). The conditions determining

redox-controlled As-bearing sulfides solubility were never found in shallow aquifers (Carraro *et al.*, 2013 and this study), further supporting the hypothesis that HFO precipitation and dissolution can be a more plausible mechanism controlling As mobility under recharge-controlled redox variations.

The SOM analysis revealed high correlation between As and other variables not directly linked to the proposed conceptual model. These correlations are still in agreement with the initial hypothesis, even if they are not connected with the mechanism of As release. One example is the high correlation value between As and Si. This can be the result of silicate solubility in aquifers where organic matter occurs (e.g. Bennett *et al.*, 1991), which is the case of the studied area. More specific dissolution processes of silicates can develop when specific types of organic ligands occur in the solution. These effects were studied for instance by Ramos *et al.* (2011), who defined a ligand-promoted over the proton-promoted dissolution process for silicates. Similarly, the inverse trend for As and Ca can be a simple dilution effects of sediment detrital carbonates, as these minerals have very low As. The poor correlation between the As and pH is consistent with the low variation of pH in the systems, corroborating that pH conditions are not expected to affect the mobility of As in the studied area.

4.3 Geochemical modeling

4.3.1 PZP036

At PZP036, we found dissolved As concentrations ranging between 2-3 µg/L during reducing conditions and progressively dropping to 1 µg/L as the system gets more oxidized. We recall that PZP036 is located in the central zone of the WAA (Figure 1), where the topping fine-grained layer of soil sediments is thinner than around the other two experimental boreholes PZP040 and PZP041.

Figure 5 compares the measured ORP and the PHREEQC model output for each of the simulated transport mechanisms (a, b, c) controlling the ingress of oxidants in the aquifer. We observe that Scenario “a” (black dashed line) underestimated the observations. This suggests that the effect of recharge solely due to the infiltrating of rainfall water was not enough to progressively increase the ORP to oxidizing

conditions. This is explained considering the limited solubility of oxygen in water. All oxygen molecules are rapidly consumed by the OM, while the contribution of other oxidizing species such as sulfate and nitrate does not seem to contribute to alter the redox status of the aquifer. An opposite situation is observed for Scenario “b” (blue dashed line), for which the model severely overestimates the observed ORP with time, resulting in an increasing trend of ORP which quickly reaches an asymptotic value close to 800 mV. The result is explained considering that the rate of oxygen molecules entering the system in gaseous form exceeds the consumption rates of oxidants by OM degradation. The model was insensitive to different OM kinetic rates, as similar results were obtained setting near-equilibrium conditions.

Scenario “c” (red line), which reproduced the mixed condition between “a” and “b”, matched well the observations. Using the *RSR* to evaluate the goodness-of-fit (Equation 4), we found a “very good” fitting (*RSR*<0.06) between observed and simulated ORP data as also shown in Figure 6(a). Such fitting is due to the *ad-hoc* tuning of the rate-limiting function, which scales the total amount of $O_{2(g)}$ and dissolved oxidizing species entering the aquifer with time. The shape of the rate-limiting function is shown in Figure 6(b). It shows mild pulses of oxidizing species entering the batch system, which agrees with the temporal occurrence of the main rainfall events recorded in the area. The correlation between rainfall events and changes in ORP supports the hypothesis that recharge events are connected to redox variations, and in turn to As mobility. The total concentrations of oxidizing species entering the aquifer increases from 10^{-5} mol/L to 4×10^{-4} mol/L, a reference value to be compared with the amount of oxidizing species computed for the other two piezometers, described below.

The resulting concentrations for As and Fe simulated by the model and embedding Scenario “c” is shown in Figure 6(c) and Figure 6 (d), respectively. In general, the model correctly reproduces the observed trends, with *RSR*<0.5 for both variables. In the case of PZP036, the model simulated a reduction in As concentrations towards the end of the experimental time, which is consistent with the minimum value of As concentrations obtained from laboratory analysis. For Fe, the decreasing concentrations were also correctly captured, supporting the hypothesis of As attenuation by precipitation of HFOs as ORP increases.

4.3.2 PZP040

The results from PZP040 displayed a different behavior compared to PZP036. In contrast with PZP036, PZP040 showed a generally stable trend in ORP between -40 and -100 mV throughout the experimental monitoring period (Figure 6(a)). Consistently, the resulting fitted function required to match the observed ORP data from PZP040 was steadily close to about 10^{-5} mol/L, which is approximately one order of magnitude lower than the amount of oxidants simulated for PZP036. The *RSR* was below 0.5.

This result is explained considering that PZP040 is located in the northern part of the site, where the fine-grained sediments topping the aquifer is thicker that around PZP036 (Section 4.1). As such, a lower influx of oxidizing species is consistent with the hypothesis that the thicker fine-grained layer at PZP040 inhibits more the connection between ground surface and the aquifer than at PZP036. A more limited connection with the surface for PZP040 is also consistent with the observed delayed response of groundwater head changes to rainfall events, when compared with the more rapid response observed at PZP036 for identical rainfall events (Figure 3). Hence, PZP040 could represent a more confined aquifer condition than around PZP036.

The geochemical change in As (Figure 6(c)) and Fe (Figure 6(d)) at PZP040 agreed with the lack of redox-perturbing recharge events. At the beginning of the experimental activity, As concentrations ranged between 12 µg/L and 20 µg/L, i.e. higher concentrations than those found at PZP036 and largely exceeding the WHO recommended threshold. These concentrations were generally stable throughout the experimental time. This is also true for Fe, which was well correlated to As, and matched well the expected behavior of the system based on the proposed conceptual model. The geochemical model reproduces steady As and Fe concentrations, achieving a “good” fitting for As (*RSR*=0.54) and “very good” fitting for Fe (*RSR*=0.17). Simulated concentrations of Fe at PZP040 were one order of magnitude higher than at the other piezometers, according to the experimental observations.

4.3.3 PZP041

The results from PZP041 provided a different way to corroborate the validity of the proposed conceptual model. At PZP041, As concentrations exceeding 40 µg/L were observed initially. Progressively, As concentrations decreased as ORP increased, according to the postulated processes. Concentrations of observed As, Fe and ORP showed well-defined peaks, suggesting a more direct and efficient ingress of oxidizing species than at the other piezometers. The amount of oxidizing species needed to model these peaks was approximately 10^{-3} mol/L (Figure 6(b)), i.e. an order of magnitude higher than at PZP036. Nonetheless, observing the stratigraphic log from the three boreholes (Section 4.1), PZP041 showed a thicker layer of fine-grained materials topping the aquifer than at PZP036. Thus, the quick fluctuation of oxidizing conditions observed at PZP041 seems to contrast with the conceptual model presented so far.

We explain this apparent anomaly considering that that PZP041 might be strongly influenced by a quick external perturbation that "unnaturally" drove oxidizing species to the aquifer. During a site visit during rainfall events, we observed ponding water the around the PZP041 wellhead. This could likely result in the direct flow of the surface runoff water into the piezometer, which might explain the very fast response of the potentiometric head to rainfall events (Figure 3(c)). This anomalous perturbation invalids the observations at PZP041 for the evaluation of the natural aquifer response around this piezometer. However, the "very good" match between the simulated values and the observations (*RSR* < 0.32 for all variables) further corroborates the hypothesis that the ingress of oxidizing species and the subsequent sequestration of As by HFOs precipitation could explain As mobility in the system.

5 Discussion

The combination of experimental observations, machine learning (SOM) and geochemical modeling (PHREEQC) consistently suggested the validity of the initially hypothesized conceptual model (Figure 2). The role of rainfall-driven recharge events as redox regulator of shallow aquifer is in line with the results from previous studies in other alluvial aquifers (e.g. Ramesh Kumar and Riyazuddin, 2012). This result has

multiple implications both form a scientific perspective and from a decision-making perspective, particularly for As risk management by local authorities.

One of these aspects is how crop irrigation could affect natural recharge. Since VAP aquifers are affected by irrigation during the dry season, irrigation-related processes could in principle trigger a reduction of As by adding oxidants to the soils (as a sort of "artificial recharge" process). However, other coexisting processes may render the role of irrigation less straightforward to disentangle. For instance, in a Bangladesh site, Harvey *et al.* (2006) indicated that irrigation water tended to pond over the low-permeable layers, thereby changing not only the hydrologic budget but also the biogeochemistry of the recharge water. We have no accurate long-term data to calculated seasonal water budgets, which could help at least inferring the volumetric contribution of irrigation-driven recharge in the aquifers. One problem is that the amount of water volumes used for irrigation is strongly farmer-dependent. Specific surveys to obtain farmer irrigation data will be one goal of future studies in the VAP.

Another aspect to be considered is about the use of multicomponent reactive transport modeling to perform As risk assessment at larger scales than those characterizing the experimental site. Such risk assessment cannot be performed using the limited number of spatially-distributed piezometers available for the present analysis. The models presented here are also limited in this sense, being based on zero-dimensional batch-like reactors. A 3D model is deemed better suited for risk analysis in the VAP. Such a model could include, for instance, the hydro-stratigraphic units of VAP subsoil reconstructed using the transition probability based code spMC (Sartore *et al.*, 2016; Fabbri *et al.*, 2020). This code allows evaluating the impact of the random spatial variability of key parameters, including peat distributions or recharge rates at regional scales. The resulting 3D model could then be used to quantify the magnitude of the As release in the VAP and the risk associated to seasonal changes in recharge patterns in all point of the system. The model could directly inherit the simple conceptual geochemical model developed in this work. Codes such as PHT3D (e.g. Prommer *et al.*, 2003) or PHAST (e.g. Parkhurst *et al.*, 2004) use PHREEQC as geochemical engine, and the geochemical system adopted for the batch models developed here could be directly used to build a multidimensional model using those

codes. The model could also help clarifying the spatial variability in confinement status of the studied aquifer.

Ultimately, we recall that a goal of this study was to propose SOM as a novel and efficient machine learning approach to perform EDA, in particular to quantify nonlinear correlation among variables. It was not a goal of this work to demonstrate that this method is superior to other methods for correlation hunting, such as regressive methods, PCA or even other machine-learning algorithm. Such an evaluation requires solid synthetic benchmarks or larger experimental datasets than those adopted in this work. We carried out a PCA (not reported) and obtained qualitatively similar results as those obtained by the SOM analysis when considering the distribution of analyzed data on the biplots of the first two principal components. However, the variance associated to the first two components was less than 50%, suggesting that a more complex analysis involving more components would be likely needed if PCA was used for EDA. Since such more complex analysis lies beyond the scope of the model, it was not analyzed further.

Conclusion

This paper presented a novel combination of exploratory data analysis (EDA) based on machine learning and geochemical modeling to infer the key processes controlling As mobility in complex systems. We applied this approach to the Western Agricultural Areas (WAA), a small area of the Venetian Alluvial Plain, VAP (near Venice, Italy), with the overall goal of identifying key mechanisms determining variations of As concentrations observed at different piezometers. Starting from an initial general conceptual model that accounts for HFOs as key redox-dependent phase controlling As mobility and recharge-controlled redox variations, the analysis of the WAA served as an experimental site to study the geochemical processes in the VAP.

The main conclusions achieved from this work are listed as follows.

- The interpretation of the key controls of As mobility in the WAA benefitted from the combined used of SOM and PHREEQC. While visual assessment of time series provided insight to correlate specific pairs of variables, the use of SOM assisted to corroborate quantitatively the nonlinear relationships among all

variables obtained from a hydrogeological and geochemical assessment of the WAA.

- SOM-based correlations between As and other variables were consistent with the initial general conceptual model. However, SOM-identified correlation between As, ORP and groundwater head changes was a key factor to link rainfall-driven recharge events to redox variations in the aquifer. In the VAP, natural attenuation of As may be thus depend on the local intensity of recharge to the aquifer and its impact to oxidize the system, tending to precipitate HFOs on which As adsorb and become immobilized.
- Process-based PHREEQC modeling was useful to test the conceptual model by simulating three possible transport mechanisms delivering oxidants to the aquifer. It was found that the infiltration of rainfall water was not sufficient to increase the aquifer oxidation potential, while it seemed unlikely that a perfect connection between aquifer and atmosphere exists in the studied domain. An intermediate mechanism between infiltration of pure rainfall water and direct input of $O_{2(g)}$ molecules is instead more plausible. The simulations generated good match with the field observations when the rate-limiting function scaling the amount of oxidants entering the aquifer scales according to the occurrence of rainfall events, which are directly connected to aquifer recharge.
- Aquifer heterogeneities determine a distinct local response of As concentrations to the effects of recharge in the study area. This is possibly occurring also at the scale in the VAP. The spatial variability of the confining or semi-confining fine-grained layer topping the aquifer could be a key limiting factor scaling the amount of oxidants entering the aquifer in the different part of the system.

Acknowledgments

We are thankful to the Editors and the anonymous Reviewers for their useful comments and suggestions to improve the quality of this paper. This study was conducted under a collaboration agreement between Università di Padova and Veneto Region Environmental Protection Agency (ARPAV). The authors thank the ARPAV laboratories for the chemical analysis performed on groundwater and soil samples. DP acknowledges MIUR grant "*Dipartimenti di Eccellenza 2018-2022*". The experimental

data and other supplementary information (SI) of this paper is available in an online Mendeley data repository (https://data.mendeley.com/datasets/v4jm2txpfw/draft?a=d28e0a1e-5413-47f2-bda0-71715e9e74f8), published under Creative Common license CC-BY-4.0

References

Aiuppa A, D'Alessandro W, Federico C, Palumbo B, Valenza M. 2003. The aquatic geochemistry of arsenic in volcanic groundwaters from southern Italy. *Applied Geochemistry*, 18(9): 1283–1296. https://doi.org/10.1016/S0883-2927(03)00051-9.

APAT. 2003. *Analytical methods for chemical water analysis, vol. 1. ("Metodi analitici per le acque - Volume I - Sezione 1000". APAT Rapporti 29). In Italian.* .

APAT. 2006. *Handbook for environmental investigation upon contaminated sites. ("Manuale per le indagini ambientali nei siti contaminati". APAT Manuali e linee guida 43). In Italian.* .

Appelo C a. J, Postma D. 2005. *Geochemistry, groundwater and pollution.* .

Artioli G, Berto D, Dalconi MC, Gò E, Peruzzo L, Tateo F. 2019. Occurrence of pararealgar in As-rich aquifer sediments of the Venetian plain (Italy). *Congresso SIMP-SGI-SOGEI Parma 2019*. Società Geologica Italiana: Parma, Italy, 343–343.

Astel A, Tsakovski S, Barbieri P, Simeonov V. 2007. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Research*, 41(19): 4566–4578. https://doi.org/10.1016/j.watres.2007.06.030.

Ball JW, Nordstrom DK. 1991. *WATEQ4F -- User's manual with revised thermodynamic data base and test cases for calculating speciation of major, trace and redox elements in natural waters*. USGS Numbered Series. .

Barreto MAS, Pérez-Uribe A. 2007. Improving the Correlation Hunting in a Large Quantity of SOM Component Planes. In: de Sá JM, Alexandre LA, Duch W and Mandic D (eds) *Artificial Neural Networks – ICANN 2007*. Springer: Berlin, Heidelberg, 379–388.

Bennett PC, Siegel DI, Hill BM, Glaser PH. 1991. Fate of silicate minerals in a peat bog. *Geology*. Geological Society of America, 19(4): 328–331.

Beretta GP, Terrenghi J. 2017. Groundwater flow in the Venice lagoon and remediation of the Porto Marghera industrial area (Italy). *Hydrogeology Journal*, 25(3): 847–861. https://doi.org/10.1007/s10040-016-1517-5.

Berg M, Tran HC, Nguyen TC, Pham HV, Schertenleib R, Giger W. 2001. Arsenic Contamination of Groundwater and Drinking Water in Vietnam: A Human Health Threat. *Environmental Science & Technology*, 35(13): 2621–2626. https://doi.org/10.1021/es010027y.

Biswas A, Gustafsson JP, Neidhardt H, Halder D, Kundu AK, Chatterjee D, Berner Z, Bhattacharya P. 2014. Role of competing ions in the mobilization of arsenic in groundwater of Bengal Basin: insight from surface complexation modeling. *Water research*, 55: 30–39.

Bondesan A, Primon S, Bassan V, Vitturi A. 2008. *Le unità geologiche della provincia di Venezia*. Provincia di Venezia, Servizio Geologico e Difesa del Suolo.

Carraro A, Fabbri P, Giaretta A, Peruzzo L, Tateo F, Tellini F. 2013. Arsenic anomalies in shallow Venetian Plain (Northeast Italy) groundwater. *Environmental earth sciences*, 70(7): 3067–3084.

Carraro A, Fabbri P, Giaretta A, Peruzzo L, Tateo F, Tellini F. 2015. Effects of redox conditions on the control of arsenic mobility in shallow alluvial aquifers on the Venetian Plain (Italy). *Science of The Total Environment*, 532: 581–594. https://doi.org/10.1016/j.scitotenv.2015.06.003.

Choudhury R, Nath B, Khan MR, Mahanta C, Ellis T, Geen A van. 2018. The Impact of Aquifer Flushing on Groundwater Arsenic Across a 35-km Transect Perpendicular to the Upper Brahmaputra River in Assam, India. *Water Resources Research*, 54(10): 8160–8173. https://doi.org/10.1029/2017WR022485.

Chowdhury TR, Basu GK, Mandal BK, Biswas BK, Samanta G, Chowdhury UK, Chanda CR, Lodh D, Roy SL, Saha KC, Roy S, Kabir S, Quamruzzaman Q, Chakraborti D. 1999. Arsenic poisoning in the Ganges delta. *Nature*, 401(6753): 545–546. https://doi.org/10.1038/44056.

Clay DE, Clapp CE, Linden DR, Molina JAE. 1992. Tillage influence on redox potentials following rainfall. *Soil and Tillage Research*, 22(3): 211–219. https://doi.org/10.1016/0167-1987(92)90038-D.

Dalla Libera N, Fabbri P, Mason L, Piccinini L, Pola M. 2017. Geostatistics as a tool to improve the natural background level definition: An application in groundwater. *Science of The Total Environment*, 598: 330–340. https://doi.org/10.1016/j.scitotenv.2017.04.018.

Dalla Libera N, Fabbri P, Mason L, Piccinini L, Pola M. 2018. A local natural background level concept to improve the natural background level: a case study on the drainage basin of the Venetian Lagoon in Northeastern Italy. *Environmental Earth Sciences*, 77(13). https://doi.org/10.1007/s12665-018-7672-3.

Dalla Libera N, Fabbri P, Piccinini L, Pola M, Mason L. 2016. Natural Arsenic in groundwater in the drainage basin to the Venice lagoon (Brenta Plain, NE Italy): the organic matter's role. *Rendiconti online Soc. Geol. It.*, 41/2016. https://doi.org/10.3301/ROL.2016.85.

Desbarats AJ, Pal T, Mukherjee PK, Beckie RD. 2017. Geochemical Evolution of Groundwater Flowing Through Arsenic Source Sediments in an Aquifer System of West Bengal, India. *Water Resources Research*, 53(11): 8715–8735. https://doi.org/10.1002/2017WR020863.

Dzombak DA, Morel FM. 1990. *Surface complexation modeling: hydrous ferric oxide*. John Wiley & Sons.

Ejarque-Gonzalez E, Butturini A. 2014. Self-Organising Maps and Correlation Analysis as a Tool to Explore Patterns in Excitation-Emission Matrix Data Sets and to Discriminate Dissolved Organic Matter Fluorescence Components. *PLOS ONE*, 9(6): e99618. https://doi.org/10.1371/journal.pone.0099618.

Erickson ML, Elliott SM, Christenson CA, Krall AL. 2018. Predicting geogenic Arsenic in Drinking Water Wells in Glacial Aquifers, North-Central USA: Accounting for Depth-Dependent Features. *Water Resources Research*, 54(12): 10,172-10,187. https://doi.org/10.1029/2018WR023106.

Fabbri P, Gaetan C, Sartore L, Dalla Libera N. 2020. Subsoil Reconstruction in Geostatistics beyond Kriging: A Case Study in Veneto (NE Italy). *Hydrology*. Multidisciplinary Digital Publishing Institute, 7(1): 15. https://doi.org/10.3390/hydrology7010015.

Fabbri P, Gaetan C, Zangheri P. 2011. Transfer function-noise modelling of an aquifer system in NE Italy. *Hydrological Processes*, 25(2): 194–206.

Fabbri P, Piccinini L. 2013. Assessing transmissivity from specific capacity in an alluvial aquifer in the middle Venetian plain (NE Italy). *Water science and technology*, 67(9): 2000–2008.

Fabbri P, Piccinini L, Marcolongo E, Pola M, Conchetto E, Zangheri P. 2016. Does a change of irrigation technique impact on groundwater resources? A case study in Northeastern Italy. *Environmental Science & Policy*. Elsevier, 63: 63–75.

Fontana A, Mozzi P, Bondesan A. 2008. Alluvial megafans in the Venetian–Friulian Plain (north-eastern Italy): Evidence of sedimentary and erosive phases during Late Pleistocene and Holocene. *Quaternary International*, 189(1): 71–90. https://doi.org/10.1016/j.quaint.2007.08.044.

Gao ZP, Jia YF, Guo HM, Zhang D, Zhao B. 2020. Quantifying Geochemical Processes of Arsenic Mobility in Groundwater from an Inland Basin Using a Reactive Transport Model. *Water Resources Research*, In Press: e2019WR025492. https://doi.org/10.1029/2019WR025492.

Harvey CF, Ashfaque KN, Yu W, Badruzzaman ABM, Ali MA, Oates PM, Michael HA, Neumann RB, Beckie R, Islam S, Ahmed MF. 2006. Groundwater dynamics and arsenic contamination in Bangladesh. *Chemical Geology*, 228(1): 112–136. https://doi.org/10.1016/j.chemgeo.2005.11.025.

Harvey CF, Swartz CH, Badruzzaman ABM, Keon-Blute N, Yu W, Ali MA, Jay J, Beckie R, Niedan V, Brabander D, Oates PM, Ashfaque KN, Islam S, Hemond HF, Ahmed MF. 2002. Arsenic Mobility and Groundwater Extraction in Bangladesh. *Science*, 298(5598): 1602–1606. https://doi.org/10.1126/science.1076978.

Healy RW, Cook PG. 2002. Using groundwater levels to estimate recharge. *Hydrogeology Journal*, 10(1): 91–109. https://doi.org/10.1007/s10040-001-0178-0.

Himberg J, Ahola J, Alhoniemi E, Vesanto J, Simula O. 2001. The Self-Organizing Map as a Tool in Knowledge Engineering. *Pattern Recognition in Soft Computing Paradigm*. WORLD SCIENTIFIC, 38–65.

Huang K, Liu Y, Yang C, Duan Y, Yang X, Liu C. 2018. Identification of Hydrobiogeochemical Processes Controlling Seasonal Variations in Arsenic Concentrations Within a Riverbank Aquifer at Jianghan Plain, China. *Water Resources Research*, 54(7): 4294–4308. https://doi.org/10.1029/2017WR022170.

Iseri Y, Matsuura T, Iizuka S, Nishiyama K, Jinno K. 2009. Comparison of pattern extraction capability between self-organizing maps and principal component analysis. *Memoirs of the Faculty of Engineering, Kyushu University*. Citeseer, 69(2): 37–47.

Jakobsen R, Kazmierczak J, Sø HU, Postma D. 2018. Spatial variability of groundwater arsenic concentration as controlled by hydrogeology; Conceptual analysis using 2-D reactive transport modeling. *Water Resources Research*, 0(ja). https://doi.org/10.1029/2018WR023685.

Jessen S, Postma D, Larsen F, Nhan PQ, Trang PTK, Long TV, Viet PH, Jakobsen R. 2012. Surface complexation modeling of groundwater arsenic mobility: Results of a forced gradient experiment in a Red River flood plain aquifer, Vietnam. *Geochimica et Cosmochimica Acta*, 98: 186–201.

Kalteh AM, Hjorth P, Berndtsson R. 2008. Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environmental Modelling & Software*, 23(7): 835–845. https://doi.org/10.1016/j.envsoft.2007.10.001.

Karafistan A, Gemikonakli E. 2020. Unsupervised classification of the intrinsic and morphological properties of quasars through self-organizing maps. *Astronomische Nachrichten*, 341(4): 471–477. https://doi.org/10.1002/asna.202013681.

Kinniburgh DG, Smedley PL (eds). 2001. *Arsenic contamination of groundwater in Bangladesh*. British Geological Survey: Keyworth.

Kohonen T. 1990. The self-organizing map. *Proceedings of the IEEE*, 78(9): 1464–1480. https://doi.org/10.1109/5.58325.

Kohonen T. 2001. *Self-Organizing Maps*. Springer-Verlag: Berlin Heidelberg.

Kohonen T. 2013. Essentials of the self-organizing map. *Neural Networks*, 37: 52–65. https://doi.org/10.1016/j.neunet.2012.09.018.

Li T, Sun G, Yang C, Liang K, Ma S, Huang L. 2018. Using self-organizing map for coastal water quality classification: Towards a better understanding of patterns and processes. *Science of The Total Environment*, 628–629: 1446–1459. https://doi.org/10.1016/j.scitotenv.2018.02.163.

Manca G, Cervone G. 2013. The case of arsenic contamination in the Sardinian Geopark, Italy, analyzed using symbolic machine learning. *Environmetrics*, 24(6): 400–406. https://doi.org/10.1002/env.2222.

Mayer KU, Frind EO, Blowes DW. 2002. Multicomponent reactive transport modeling in variably saturated porous media using a generalized formulation for kinetically controlled reactions: REACTIVE TRANSPORT MODELING IN VARIABLY SATURATED MEDIA. *Water Resources Research*, 38(9): 13-1-13–21. https://doi.org/10.1029/2001WR000862.

McArthur JM, Ravenscroft P, Safiulla S, Thirlwall MF. 2001. Arsenic in groundwater: Testing pollution mechanisms for sedimentary aquifers in Bangladesh. *Water Resources Research*, 37(1): 109–117. https://doi.org/10.1029/2000WR900270.

Melo DS, Gontijo ESJ, Frascareli D, Simonetti VC, Machado LS, Barth JAC, Moschini-Carlos V, Pompêo ML, Rosa AH, Friese K. 2019. Self-Organizing Maps for Evaluation of Biogeochemical Processes and Temporal Variations in Water Quality of Subtropical Reservoirs. *Water Resources Research*, 55(12): 10268–10281. https://doi.org/10.1029/2019WR025991.

Michael HA, Khan MR. 2016. Impacts of physical and chemical aquifer heterogeneity on basin-scale solute transport: Vulnerability of deep groundwater to arsenic contamination in Bangladesh. *Advances in Water Resources*, 98: 147–158. https://doi.org/10.1016/j.advwatres.2016.10.010.

Molinari A, Guadagnini L, Marcaccio M, Straface S, Sanchez-Vila X, Guadagnini A. 2013. Arsenic release from deep natural solid matrices under experimentally controlled redox conditions. *Science of The Total Environment*, 444: 231–240. https://doi.org/10.1016/j.scitotenv.2012.11.093.

Monod J. 1949. The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1): 371–394. https://doi.org/10.1146/annurev.mi.03.100149.002103.

Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*. American society of agricultural and biological engineers, 50(3): 885–900.

Park Y, Ligaray M, Kim YM, Kim JH, Cho KH, Sthiannopkao S. 2016. Development of enhanced groundwater arsenic prediction model using machine learning approaches in Southeast Asian countries. *Desalination and Water Treatment*. Taylor & Francis, 57(26): 12227–12236. https://doi.org/10.1080/19443994.2015.1049411.

Parkhurst DL, Appelo CAJ. 2013. Description of Input and Examples for PHREEQC Version 3—A Computer Program for Speciation, Batch-Reaction, One-Dimensional Transport, and Inverse Geochemical Calculations. Chapter 43 of Section A, Groundwater Book 6, Modeling Techniques. U.S. Geological Survey, Denver, Colorado. .

Parkhurst DL, Kipp KL, Engesgaard P, Charlton SR. 2004. Phast, a program for simulating ground-waterflow, solute transport, and multicomponent geochemical reactions. *US Geological survey techniques and methods*, 6: A8.

Pedretti D, Fiori A. 2013. Travel time distributions under convergent radial flow in heterogeneous formations: Insight from the analytical solution of a stratified model.

Advances in Water Resources, 60: 100–109. https://doi.org/10.1016/j.advwatres.2013.07.013.

Pedretti D, Luoma S, Ruskeeniemi T, Backman B. 2019. A geologically-based approach to map arsenic risk in crystalline aquifers: Analysis of the Tampere region, Finland. *Geoscience Frontiers*. https://doi.org/10.1016/j.gsf.2018.12.004.

Pedretti D, Marco Masetti, Giovanni P. Beretta, Mariangela Vitiello. 2013. A Revised Conceptual Model to Reproduce the Distribution of Chlorinated Solvents in the Rho Aquifer (Italy). *Groundwater Monitoring & Remediation*, 33(3): 69–77. https://doi.org/10.1111/gwmr.12017.

Pedretti D, Molinari A, Fallico C, Guzzi S. 2016. Implications of the change in confinement status of a heterogeneous aquifer for scale-dependent dispersion and mass-transfer processes. *Journal of Contaminant Hydrology*, 193: 86–95. https://doi.org/10.1016/j.jconhyd.2016.09.005.

Piccinini L, Fabbri P, Pola M. 2016. Point dilution tests to calculate groundwater velocity: an example in a porous aquifer in northeast Italy. *Hydrological Sciences Journal*, 61(8): 1512–1523.

Piccinini L, Fabbri P, Pola M, Marcolongo E. 2017. An example of aquifer heterogeneity simulation to modeling well-head protection areas. *Ital J Eng Geol Environ Special*, (103–115).

Postma D, Larsen F, Minh Hue NT, Duc MT, Viet PH, Nhan PQ, Jessen S. 2007. Arsenic in groundwater of the Red River floodplain, Vietnam: Controlling geochemical processes and reactive transport modeling. *Geochimica et Cosmochimica Acta*, 71(21): 5054–5071. https://doi.org/10.1016/j.gca.2007.08.020.

Prommer H, Barry D a., Zheng C. 2003. MODFLOW/MT3DMS-Based Reactive Multicomponent Transport Modeling. *Ground Water*, 41(2): 247–257. https://doi.org/10.1111/j.1745-6584.2003.tb02588.x.

Ramesh Kumar A, Riyazuddin P. 2012. Seasonal variation of redox species and redox potentials in shallow groundwater: A comparison of measured and calculated redox potentials. *Journal of Hydrology*, 444–445: 187–198. https://doi.org/10.1016/j.jhydrol.2012.04.018.

Ramos ME, Cappelli C, Rozalén M, Fiore S, Huertas FJ. 2011. Effect of lactate, glycine, and citrate on the kinetics of montmorillonite dissolution. *American Mineralogist*. Mineralogical Society of America, 96(5–6): 768–780.

Rathi B, Siade AJ, Donn MJ, Helm L, Morris R, Davis JA, Berg M, Prommer H. 2017. Multiscale Characterization and Quantification of Arsenic Mobilization and Attenuation During Injection of Treated Coal Seam Gas Coproduced Water into Deep Aquifers. *Water Resources Research*, 53(12): 10779–10801. https://doi.org/10.1002/2017WR021240.

Rawson J, Prommer H, Siade A, Carr J, Berg M, Davis JA, Fendorf S. 2016. Numerical modeling of arsenic mobility during reductive iron-mineral transformations. *Environmental science & technology*, 50(5): 2459–2467.

Reusch DB, Alley RB, Hewitson BC. 2005. Relative Performance of Self-Organizing Maps and Principal Component Analysis in Pattern Extraction from Synthetic Climatological Data. *Polar Geography*. Taylor & Francis, 29(3): 188–212. https://doi.org/10.1080/789610199.

Richards LA, Magnone D, Sültenfuß J, Chambers L, Bryant C, Boyce AJ, van Dongen BE, Ballentine CJ, Sovann C, Uhlemann S, Kuras O, Gooddy DC, Polya DA. 2019. Dual in-aquifer and near surface processes drive arsenic mobilization in Cambodian groundwaters. *Science of The Total Environment*, 659: 699–714. https://doi.org/10.1016/j.scitotenv.2018.12.437.

Rotiroti M, Sacchi E, Fumagalli L, Bonomi T. 2014. Origin of Arsenic in Groundwater from the Multilayer Aquifer in Cremona (Northern Italy). *Environmental Science & Technology*, 48(10): 5395–5403. https://doi.org/10.1021/es405805v.

Sahoo S, Russo TA, Elliott J, Foster I. 2017. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US. *Water Resources Research*, 53(5): 3878–3895.

Sartore L, Fabbri P, Gaetan C. 2016. spMC: an R-package for 3D lithological reconstructions based on spatial Markov chains. *Computers & Geosciences*, 94: 40–47. https://doi.org/10.1016/j.cageo.2016.06.001.

Saunders JA, Lee M-K, Shamsudduha M, Dhakal P, Uddin A, Chowdury MT, Ahmed KM. 2008. Geochemistry and mineralogy of arsenic in (natural) anaerobic groundwaters. *Applied Geochemistry*, 23(11): 3205–3214.

Singh SK, Taylor RW, Rahman MM, Pradhan B. 2018. Developing robust arsenic awareness prediction models using machine learning algorithms. *Journal of Environmental Management*, 211: 125–137. https://doi.org/10.1016/j.jenvman.2018.01.044.

Smedley PL, Kinniburgh DG. 2013. Arsenic in Groundwater and the Environment. *Essentials in Medical Geology*, 279–310.

Smith R, Knight R, Fendorf S. 2018. Overpumping leads to California groundwater arsenic threat. *Nature Communications*, 9(1). https://doi.org/10.1038/s41467-018-04475-3.

Solomatine DP, Shrestha DL. 2009. A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Research*, 45(12).

Sracek O, Bhattacharya P, Jacks G, Gustafsson J-P, Brömssen M von. 2004. Behavior of arsenic and geochemical modeling of arsenic enrichment in aqueous environments. *Applied Geochemistry*, 19(2): 169–180. https://doi.org/10.1016/j.apgeochem.2003.09.005.

Steefel CI, DePaolo DJ, Lichtner PC. 2005. Reactive transport modeling: An essential tool and a new research approach for the Earth sciences. *Earth and Planetary Science Letters*, 240(3): 539–558. https://doi.org/10.1016/j.epsl.2005.09.017.

Ungaro F, Ragazzi F, Cappellin R, Giandon P. 2008. Arsenic concentration in the soils of the Brenta Plain (Northern Italy): Mapping the probability of exceeding contamination thresholds. *Journal of Geochemical Exploration*, 96(2): 117–131. https://doi.org/10.1016/j.gexplo.2007.03.006.

Vesanto J, Ahola J. 1999. Hunting for correlations in data using the self-organizing map. *International ICSC congress on computational intelligence methods and applications, 1999*. ICSC Academic Press.

Vesanto J, Himberg J, Alhoniemi E, Parhankangas J. 1999. Self-organizing map in Matlab: the SOM Toolbox. *Proceedings of the Matlab DSP conference*, 16–17.

Vorlicek PA, Antonelli R, Fabbri P, Rausch R. 2004. Quantitative hydrogeological studies of the Treviso alluvial plain, NE Italy. *Quarterly Journal of Engineering Geology and Hydrogeology*. Geological Society of London, 37(1): 23–29.

Wallis I, Prommer H, Pichler T, Post V, B. Norton S, Annable MD, Simmons CT. 2011. Process-based reactive transport model to quantify arsenic mobility during aquifer storage and recovery of potable water. *Environmental science & technology*, 45(16): 6924–6931.

Wehrens R, Buydens LM. 2007. Self-and super-organizing maps in R: the Kohonen package. *Journal of Statistical Software*, 21(5): 1–19.

Wehrens R, Kruisselbrink J. 2018. Flexible Self-Organising Maps in kohonen 3.0. *Journal of Statis*.

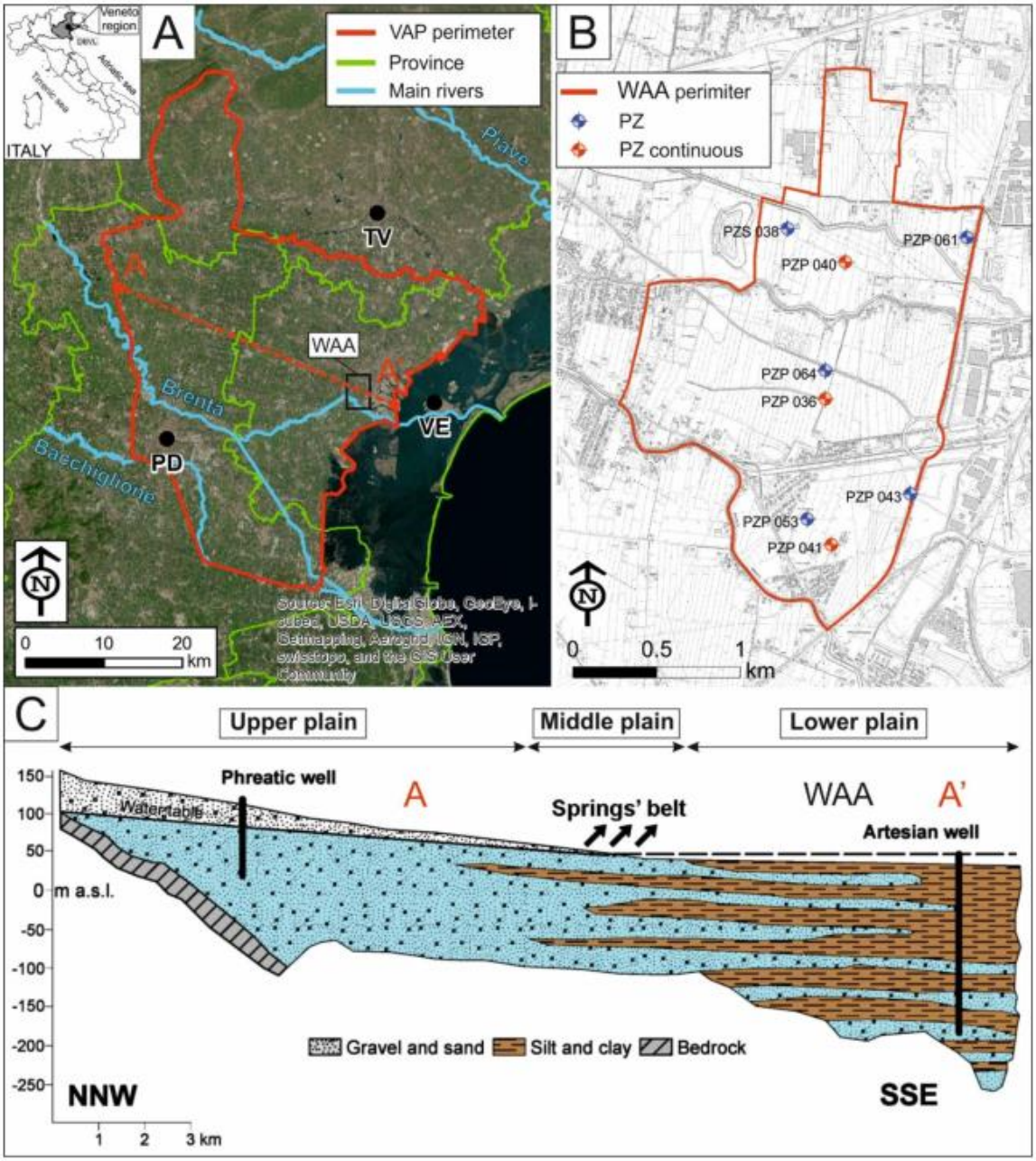

Figure 1 A) The red line marks the limits of the Venetian Alluvial Plain (VAP), located between Padova (PD), Treviso (TV) and Venice (VE). The VAP is crossed by the Brenta River near its entrance to the Venice Lagoon. The small black square shows the location of the Western Agricultural Area (WAA). B) A topographic map of the WAA, showing the eight piezometers used in this study, and the three red piezometers (PZP036, PZP040 and PZP041) with more frequent analyses. C) A schematic geological cross-section of the VAP (modified from Piccinini et al., 2017).

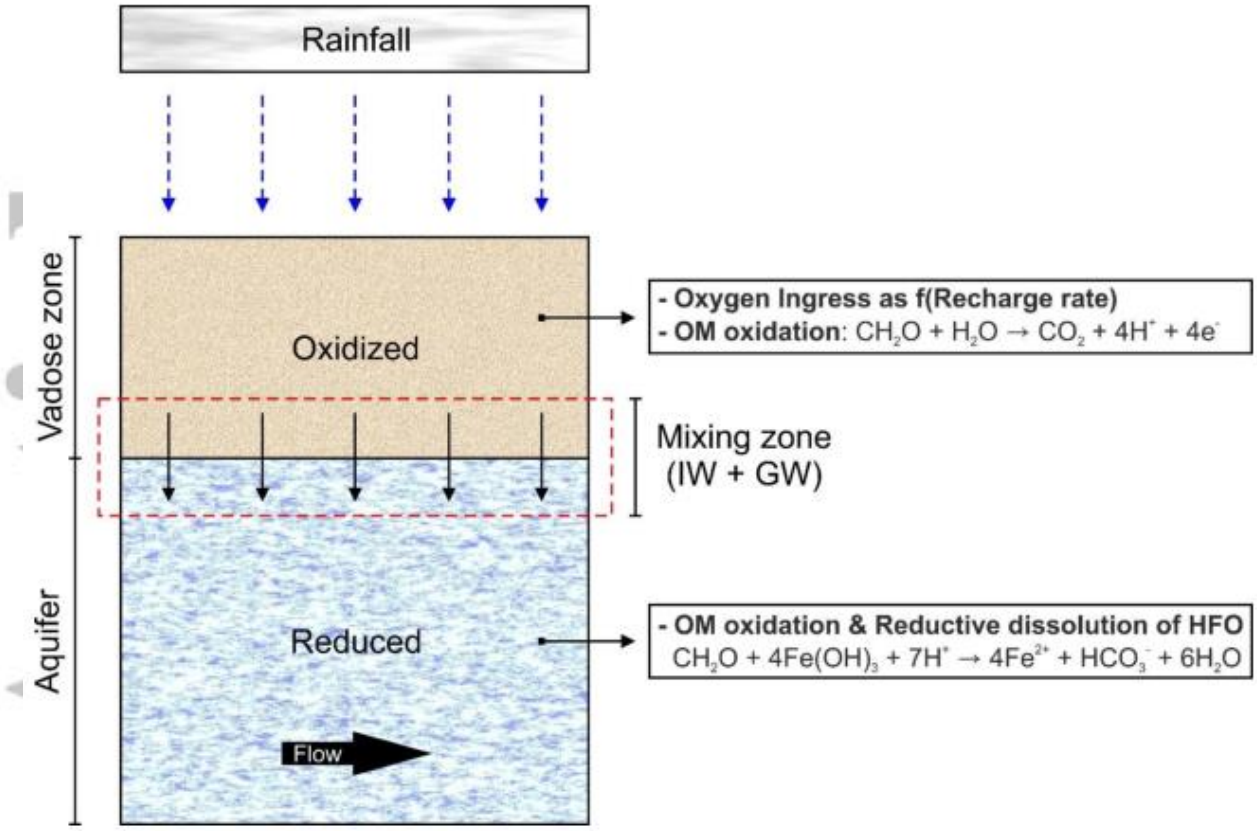

Figure 2 General conceptual model proposed to explain the main processes controlling As mobility in the studied domain for the shallow aquifers at the scale of the VAP.

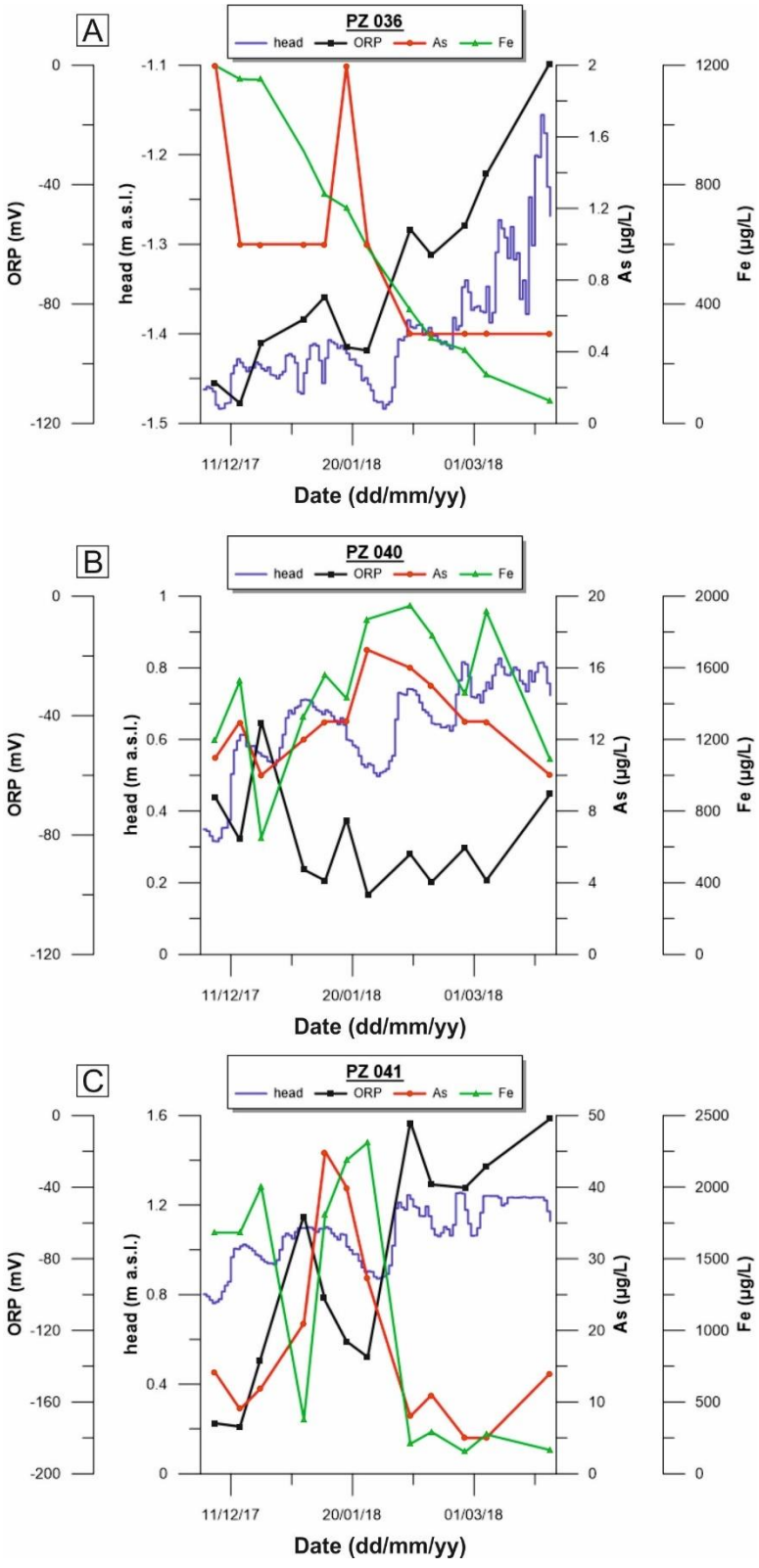

Figure 3 Time series of measured variations of groundwater head, oxidizing-reduction potential (ORP), arsenic (As) and iron (Fe) concentrations for the three piezometers PZP 036 (A), 040 (B) and 041 (C). with weekly geochemical monitoring and continuous groundwater head level logging.

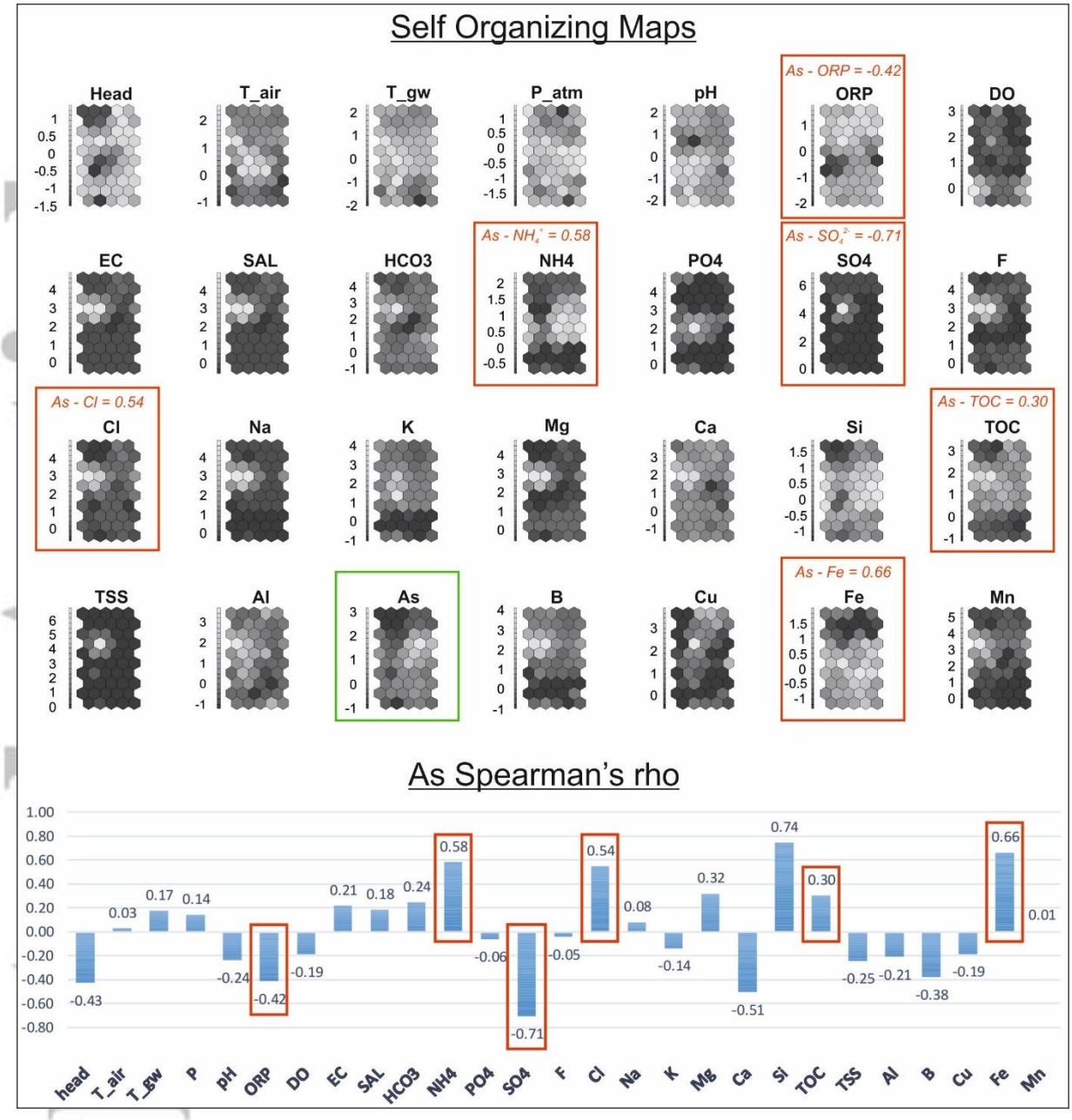

Figure 4 At the top, component planes of the Self-Organized Map (SOM) analysis for each input variable. The vector space of the SOM is represented through a toroidal lattice (5x10) made of hexagonal cells. The maps show the z-score normalized values of the variables with a grayscale. The Spearman's correlation coefficient ρ values for selected variables postulated to be involved in the geochemical process controlling As mobility. At the bottom, ρ is reported for all variables.

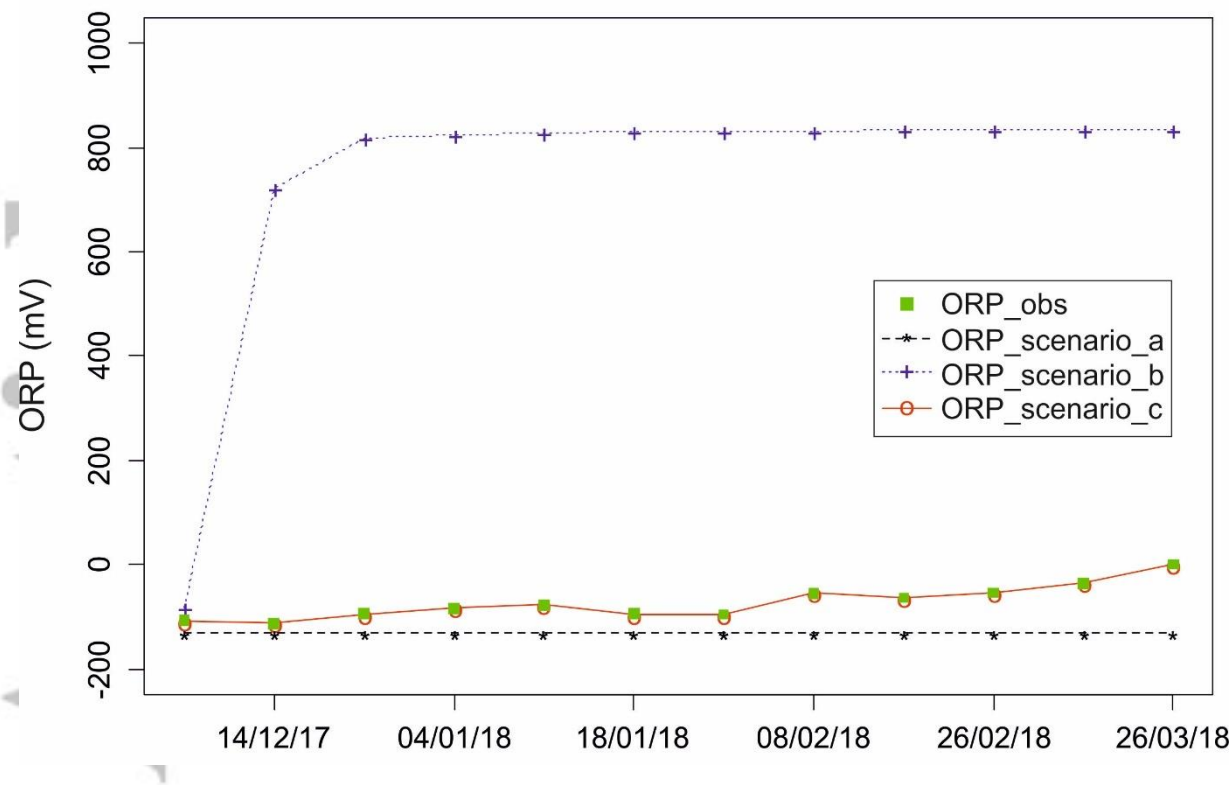

Figure 5 Comparison between measured ORP at PZP036 and simulated using PHREEQC models embedding the three different scenarios (a,b,c). Each scenario describes a specific modality of oxidants transfer from the surface to the aquifer.

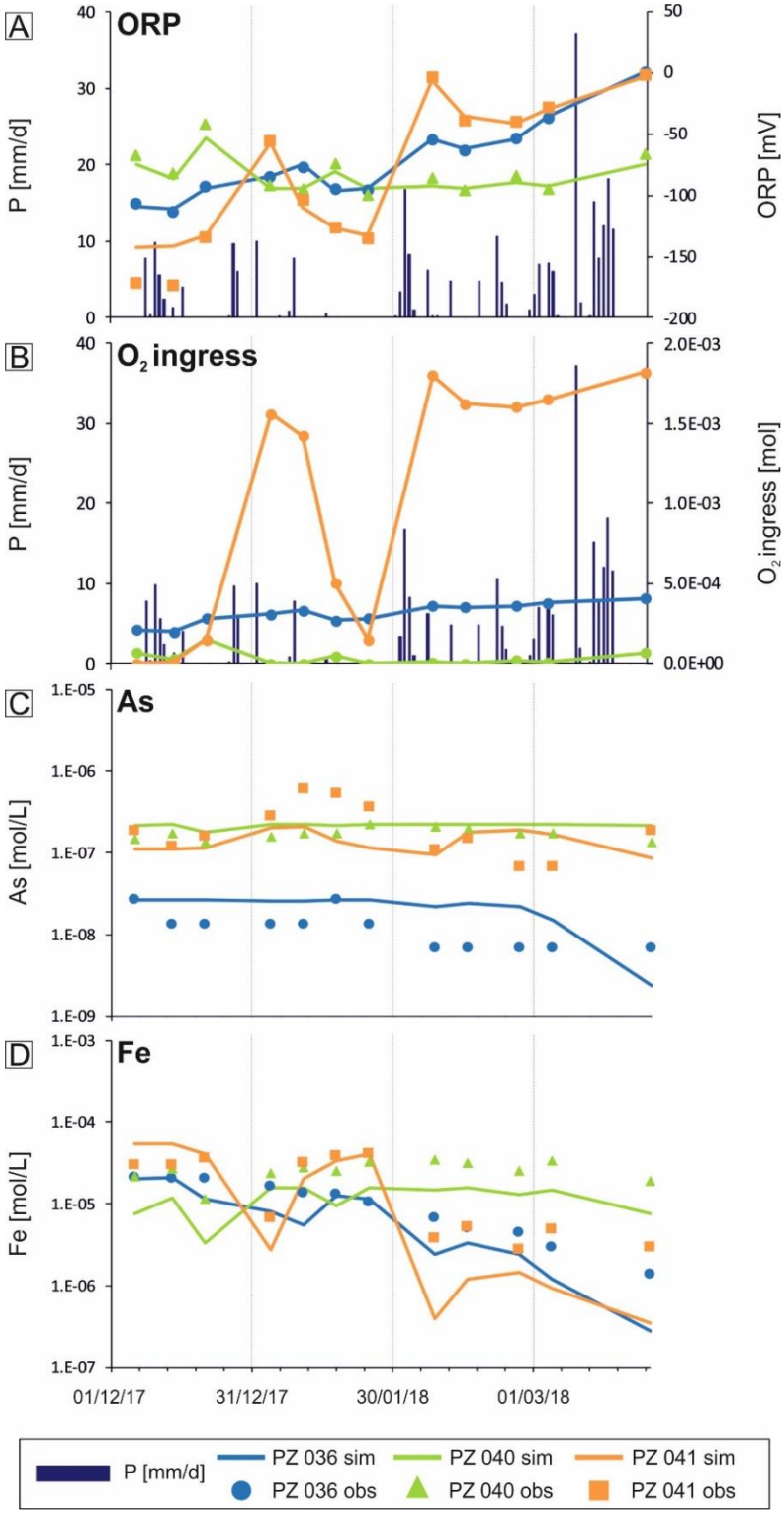

Figure 6 Comparison between observed (symbols) and simulated (lines) variables for PZP036, PZP040 and PZP041. Panel (A) shows the comparison for ORP. Panel (B) reproduces the shape of the rate-limiting function that enables reproducing the observed ORP at the three piezometers. The observed daily rainfall depths (P) is reported in both panels A and B. Panel (C) shows the comparisong for As, and Panel (D) for Fe.

Table 1 Summary table of the WAA dataset. The column "head" represents the hydraulic head. The extended name of the variables can be found in the text.

	m a.s.l.	*°C*	*°C*	*mbar*		*mV*	*mg/L*	*µS/cm*	*ppt*	*mg/L*	*mg/L*	*mg/L*	*mg/L*	*mg/L*
	head	**T_air**	**T_gw**	**P_atm**	**pH**	**ORP**	**DO**	**EC**	**SAL**	**HCO3**	**NH4**	**PO4**	**SO4**	**F**
Min	-1.28	-0.60	11.20	998.90	6.68	-173.70	0.09	734.00	0.41	475.00	0.04	0.02	0.50	0.10
1st_Q	-1.03	5.83	12.93	1010.20	6.88	-102.62	0.13	876.80	0.49	522.40	0.14	0.02	20.50	0.18
Median	0.23	8.15	13.95	1018.80	6.95	-84.70	0.17	881.50	0.49	543.40	0.46	0.02	42.00	0.20
Mean	-0.14	9.39	13.66	1019.10	6.96	-83.35	0.26	947.70	0.53	553.70	1.04	0.05	66.61	0.28
3rd_Q	0.44	10.45	14.20	1028.80	7.01	-55.15	0.26	887.50	0.50	565.40	2.07	0.07	73.75	0.29
Max	0.85	23.90	17.00	1035.00	7.23	0.90	1.12	2481.00	1.45	827.00	3.73	0.34	1125.00	1.60
	mg/L	*mg/L*	*mg/L*	*mg/L*	*mg/L*	*mg/L*	*mg/L*	*mg/L*	*µg/L*	*µg/L*	*µg/L*	*µg/L*	*µg/L*	*µg/L*
	Cl	**Na**	**K**	**Mg**	**Ca**	**Si**	**TOC**	**TSS**	**Al**	**As**	**B**	**Cu**	**Fe**	**Mn**
Min	21.00	18.80	0.50	29.00	46.30	3.95	0.60	2.50	0.50	0.50	8.00	0.50	75.00	109.00
1st_Q	24.00	19.73	0.80	34.20	113.80	5.07	1.10	5.00	1.00	2.25	14.25	0.50	487.50	129.00
Median	33.50	26.50	2.00	36.75	123.00	6.81	1.45	9.00	1.00	11.00	64.50	0.50	1248.50	160.00
Mean	36.66	30.78	1.90	43.27	123.40	6.91	1.98	189.60	1.17	11.33	61.50	0.86	1232.00	183.90
3rd_Q	43.00	29.73	2.30	46.95	137.60	8.44	2.70	30.00	1.00	14.75	82.75	1.00	1800.20	198.20
Max	135.00	151.60	7.20	140.80	294.00	10.41	5.90	5925.00	3.00	45.00	266.00	3.00	3222.00	696.00

Table 2 Chemical composition of the background groundwater and the rainfall used as input for the PHREEQC models.

Parameter	PZP 036	PZP 040	PZP 041	Rainfall (Pieri et al., 2012)
T [°C]	14.200	15.000	14.300	14.2
pH	7.220	7.140	7.100	6.55
pe	-1.870	-1.180	-3.010	4 (function of O2 atm)
Chemicals	*[mg/L]*	*[mg/L]*	*[mg/L]*	*[µmol/L]*
Al	0.001	0.002	0.001	\
Alkalinity (as HCO_3^-)	533.600	555.600	572.700	\
As	0.002	0.011	0.014	\
B	0.083	0.013	0.077	\
Ca	136.700	123.700	117.900	105.22
Cl	23.000	32.000	45.000	63.3
F	0.180	0.280	0.160	\
Fe	1.198	1.203	1.682	\
H(1)	\	\	\	0.3
Mg	33.700	46.000	34.900	37.95
Mn	0.160	0.139	0.194	\
N(+5)	\	\	\	62.25
N(-3)	0.770	0.150	3.730	\
Na	27.800	18.800	31.700	97.8
P	0.030	0.015	0.030	\
S(6)	80.000	24.000	17.000	41.95
Si	5.190	7.330	10.410	\