Oscillatory Pumping Test to Estimate Aquifer Hydraulic Parameters in a Bayesian Geostatistical Framework

Mathematical Geosciences

# Oscillatory pumping test to estimate aquifer hydraulic parameters in a Bayesian Geostatistical framework

--Manuscript Draft--

| | |
|---|---|
| Manuscript Number: | MATG-D-16-00175R2 |
| Full Title: | Oscillatory pumping test to estimate aquifer hydraulic parameters in a Bayesian Geostatistical framework |
| Article Type: | S.I. : geoENV2016 |
| Keywords: | Oscillatory pumping test;  Bayes, geostatistical approach, hydraulic parameter estimation, groundwater |
| Corresponding Author: | Marco D'Oria<br>Universita degli Studi di Parma<br>Parma, ITALY |
| Corresponding Author Secondary Information: | |
| Corresponding Author's Institution: | Universita degli Studi di Parma |
| Corresponding Author's Secondary Institution: | |
| First Author: | Marco D'Oria |
| First Author Secondary Information: | |
| Order of Authors: | Marco D'Oria |
| | Andrea Zanini |
| | Fausto Cupola |
| Order of Authors Secondary Information: | |
| Funding Information: | |
| Abstract: | Comprehensive information about the spatial distribution of the subsurface hydraulic properties is crucial to model groundwater flow, to predict solute transport in aquifers and to design remediation actions. In this work, a Bayesian Geostatistical approach, as implemented in bgaPEST, was adopted to estimate the hydraulic properties of a well field located at the Campus of Science and Technology of the University of Parma (Northern Italy), in a contest of a highly parameterized inversion. Head data, collected by means of multi frequency oscillatory pumping tests, were used to both estimate the hydraulic parameters and validate the results. The groundwater flow processes were modelled by means of MODFLOW 2005 and an adjoint-state formulation of the same software was used to efficiently calculate the sensitivity matrix, required by the inverse procedure. The Bayesian Geostatistical approach estimated the hydraulic conductivity and specific storage fields, handling a large number of parameters. The results of the inversion are consistent with the alluvial nature of the investigated aquifer and the preliminary traditional pumping tests carried out at the site. |

# Title: Oscillatory pumping test to estimate aquifer hydraulic parameters in a Bayesian Geostatistical framework

Received: / Accepted:

*Authors*: D'Oria Marco, Zanini Andrea, Cupola Fausto

*Affiliations*: Department of Engineering and Architecture, University of Parma. Parma, Italy

marco.doria@unipr.it; andrea.zanini@unipr.it; fausto.cupola@gmail.com

*Corresponding author*: Marco D'Oria, Email: marco.doria@unipr.it, Phone: 0039 0521 906335, Fax: 0039 0521 905912

## Abstract

Comprehensive information about the spatial distribution of the subsurface hydraulic properties is crucial to model groundwater flow, to predict solute transport in aquifers and to design remediation actions. In this work, a Bayesian Geostatistical approach, as implemented in bgaPEST, was adopted to estimate the hydraulic properties of a well field located at the Campus of Science and Technology of the University of Parma (Northern Italy), in a contest of a highly parameterized inversion. Head data, collected by means of multi frequency oscillatory pumping tests, were used to both estimate the hydraulic parameters and validate the results. The groundwater flow processes were modelled by means of MODFLOW 2005 and an adjoint-state formulation of the same software was used to efficiently calculate the sensitivity matrix, required by the inverse procedure. The Bayesian Geostatistical approach estimated the hydraulic conductivity and specific storage fields, handling a large number of parameters. The results of the inversion are consistent with the alluvial nature of the investigated aquifer and the preliminary traditional pumping tests carried out at the site.

## Keywords

Oscillatory pumping test; Bayes; geostatistical approach; hydraulic parameter estimation; groundwater

# 1 Introduction

The knowledge of spatial distribution of the hydraulic parameters, such as hydraulic conductivity and specific storage, is of main importance in groundwater modeling (Zhou et al. 2014). In particular, it allows to accurately forecast the fate of solutes in aquifers (Alberti et al. 2011), to estimate the source location and its release in time and to design remediation actions (Snodgrass and Kitanidis 1997; Cupola et al. 2015; Zanini and Woodbury 2016). In the last decades, different methods have been proposed and adopted to collect data useful for aquifer characterization and an extensive review is available in Cardiff et al. (2012). Pumping tests, and measurement of the associated changing in head, are still largely used to obtain such information. In addition to traditional pumping tests, with constant flow rate, sequential aquifer tests (hydraulic tomography) have been used in the last 20 years. A comprehensive literature review about hydraulic tomography, that analyzes advantages and drawbacks, is proposed by Cardiff and Barrash (2011). According to the Authors, most of the studies consider pumping with constant flow rate. On the other hand, recently, Cardiff et al. (2013) used oscillatory pumping tests for aquifer heterogeneity characterization; the data were collected using a periodic signal, in which water is extracted during half a period and then reinjected in the other half. The Authors propose to use several pumping frequencies in order to provide information about aquifer heterogeneity. Recently Zhou et al. (2016) have tested this approach in a laboratory controlled environment. The Authors stimulated the "laboratory aquifer" through injection and extraction of water with a sinusoidal wave. Only few applications on real case study are available in the literature (Rabinovich et al. 2015).

Once the data are collected, solving an inverse problem is a widely used method to characterize the aquifer hydraulic parameters (e.g. hydraulic conductivity, specific storage, etc.). In the literature, many approaches, both deterministic and stochastic, have been adopted to this aim and several review papers are available on the inverse problems. The works of Yeh (1986), McLaughlin and Townley (1996), Zimmerman et al. (1998), Carrera et al. (2005), Hendricks Franssen et al. (2009), Zhou et al. (2012, 2014) and Zanini et al. (2017) provide a complete, even if not exhaustive, overview of the available inverse procedures (Kitanidis 1997; Tarantola 2005).

In this work, the hydraulic parameters (hydraulic conductivity and specific storage) of a well field located at the University of Parma (Northern Italy) were estimated. At this aim, the Bayesian Geostatistical approach, developed by Kitanidis (1995), as implemented in bgaPEST (Fienen et al. 2013) was adopted.

Bayesian approaches using geostatistical simulations have been presented by several authors, such as Hansen et al. (2006), Rimstad et al. (2012), Grana et al. (2017) and Soares et al. (2017). Head data collected by means of multi frequency oscillatory pumping tests (Cardiff et al. 2013) were used in the inversion process. The oscillatory pumping tests were performed using only commercial equipment; this is a novelty with respect to previous studies. Two series of data were collected: the first (which has a period of 10 minutes) was used for estimation, while the second (with a period of 5 minutes) for validation.

The remainder of this paper is organized as follows. After a description of the inverse methodology in Sect. 2, information about the test site, the pumping tests, the numerical model and the observations and parameters is reported in Sect. 3, results are presented in Sect. 4 and finally conclusions are drawn in Sect. 5.

# 2 Methods

In this work, a Bayesian approach to the inverse problem combining prior information on the unknown parameters (aquifer hydraulic parameters) with available field data (observed head data) has been used. Different Bayesian methods are available in literature to solve inverse problems, such as the ensemble Kalman filter (EnKF, Evensen 2009; Xu et al. 2016) and its smoother variants (Chen and Oliver 2012; Emerick and Reynolds 2013; Li et al. 2015), the method of anchored distributions (MAD, Rubin et al. 2010), the maximum a posteriori method (MAP, McLaughlin and Townley 1996) and the Bayesian Geostatistical approach (Kitanidis and Vomvoris 1983), among others.

In the Bayesian Geostatistical (BG) approach (Kitanidis 1995), the term Bayesian refers to the inverse theory aspect, whereas Geostatistical deals with the way in which prior information about the unknown parameters is used (Fienen et al. 2013). In particular, the BG approach makes use of autocorrelated spatial or temporal functions (the correlation is specified by means of a prior covariance matrix or a generalized covariance function, see Kitanidis (1993)) to provide prior information about the parameters. Usually, the covariance model is a function of the separation distance (in space or time) between the parameters and describes their deviations from the mean behavior (Michalak et al. 2004).

Hence, BG is a flexible highly parameterized inversion method suitable when the unknown parameters are distributed in space or time and are correlated with one another; highly parameterized models are characterized by having more parameters than can be estimated uniquely on the basis of a given observation dataset, making the inverse problem ill-posed (Doherty and Hunt 2010). The inverse method was developed in the context of estimating spatial parameter fields (Kitanidis and Vomvoris 1983; Hoeksema and Kitanidis

1984; Fienen et al. 2009, among others) but has also been successfully used to evaluate auto-correlated time functions in different fields (e.g. Snodgrass and Kitanidis 1997; Michalak et al. 2004; Butera et al. 2013; D'Oria et al. 2014; D'Oria et al. 2015; Leonhardt et al. 2014). In the following, a brief description of the inverse procedure is introduced; detailed information is available in Kitanidis (1995) and Nowak and Cirpka (2004), for example.

The estimation problem is usually expressed considering the measurement equation $\mathbf{y} = \mathbf{h}(\mathbf{s}) + \mathbf{r}$ that relates the vector of the available data $\mathbf{y}$ [$m \times 1$], with $\mathbf{s}$ [$n \times 1$], the state vector (unknown parameters in the inverse problem); $\mathbf{h}(\mathbf{s})$ represents the modeled values at the same locations and times of the observed data $\mathbf{y}$ (using the forward model) and $\mathbf{r}$ [$m \times 1$] is the epistemic error vector. According to the Bayes' theorem, the posterior probability density function (*pdf*) of the parameters, for given data, $p(\mathbf{s}|\mathbf{y})$ may be written as

$$p(\mathbf{s}|\mathbf{y}) \propto p(\mathbf{s}) L(\mathbf{y}|\mathbf{s}), \tag{1}$$

where $p(\mathbf{s})$ is the prior *pdf* of $\mathbf{s}$ and $L(\mathbf{y}|\mathbf{s})$ is the likelihood function, $\propto$ indicates proportionality. In the BG approach, the unknown parameters are a-priori assumed normally distributed and characterized by their mean $E[\mathbf{s}] = \mathbf{X}\boldsymbol{\beta}$ and covariance $E\left[(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\right] = \mathbf{Q}_{\mathbf{ss}}$, where $E$ means the expected value, $\mathbf{X}$ [$n \times p$] is a known matrix of basis functions, $\boldsymbol{\beta}$ [$p \times 1$] is a vector of drift coefficients and $\mathbf{Q}_{\mathbf{ss}}$ [$n \times n$] indicates the prior covariance matrix. Multiple values of $\boldsymbol{\beta}$ can be used to both introduce a partitioning of the estimated area by imposing discontinuities in the stochastic field and allow the estimation of different parameter types (for example, hydraulic conductivity and specific storage). In this way different parameter types and/or different regions of the field (homogeneous or not) correspond to distinct mean values. The matrix of basis function, $\mathbf{X}$, maps each unknown parameter with the appropriate mean value and if necessary can be used to express a trend in the estimated fields. The prior covariance matrix is therefore censored by assigning zero values where elements of different regions or parameter types are considered. Considering that no information is available on the spatial variability of the hydraulic parameters and the alluvial nature of the investigated aquifer, the exponential covariance function (Eq. (2)) has been selected as the best compromise to characterize the hydraulic parameters of the well field.

$$\mathbf{Q}_{\mathbf{ss}}(\sigma^2, l, \mathbf{d}) = \sigma^2 \exp\left(-\frac{|\mathbf{d}|}{l}\right), \tag{2}$$

where $\sigma^2$ is the variance, $l$ is the integral scale (range parameter) and $|\mathbf{d}|$ is the vector that contains the separation distances between nodes. The exponential model enforces continuity to the solution and provides a sill in the overall variance; it does not introduce smoothness constraints, allowing short range variability (Michalak et al. 2004).

The likelihood function is also assumed normally distributed and characterizes the errors, which are modeled as a random process with zero mean and covariance matrix $\mathbf{R}$. In case the epistemic errors are independent and identically distributed, the covariance matrix assumes the form $\mathbf{R} = \sigma_R^2 \mathbf{I}$, where $\sigma_R^2$ is the variance and $\mathbf{I}$ [$m \times m$] is the identity matrix.

With the assumptions made, the posterior *pdf* assumes the form

$$p\left(\mathbf{s} \middle| \mathbf{y}, \boldsymbol{\theta}, \sigma_R^2\right) \propto \underbrace{\exp\left(-\frac{1}{2}(\mathbf{s}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}} \mathbf{Q}_{\mathbf{ss}}^{-1}(\mathbf{s}-\mathbf{X}\boldsymbol{\beta})\right)}_{\text{Prior term}} \underbrace{\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{h}(\mathbf{s}))^{\mathrm{T}} \mathbf{R}^{-1}(\mathbf{y}-\mathbf{h}(\mathbf{s}))\right)}_{\text{Likelihood term}}, \quad (3)$$

where $\boldsymbol{\theta}$ is the vector of the structural parameters (parameters of the prior covariance models, $\sigma^2$ and $l$ in this case). The solution of the inverse problem requires to obtain the $\boldsymbol{\beta}$ and $\mathbf{s}$ vectors; in addition, also the parameters of $\mathbf{Q}_{\mathbf{ss}}$ and $\mathbf{R}$ ($\boldsymbol{\theta}$ and $\sigma_R^2$) can be estimated (Michalak et al. 2004); this is accomplished in two distinct steps.

In the first step, the posterior best estimate of $\mathbf{s}$ ($\hat{\mathbf{s}}$) and $\boldsymbol{\beta}$ ($\hat{\boldsymbol{\beta}}$) is obtained maximizing, for given structural parameters and epistemic error variance, the Eq. (3) and represents the most likely values obtainable conditional on the data. In case the relation between the parameters and the observations is linear ($\mathbf{Hs}$ replaces $\mathbf{h}(\mathbf{s})$), the solution can be found solving the system of Eqs. (4) and (5) in the form of the so-called cokriging equations (Kitanidis and Lee 2014)

$$\hat{\mathbf{s}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Q}_{\mathbf{ss}}\mathbf{H}^{\mathrm{T}}\boldsymbol{\xi}, \quad (4)$$

$$\begin{bmatrix} \mathbf{H}\mathbf{Q}_{\mathbf{ss}}\mathbf{H}^{\mathrm{T}} + \mathbf{R} & \mathbf{HX} \\ \mathbf{X}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix}. \quad (5)$$

When the forward problem is nonlinear or nonlinearities are introduced working in a transformed estimation space (in this work a logarithm transformation was adopted) the relation between the parameters and the observations, $\mathbf{h}(\mathbf{s})$, can be successively linearized following the quasi-linear extension of the BG approach (Kitanidis 1995). Indicating with $\mathbf{s}_k$ the current estimated parameter vector, $\mathbf{h}(\mathbf{s})$ is approximated with $\mathbf{h}(\mathbf{s}_k)+\widetilde{\mathbf{H}}(\mathbf{s}-\mathbf{s}_k)$ and the measurements are corrected as $\mathbf{y}_k=\mathbf{y}-\mathbf{h}(\mathbf{s}_k)+\widetilde{\mathbf{H}}\mathbf{s}_k$. The matrix $\widetilde{\mathbf{H}}$ represents the sensitivity of each observation with respect to each unknown parameters (Jacobian matrix) and is evaluated at each iteration in the linearization process. The quasi-linear extension of the BG approach presumes that the nonlinearities are not too extreme so that the linear approximation is effectively applicable in the proximity of the best estimate and the posterior is then almost Gaussian (Kitanidis and Lee 2014). In some cases, especially in strongly nonlinear optimization tasks, overshooting and numerical instabilities can make the algorithm convergence difficult (Nowak and Cirpka 2014). In this work, as suggested by Zanini and Kitanidis (2009) an optimization procedure, "line search", was used to drive the solution and address its oscillations.

The right selection of the structural parameters is important to obtain a proper estimation of the spatial fields; this represents the second step in the estimation process. In the perspective of the "empirical Bayes", the structural parameters $\boldsymbol{\theta}$ (and, optionally, the epistemic error variance), which regulate the level of fit between modeled data and observations, are inferred from the data. In this work, a Bayesian adaptation (Fienen et al. 2013) of the Restricted Maximum Likelihood method of Kitanidis (1995) was used. The nonlinearity of the estimation problem requires to evaluate structural parameters and unknown parameters ($\mathbf{s}$) iteratively (Fienen et al. 2009).

Once the optimal solution has been achieved, the linearized uncertainty of the unknowns (i.e. the uncertainty provided after the linearization process of the nonlinear forward problem) can be evaluated in terms of the posterior covariance matrix of the parameter estimates (Nowak and Cirpka 2004). The posterior uncertainty (in the nonlinear case) is only approximated (Cardiff and Barrash 2011) but it has an acceptable accuracy when compared with more computational expensive methods such as the Markov chain Monte Carlo approach (Liu et al. 2012).

The free software package bgaPEST (Fienen et al. 2013), developed according to the PEST software concept (Doherty 2010), which implements the Bayesian geostatistical inversion above described has been adopted in this work. The inverse code is independent from, and can be linked to, the majority of forward models.

# 3 Application

## 3.1 Description of the test site

The well field is situated inside the Campus of Science and Technology of the University of Parma (see Fig. 1); it is located in an alluvial deposit between the Parma and Baganza rivers (with a mean elevation of 82 m asl). Starting from literature works (Regione Emilia-Romagna 1998), the geology of the area has been refined according to a stratigraphic survey up to the depth of 30 m. The first 5 m (from the ground surface) are mainly constituted by a clay layer; from 5 m to 22 m there is a succession of gravel and clay that constitutes the shallow aquifer; from 22 m to 23 m there is a clay layer (aquitard) which separates the two aquifers; from 23 m to 27 m there are deposits of gravel which constitute the investigated aquifer (in the following named target aquifer); and from 27 m to 30 m there is a clay layer which is an impervious boundary. According to the regional geology (Regione Emilia-Romagna 1998), the two aquifers, several hundred meters upstream (South-West direction), are connected due to the thinning of the aquitard (see Fig. 2 for a sketch). For this reason, during the recharge period (autumn/winter) the two aquifers present the same hydraulic head levels, while during the summer (irrigation period) the shallow aquifer has higher drawdowns with respect to the target one. Close to the investigated site, there is a small river (Cinghio River) that flows from South West to North East (see Fig. 1); from the 1980s the river is regulated in order to secure, from flooding, the South part of the city of Parma. Its riverbed is fully contained in the shallow clay layer; so it does not present any interaction with the underneath aquifers.

*Here Figure 1*

*Here Figure 2*

The test site was completed in 2014 and consists of 8 wells: 4 are screened at the top of the shallow aquifer, between 8 m and 11 m below the ground surface (named with S in Fig. 2), and 4 are screened in the target aquifer, between 23 m and 27 m below the ground surface (named with P in Fig. 2). The wells have a diameter of 10 cm and they are surrounded by a filter pack of uniform gravel. Figure 3 shows the location of the wells and their mutual distances.

The well field has been equipped with several instruments, which allow performing traditional and advanced pumping tests: a submersible pump, five pressure and temperature probes with integrated data logging, a magnetic flow meter, a manual water level meter, a volumetric pump and a tank. Four pressure probes are installed in all the P wells to collect the data during the pumping tests in the investigated aquifer; the other probe (installed in one of the S wells) quantifies the influence of the withdrawal on the shallow aquifer. Several preliminary tests were carried out pumping from different locations in order to understand the response of each well. During these tests, it was noticed a clogging of the well P2; for this reason, it was not used in the further investigations.

The data collected by the Regional Agency for Environmental Protection (ARPA Emilia Romagna), available online at http://webbook.arpae.it/acque/acque-sotterranee/index.html (November 2016), were used to create water level contour maps (e.g. the blue lines in Fig. 3) to estimate the flow direction and the mean hydraulic gradient.

*Here Figure 3*

## 3.2 Pumping tests

### 3.2.1 Constant rate pumping tests

Several traditional pumping tests (Theis 1935), with different constant flow rates and different pumping locations, were carried out in order to evaluate the aquifer response to a stress. The results of these tests showed that the water level inside the monitoring wells decreases rapidly and then it stabilizes to a steady value according to the behavior of a leaky aquifer. Consequently, these traditional tests were evaluated through the Hantush (1956) method. The estimated average hydraulic conductivity was $5 \cdot 10^{-5}$ m/s and the storativity $1.5 \cdot 10^{-4}$; in addition, following the approach proposed by Veling and Maas (2010), the hydraulic conductivity of the aquitard was estimated in $4 \cdot 10^{-7}$ m/s. The shallow aquifer was not affected by the pumping from the target aquifer.

### 3.2.2 Oscillatory pumping tests

A second set of pumping tests have been performed following the approach proposed by Cardiff et al. (2013). The method (named oscillatory hydraulic tomography, OHT) is an improvement of the hydraulic tomography (Yeh and Liu 2000; Fienen et al. 2008; Cardiff et al. 2011, 2012) and consists in characterizing the aquifer heterogeneities injecting and extracting water, according to a known periodic function, from a well and observing the response, in terms of frequency and amplitude, in the available monitoring wells. As noted by Cardiff et al. (2013), that used sinusoidal signals, this approach has several advantages relative to traditional characterization tests: 1. the water balance between injection and extraction is zero; 2. in case of polluted sites, there is not the necessity to treat significant quantities of contaminated water and plumes have less movement; 3. the use of signals with known frequencies allows to distinguish the responses at monitoring locations from the other hydrological processes or noises.

Rabinovich et al. (2015) showed that using a sinusoidal wave in a real aquifer is difficult using commercial equipment. For this reason, in the present work, the injection/extraction process was simplified using a square wave with a constant flow rate. This assumption has allowed to greatly simplify the field equipment in: two pumps (one submersible and one centrifugal) and a tank (see Fig. 4). The equipment was completed by a flow meter, some regulation valves, the pressure sensors and a level probe to record the water level inside the tank.

*Here Figure 4*

When the test starts, the tank is full of water; the centrifugal pump injects this water into the aquifer, with a constant flow rate, for a fixed time interval (half period). After this period, the injection ceases and the submersible pump extracts water from the aquifer at the same flow rate of the injection and for the same time interval; the extracted water is temporarily stored in the tank. This sequence of injection/extraction is repeated several times. Figure 5 shows the data collected at two monitoring wells (P1 and P3) using two different square waves (half period equal to 5 min and 10 min) with the same extracting/injecting flow rate of 0.24 l/s (from well P4). All the data are accessible by contacting the corresponding author.

*Here Figure 5*

### 3.3 Numerical flow model

The groundwater flow processes of the test site were modelled numerically using MODFLOW 2005 (Harbaugh et al. 2005). The main objective of the model was the estimation of the hydraulic parameters (hydraulic conductivity and specific storage) at the well field. For this reason, the site was discretized with a small finite difference grid (0.2×0.2 $m^2$) in the area between the wells (black square in the inset of Fig. 3); the grid size increases up to 16 m × 18 m at the borders, with a ratio of 1.2 between adjacent cells. The model represents with three layers: the shallow aquifer, the aquitard and the target aquifer. Each layer has a total of 105×93 cells, covering an area of 530×350 $m^2$. The top of the surface layer was modeled according to a regional topographic survey while the top and bottom of the other two layers follows the local stratigraphy. Constant head boundary conditions (according to the regional hydraulic head contours) were imposed at South-West and North-East boundaries. No flow boundary conditions were assigned at the other borders. As showed in Fig. 3, the model domain extends far from the well field to avoid influence of the boundary conditions on the solutions.

The final pumping test configuration consisted in pumping from well P4 and observing the drawdowns in P1 and P3. In MODFLOW 2005, the pumping well P4 (placed in the third layer) was simulated using the well package, while the wells P1 and P3 were considered observation points.

An adjoint-state formulation of MODFLOW 2005 (Clemo 2007) was used to efficiently calculate the

Jacobian matrix required by the inverse procedure. By means of the adjoint-state approach rather than calculating the sensitivity matrix using a minimum of one forward run per parameter (plus one base run), the number of forward runs is associated to the number of observations. The suggestions of D'Oria and Fienen (2012) were also taken into account to speed up the computation.

### 3.4 Observations and parameters

Besides boundary conditions, the numerical model needs the definition of the hydraulic parameters of the three layers used to represent the study area stratigraphy. In particular, the hydraulic conductivity must be specified for each layer, the specific yield for the shallow unconfined aquifer and the specific storage for the aquitard and the target aquifer. To estimate these unknown parameters, the head data collected during the oscillatory pumping test with square wave and half period equal to 10 min were considered. Two observation points (P1 and P3), two stimulation periods and one observation per minute were used. In summary, a total of 40 heads for each observation points represent the observed data. The observations are expressed in terms of water level differences with respect to the initial values (drawdowns or positive increments).

By means of PEST (Doherty 2010), a preliminary estimation of the unknown parameters was performed using homogeneous layers with the aim at finding initial values for the parameters. Then, in applying the Bayesian Geostatistical approach, only the focused area around the wells (inset of Fig. 3) and inside the target aquifer was considered as highly parameterized allowing at each cell to have a different parameter value; the outer zone and the other two layers were kept lumped. As a result, 1225 hydraulic conductivity values (35x35 cells) and 1225 specific storage values (35x35 cells) were estimated for the focused area together with the values of three homogenous zones: the hydraulic conductivity and specific storage of the outer zone of the target aquifer and the hydraulic conductivity of the aquitard; in total the number of unknown parameters was 2453. The other parameters were kept constant and equal to the values estimated through the preliminary calibration.

## 4 Results

Using only the data collected by the oscillatory pumping test with square wave and half period equal to 10 min, the best estimates, $\hat{\mathbf{s}}$ , of both the hydraulic conductivity and specific storage fields of the focused area were evaluated (Fig. 6); mean values of all the estimated parameters are reported in Table 1. With reference to Fig. 6, the hydraulic conductivity slightly increases from West to East and the mean value of

Table 1 is consistent with the alluvial nature of the aquifer and the value estimated by means of the traditional aquifer tests reported at Sect. 3.2.1. In natural logarithm scale the hydraulic conductivity field presents a mean equal to -10.07 and a variance equal to 0.31 (i.e. -4.37 and 0.06 in $\log_{10}$ scale, respectively). The specific storage field is generally homogeneous with variations smaller than half an order of magnitude; also in this case the mean value reported in Table 1 agrees with the one derived from traditional tests. The estimated specific storage field presents a mean of -9.63 and variance equal to 0.005 in natural logarithm scale (i.e. -4.18 and $9.95 \cdot 10^{-4}$ in $\log_{10}$ scale, respectively). Both the variances of the estimated hydraulic conductivity and specific storage fields are small and are indicative of an almost homogeneous area.

*Here Figure 6*

*Here Table 1*

Figure 7 reports the comparison between the observed and modeled values at the monitoring wells P1 and P3; Fig. 8 shows the observed values versus the modeled ones together with some performance metrics: mean error, root mean square error and the normalized one and the coefficient of determination based on the 1:1 line. Overall, there is a good agreement between observations and model outputs.

*Here Figure 7*

*Here Figure 8*.

The starting values of structural parameters, for both the hydraulic conductivity and the specific storage fields, were set such that the first solution was flat. Complexity is introduced during the optimization process only if supported by the data (D'Oria and Tanda 2012). After three iterations of the structural parameters estimation step, they almost became stable. Table 2 reports the initial values and the ones estimated at the end of the procedure (five iterations).

The linearized uncertainties of the estimates have been evaluated on the basis of the posterior covariance matrix. Figure 9 shows the logarithm of the posterior standard deviation respectively for the hydraulic conductivity and the specific storage fields. With reference to the hydraulic conductivity field, it is clear that the area with the smallest variance is between the pumping wells and the observation points. This is due to the higher amount of information provided during the stimulation of the aquifer through the oscillatory pumping tests. Meanwhile, considering the specific storage field, the posterior standard deviation is almost constant and relatively small.

*Here Figure 9*

Once estimated the hydraulic parameters, the results were validated by means of the head data collected during the oscillatory pumping test with a square wave and half period equal to 5 min. Four periods and a total of 80 heads for each of the two observation points (P1 and P3) were considered. Fig. 10 shows the comparison between observations and model outputs; in this case the mean error is -3.8 mm, the RMSE 19.9 mm, the nRMSE 3.8% and the coefficient of determination is 0.9896. These values, even if are larger than the one obtained in the calibration procedure, highlight the goodness of the parameter estimation.

*Here Figure 10*

# 5 Discussion and Conclusions

The objective of the present work was the estimation of the hydraulic parameters of a leaky confined aquifer by means of a Bayesian Geostatistical approach using the bgaPEST inverse code (Fienen et al. 2013). The BG approach estimated the hydraulic conductivity and the specific storage fields with a large number of parameters using the observations taken at two monitoring wells at several times. In order to collect the field data, the aquifer has been stimulated through oscillatory pumping tests using commercial equipment. The results of the inversion are consistent with the alluvial nature of the investigated aquifer and with the preliminary traditional pumping test carried out at the site.

Despite the high number of estimable parameters, the hydraulic conductivity and specific storage fields obtained by means of the Bayesian Geostatistical inversion are smooth. In fact, it noteworthy that, in the contest of a highly parameterized inversion, the excess of flexibility can introduce the possibility of over-fitting (Hill 2006). The use of prior information can control this issue that can be also mitigated by smoothness constraints. The BG approach and the Bayesian adaptation of restricted maximum likelihood, estimating the structural parameters of the covariance matrix and/or the epistemic error variance, provides a best estimate which represents the smoothest solution consistent with the available observations (Fienen et al. 2013). Since complexity is introduced only if supported by the data, the solution is probably smoother than reality (Fienen et al. 2009); however conditional realizations, that are samples from the posterior distribution (Kitanidis and Lee 2014; Kitanidis 1995; Zanini and Kitanidis 2009), can be obtained in addition to the best estimate. These equally likely realizations follow the spatial structure imposed by the covariance matrix and reproduce the observations within the level of fit imposed or estimated through the epistemic uncertainty. Conditional realizations can characterize the small scale variability of the solution and their mean, if the number of generations is large enough, would reproduce the best estimate ( $\hat{\mathbf{s}}$ in Eq. (4), Michalak et al. 2004).

## Acknowledgments

The authors are grateful to the editor and the anonymous reviewers for their valuable comments and suggestions on this work.

## References

Alberti L, Lombi S, Zanini A, (2011) Identifying sources of chlorinated aliphatic hydrocarbons in a residential area in Italy using the integral pumping test method, Hydrogeol J, 19(6):1253-1267.doi:10.1007/s10040-011-0742-1

Butera I, Tanda MG, Zanini A (2013) Simultaneous identification of the pollutant release history and the source location in groundwater by means of a geostatistical approach. Stochastic Environ Res Risk Assess, 27(5), 1269–1280.doi:10.1007/s00477-012-0662-1

Cardiff M, Bakhos T, Kitanidis PK, Barrash W (2013) Aquifer heterogeneity characterization with oscillatory pumping: Sensitivity analysis and imaging potential. Water Resour Res 49:5395–5410.doi:10.1002/wrcr.20356

Cardiff M, Barrash W (2011) 3-D transient hydraulic tomography in unconfined aquifers with fast drainage response. Water Resour Res 47:W12518.doi:10.1029/2010WR010367

Cardiff M, Barrash W, Kitanidis PK (2012) A field proof-of-concept of aquifer imaging using 3-D transient hydraulic tomography with modular, temporarily emplaced equipment. Water Resour Res 48: W05531.doi:10.1029/2011WR011704

Carrera J, Alcolea A, Medina A, Hidalgo J, Slootenet LJ (2005) Inverse problem in hydrogeology. Hydrogeol J 13:206.doi:10.1007/s10040-004-0404-7

Chen Y, Oliver DS (2012) Ensemble Randomized Maximum Likelihood Method as an Iterative Ensemble Smoother. Math Geosci 44:1-26.doi:10.1007/s11004-011-9376-z

Clemo T (2007) MODFLOW-2005 ground water model - User guide to the adjoint state based sensitivity process (ADJ). Center for the Geophysical Investigation of the Shallow Subsurface, Boise State University, Boise, Idaho

Cupola F, Tanda MG, Zanini A (2015) Contaminant release history identification in 2-D heterogeneous aquifers through a minimum relative entropy approach, SpringerPlus 4:656.doi:10.1186/s40064-015-1465-x

Doherty JE (2010) PEST, Model-independent parameter estimation - User manual (5th ed., with slight additions). Watermark Numerical Computing, Brisbane, Australia

Doherty JE, Hunt RJ (2010) Approaches to highly parameterized inversion - A guide to using PEST for groundwater-model calibration: U.S. Geological Survey Scientific Investigations Report 5169

D'Oria M, Fienen MN (2012) MODFLOW-Style Parameters in Underdetermined Parameter Estimation. Ground Water 50:149–153.doi:10.1111/j.1745-6584.2011.00803.x

D’Oria M, Mignosa P, Tanda MG (2014) Bayesian estimation of inflow hydrographs in ungauged sites of multiple reach systems. Adv Water Resour 63:143-151.doi:10.1016/j.advwatres.2013.11.007

D’Oria M, Mignosa P, Tanda MG (2015) An inverse method to estimate the flow through a levee breach. Adv Water Resour 82:166-175.doi: 10.1016/j.advwatres.2015.05.002

D’Oria M, Tanda MG (2012) Reverse flow routing in open channels: A Bayesian Geostatistical Approach, J Hydrol 460:130-135.doi:10.1016/j.jhydrol.2012.06.055

Emerick AA, Reynolds AC (2013) Ensemble smoother with multiple data assimilation. Comput Geosci 55:3-15.doi:10.1016/j.cageo.2012.03.011

Evensen G (2009) Data assimilation: the ensemble Kalman filter 2nd edn. Springer, Berlin.doi: 10.1007/978-3-642-03711-5

Fienen MN, Clemo TM, Kitanidis PK (2008) An interactive Bayesian geostatistical inverse protocol for hydraulic tomography. Water Resour Res 44:W00B01.doi:10.1029/2007WR006730

Fienen M, Hunt R, Krabbenhoft D, Clemo T (2009) Obtaining parsimonious hydraulic conductivity fields using head and transport observations: A Bayesian geostatistical parameter estimation approach. Water Resour Res 45:W08405.doi:10.1029/2008wr007431

Fienen MN, D’Oria M, Doherty JE, Hunt RJ (2013) Approaches in highly parameterized inversion: bgaPEST, a Bayesian geostatistical approach implementation with PEST-Documentation and instructions. In: U.S. Geological Survey Techniques and Methods 7-C9

Grana D, Fjeldstad T, Omre H. (2017) Bayesian Gaussian Mixture Linear Inversion for Geophysical Inverse Problems. Math Geosci 49:493.doi:10.1007/s11004-016-9671-9

Hansen TM, Journel AG, Tarantola A, Mosegaard K (2006) Linear inverse Gaussian theory and geostatistics. Geophysics 71(6), R101-R111.doi:10.1190/1.2345195

Hantush MS (1956) Analysis of Data from Pumping Tests in Leaky Aquifers. Transactions, American Geophysical Union 37(6):702-714

Harbaugh AW (2005) MODFLOW-2005, The U.S. Geological Survey modular ground-water model - the

Ground-Water Flow Process. In: U.S. Geological Survey Techniques and Methods 6-A16

Hendricks Franssen HJ, Alcolea A, Riva M, Bakr M, van der Wiel N, Stauffer F, Guadagnini A (2009) A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments. Adv Water Resour 32(6):851–872.doi.org/10.1016/j.advwatres.2009.02.011

Hill MC (2006) Using models to manage systems subject to sustainability indicators. In: Webb B, Hirata R, Kruse E, Vrba J (ed) Sustainability of Groundwater Resources and its Indicators, IAHS Publication 302. IAHS Press, Wallingford, UK.

Hoeksema RJ, Kitanidis PK (1984) An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling. Water Resour Res 20(7):1003–1020.doi:10.1029/WR020i007p01003

Kitanidis PK, Vomvoris EG (1983) A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. Water Resour Res 19(3):677–690.doi:10.1029/WR019i003p00677

Kitanidis PK (1993) Generalized covariance functions in estimation. Math Geol 25(5):525-540.doi:10.1007/BF00890244

Kitanidis PK (1995) Quasi-linear geostatistical theory for inversing. Water Resour Res 31(10):2411–2419.doi:10.1029/95WR01945

Kitanidis PK (1997) Introduction to Geostatistics, Applications in Hydrogeology. Cambridge University Press

Kitanidis PK, Lee J (2014) Principal Component Geostatistical Approach for large-dimensional inverse problems, Water Resour Res 50:5428-5443.doi:10.1002/2013WR014630

Leonhardt G, D'Oria M, Kleidorfer M, Rauch W. (2014) Estimating inflow to a combined sewer overflow structure with storage tank in real time: evaluation of different approaches. Water Sci Technol 70(7):1143-1151.doi:10.2166/wst.2014.331

Li L, Srinivasan S, Zhou H, Gomez-Hernandez JJ (2015) Two-point or multiple point statistics? A comparison between the ensemble Kalman filtering and the ensemble pattern matching inverse methods. Adv Water Resour 86(B):297-310.doi:10.1016/j.advwatres.2015.05.014

Liu X, Lee J, Kitanidis PK, Parker J, Kim U (2012) Value of information as a context-specific measure of uncertainty in groundwater remediation. Water Resour Manage 26(6):1513-1535.doi:10.1007/s11269-011-9970-3

McLaughlin D, Townley LR (1996) A Reassessment of the Groundwater Inverse Problem. Water Resour Res 32(5):1131–1161.doi:10.1029/96WR00160

Michalak AM, Bruhwiler L, Tans PP (2004) A geostatistical approach to surface flux estimation of atmospheric trace gases. J of Geophys Res 109:D14.doi:10.1029/2003jd004422

Nowak W, Cirpka OA (2004) A modified Levenberg-Marquardt algorithm for quasi-linear geostatistical inversing. Adv Water Resour 27(7):737–750.doi:10.1016/j.advwatres.2004.03.004

Rabinovich A, Barrash W, Cardiff M, Hochstetler DL, Bakhos T, Dagan G, Kitanidis PK (2015) Frequency dependent hydraulic properties estimated from oscillatory pumping tests in an unconfined aquifer, J Hydrol 531(1):2-16.doi:10.1016/j.jhydrol.2015.08.021

Regione Emilia-Romagna, Servizio Geologico Sismico e dei Suoli, ENI – AGIP (1998) Riserve idriche sotterranee della Regione Emilia-Romagna, pp 119.

Rimstad K, Avseth P, Omre H (2012) Hierarchical Bayesian lithology/fluid prediction: a North Sea case study. Geophysics 77(2):B69–B85

Rubin Y, Chen X, Murakami H, Hahn M (2010) A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields, Water Resour Res 46(W10523).doi:10.1029/2009WR008799

Snodgrass MF, Kitanidis PK (1997) A geostatistical approach to contaminant source identification. Water Resour Res 33(4):537–546.doi:10.1029/96WR03753

Soares A, Nunes R, Azevedo L. (2017) Integration of Uncertain Data in Geostatistical Modelling. Math Geosci 49:253.doi:10.1007/s11004-016-9667-5

Tarantola A (2005) Inverse problem theory and methods for model parameter estimation. SIAM, Philadelphia

Theis CV (1935) The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using groundwater storage. Transactions of the American Geophysical Union 2:519-524

Yeh WWG (1986) Review of Parameter Identification Procedures in Groundwater Hydrology: The Inverse Problem. Water Resour Res 22(2):95–108.doi:10.1029/WR022i002p00095

Yeh TCJ, Liu S (2000) Hydraulic tomography: Development of a new aquifer test method. Water Resour Res 36(8):2095–2105.doi:10.1029/2000WR900114

Veling EJM, Maas C (2010) Hantush Well Function revisited. J Hydrol 393(3):381-388.doi: 10.1016/j.jhydrol.2010.08.033

Xu T, Gomez-Hernandez JJ (2016) Joint identification of contaminant source location, initial release time, and initial solute concentration in an aquifer via ensemble Kalman filtering. WaterResour. Res. 52: 6587–6595.doi:10.1002/2016WR019111

Zanini A, Kitanidis PK (2009) Geostatistical inversing for large-contrast transmissivity fields. Stoch Environ Res Risk Assess 23:565.doi:10.1007/s00477-008-0241-7

Zanini A, Woodbury AD (2016) Contaminant source reconstruction by empirical Bayes and Akaike's Bayesian Information Criterion. J Contam Hydrol, 185-186:74-86.doi:10.1016/j.jconhyd.2016.01.006.

Zanini A, Tanda MG, Woodbury AD (2017) Identification of transmissivity fields using a Bayesian strategy and perturbative approach. Adv Water Resour 108:69-82.doi:10.1016/j.advwatres.2017.07.022

Zhou H, Gómez-Hernández JJ, Liangping L (2014) Inverse methods in hydrogeology: Evolution and recent trends. Adv Water Resour 63:22–37.doi:10.1016/j.advwatres.2013.10.014

Zhou H, Li L, Hendricks Franssen, HJ, Gómez-Hernández JJ (2012) Pattern Recognition in a Bimodal Aquifer Using the Normal-Score Ensemble Kalman Filter. Math Geosci 44:169.doi:10.1007/s11004-011-9372

Zhou YQ, Lim D, Cupola F, Cardiff M (2016) Aquifer imaging with pressure waves - Evaluation of low-impact characterization through sandbox experiments. Water Resour Res 52:2141–2156.doi:10.1002/2015WR017751

Zimmerman DA, de Marsily G, Gotway CA, Marietta MG, Axness CL, Beauheim RL, Bras RL, Carrera J, Dagan G, Davies PB, Gallegos DP, Galli A, Gómez-Hernández J, Grindrod P, Gutjahr AL, Kitanidis PK, Lavenue AM, McLaughlin D, Neuman SP, RamaRao BS, Ravenne C, Rubin Y (1998) A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Water Resour Res 34(6):1373–1413.doi:10.1029/98WR00003

# Figure Captions

**Fig. 1** Study area and location of the well field

**Fig. 2** Sketch of the well field conceptual model; not to scale

**Fig. 3** Map of the area surrounding the well field with the finite difference grid of the numerical flow model; the blue lines are the regional hydraulic head contours; the values are in m asl

**Fig. 4** Sketch of the oscillatory pumping test and scheme of the field equipment, not to scale

**Fig. 5** Observed hydraulic heads at the monitoring points P1 and P3

**Fig. 6** Estimated hydraulic conductivity and specific storage fields in natural logarithm scale

**Fig. 7** Modeled and observed drawdown at P1 and P3, half period of 10 minutes.

**Fig. 8** Modeled versus observed drawdown and performance metrics: ME mean error, RMSE root mean square error and normalized one (nRMSE), $R^2$ coefficient of determination based on the 1:1 line

**Fig. 9** Standard deviation of the estimated fields in natural logarithm scale

**Fig. 10** Modeled and observed drawdown at P1 and P3, half period of 5 minutes

# Table Captions

**Table 1** Mean values of the estimated hydraulic parameters: HK hydraulic conductivity, SS specific storage and SY specific yield

**Table 2** Initial and estimated structural parameters for the hydraulic conductivity and the specific storage fields; $\sigma^2$ is the variance and $l$ is the correlation length

**Table 1**

| | Shallow aquifer (Layer 1) | Aquitard (Layer 2) | Target aquifer focused area (Layer 3) | Target aquifer outer zone (Layer 3) |
|---|---|---|---|---|
| HK (m/s) | $1.53 \cdot 10^{-5}$ | $2.71 \cdot 10^{-7}$ | $4.93 \cdot 10^{-5}$ | $2.18 \cdot 10^{-4}$ |
| SS (1/m) | - | $1.10 \cdot 10^{-5}$ | $6.56 \cdot 10^{-5}$ | $6.18 \cdot 10^{-7}$ |
| SY (-) | 0.25 | - | - | - |

**Table 2**

| | Hydraulic conductivity | | Specific storage | |
|---|---|---|---|---|
| | $\sigma^2$ | $l$ [m] | $\sigma^2$ | $l$ [m] |
| Initial | 0.0001 | 5.00 | 0.0001 | 5.00 |
| Estimated | 0.0507 | 1.41 | 0.0134 | 10.88 |

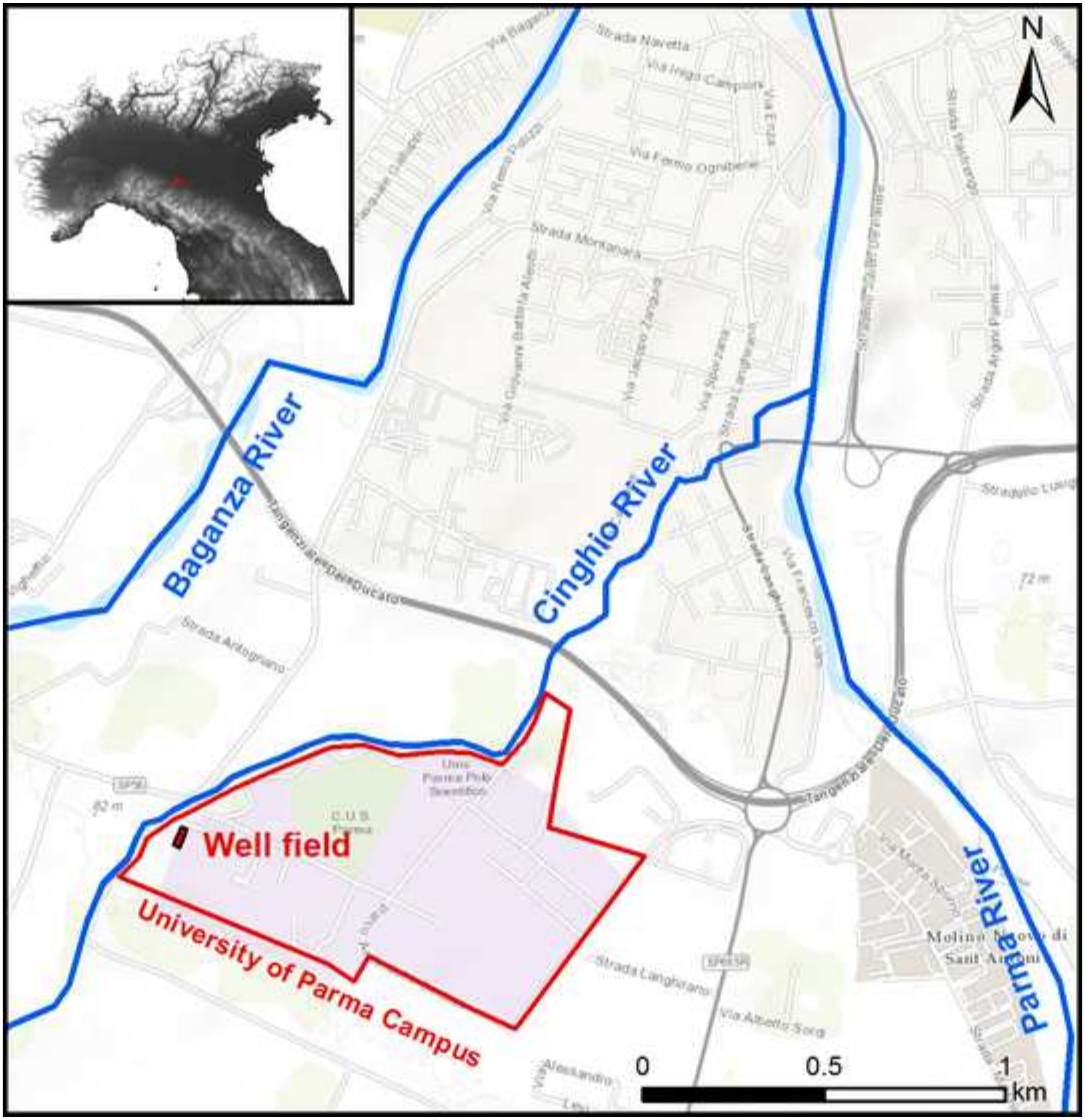

N
Baganza River
Cinghio River
Parma River
Well field
University of Parma Campus
0
0.5
1
km

Figure 2

Click here to download Figure Fig2.tif

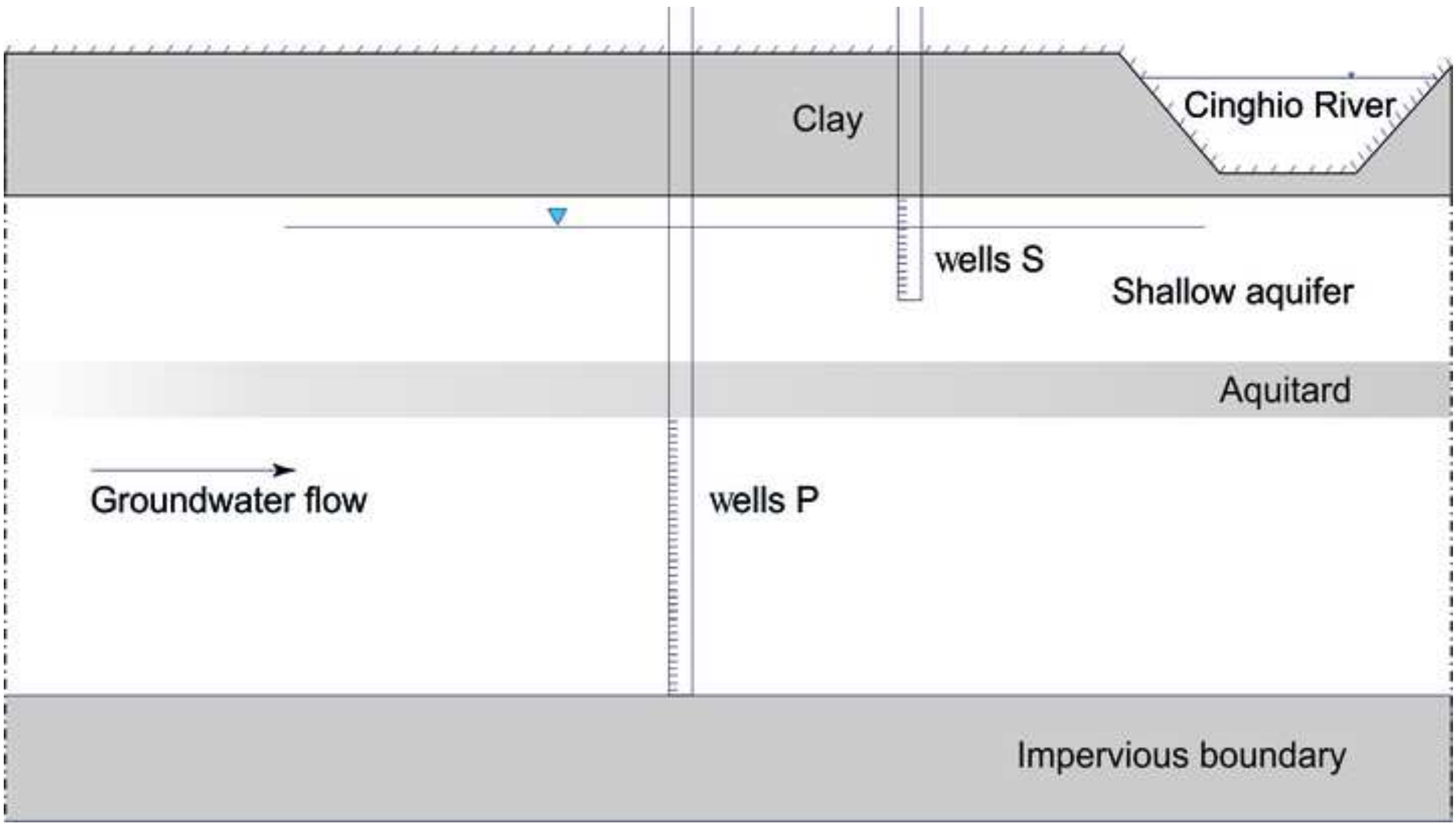
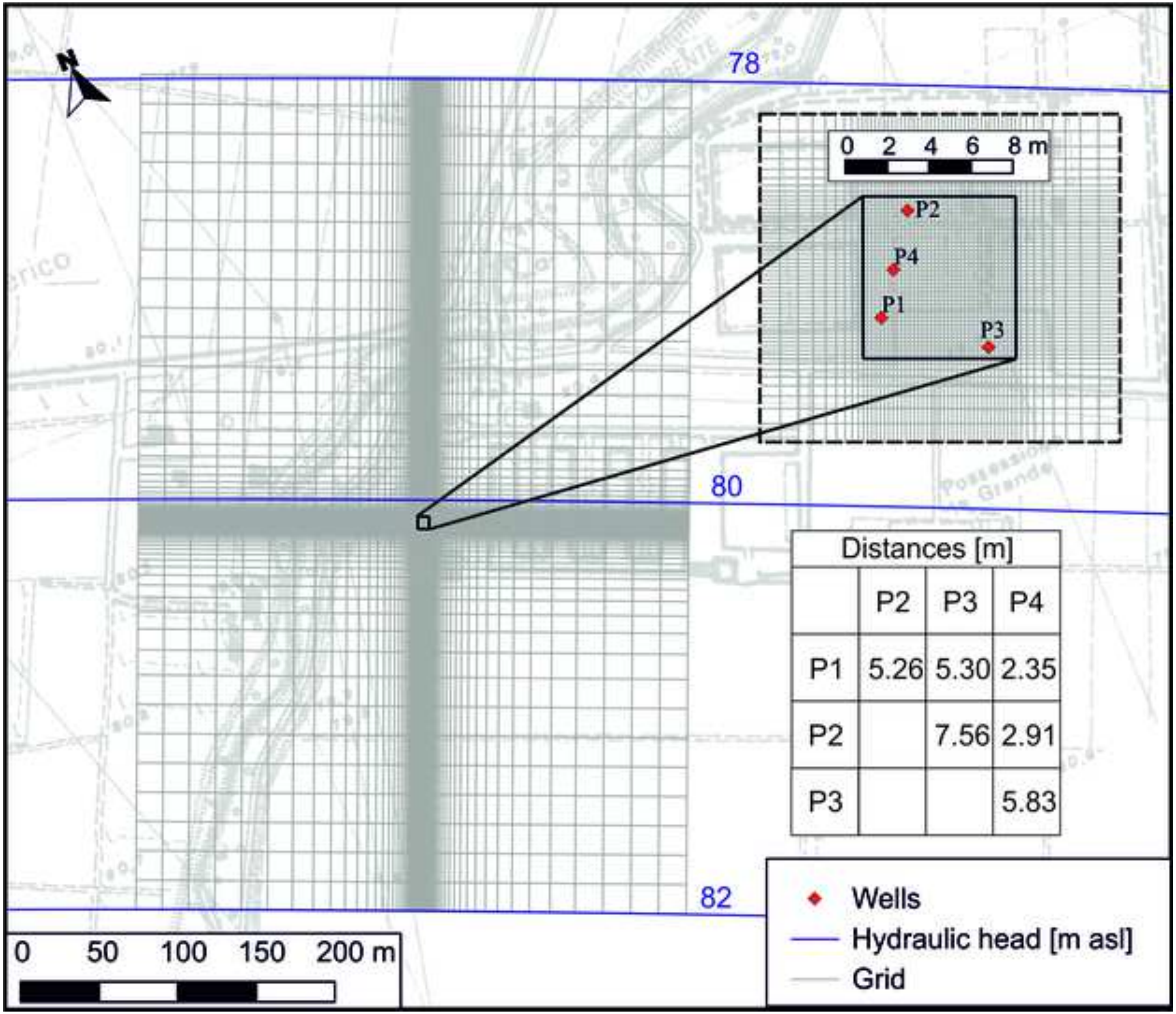


Figure 3






| Distances [m] | | | |
|---|---|---|---|
| | P2 | P3 | P4 |
| P1 | 5.26 | 5.30 | 2.35 |
| P2 | | 7.56 | 2.91 |
| P3 | | | 5.83 |

Figure 4

Click here to download Figure Fig4.tif

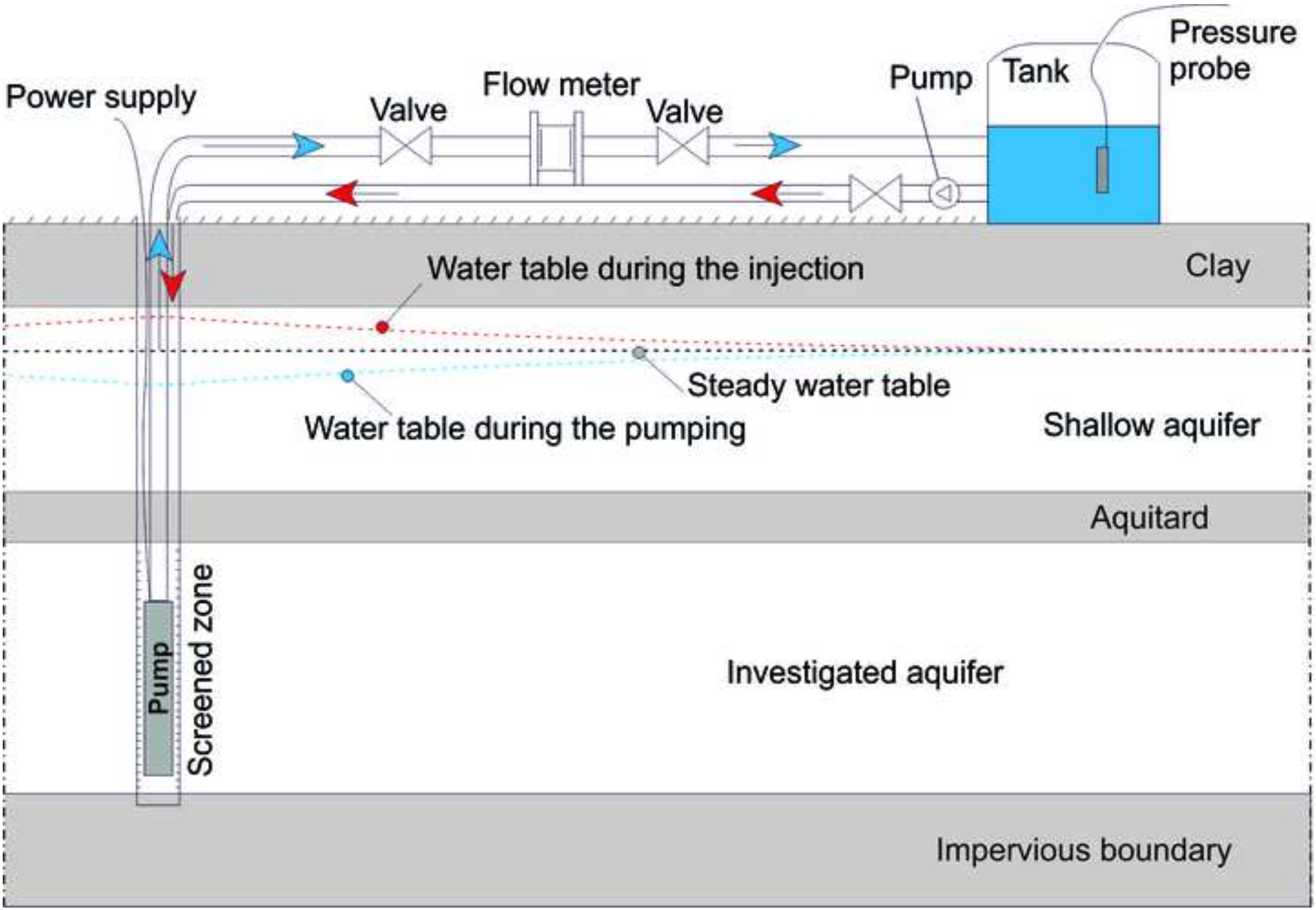


Figure 5

Click here to download Figure Fig5.tif

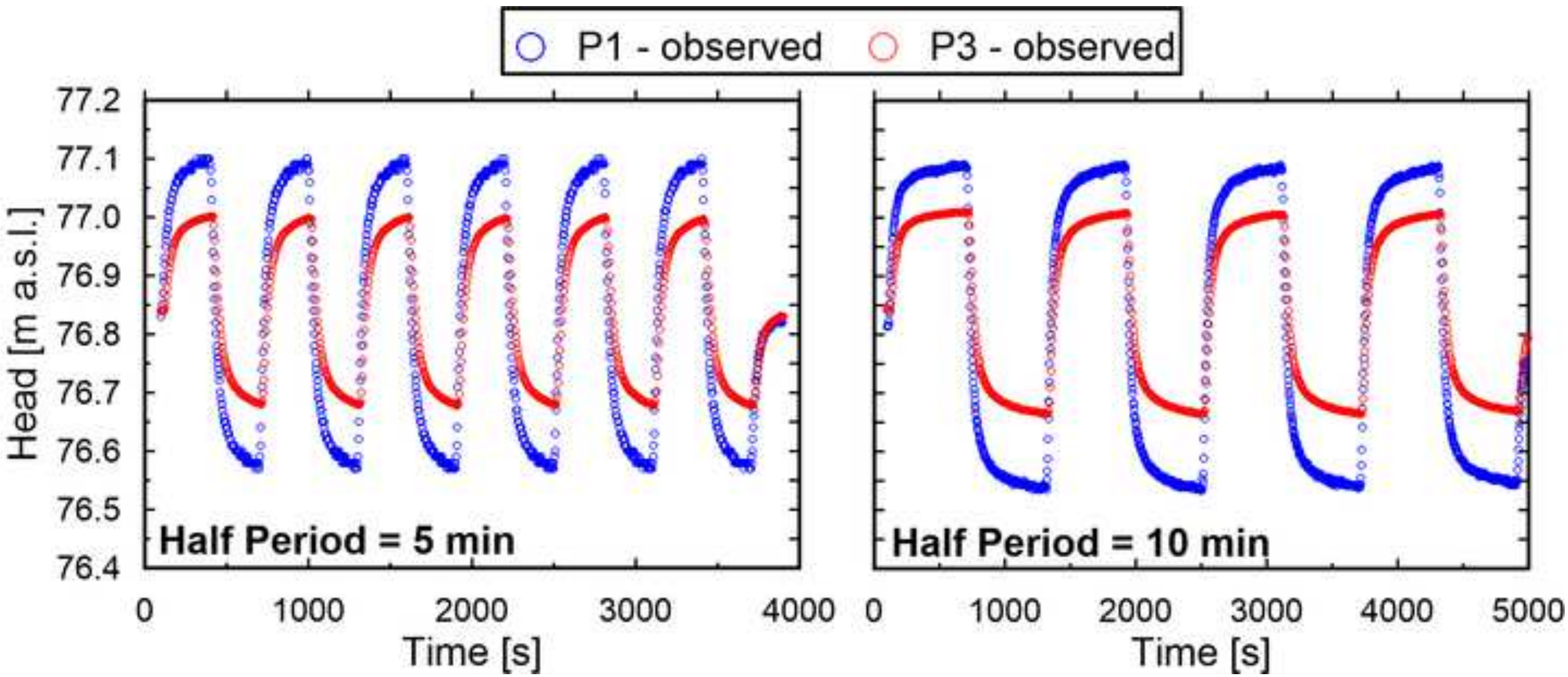


Figure 6

Click here to download Figure Fig6.tif

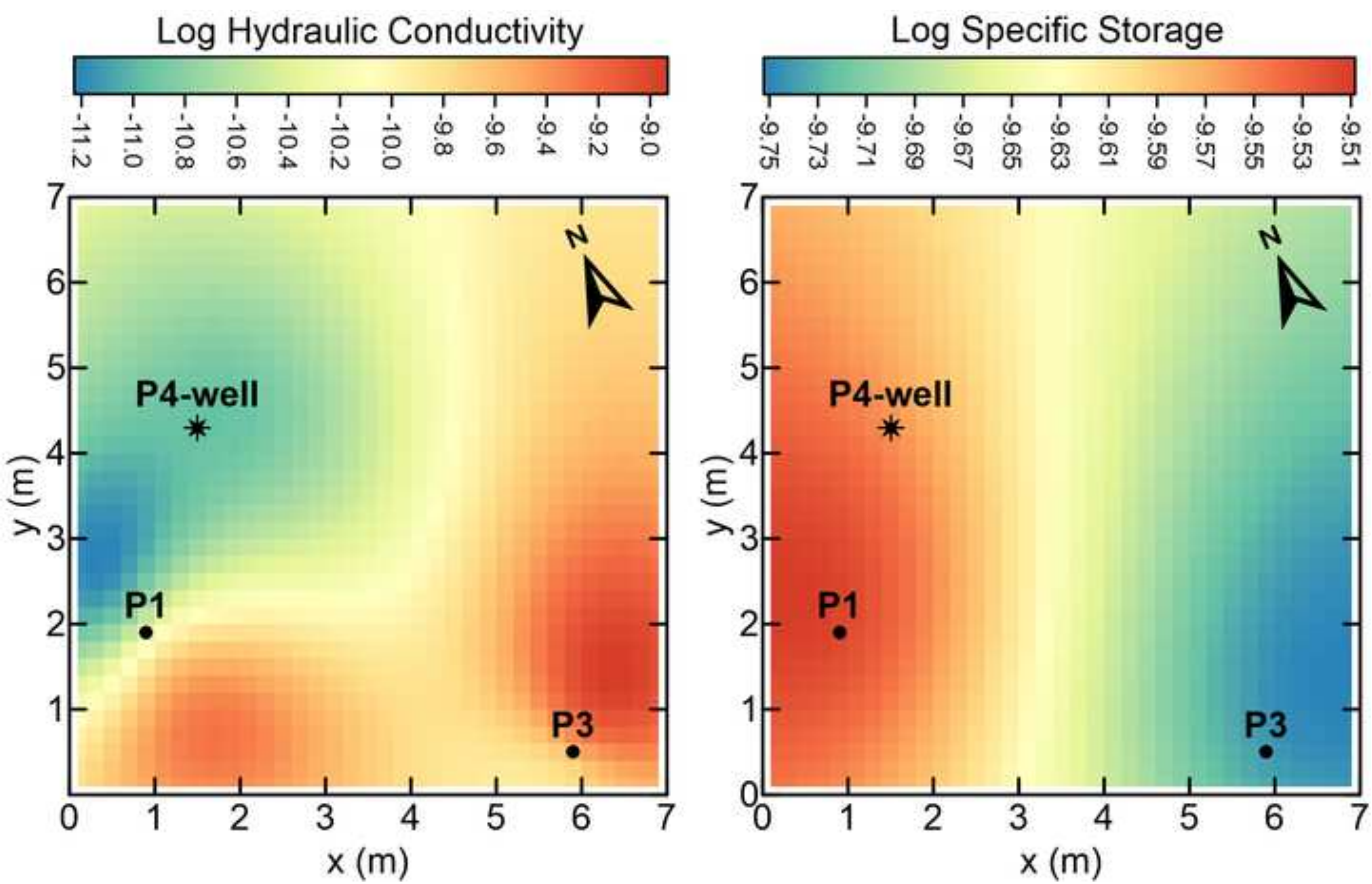
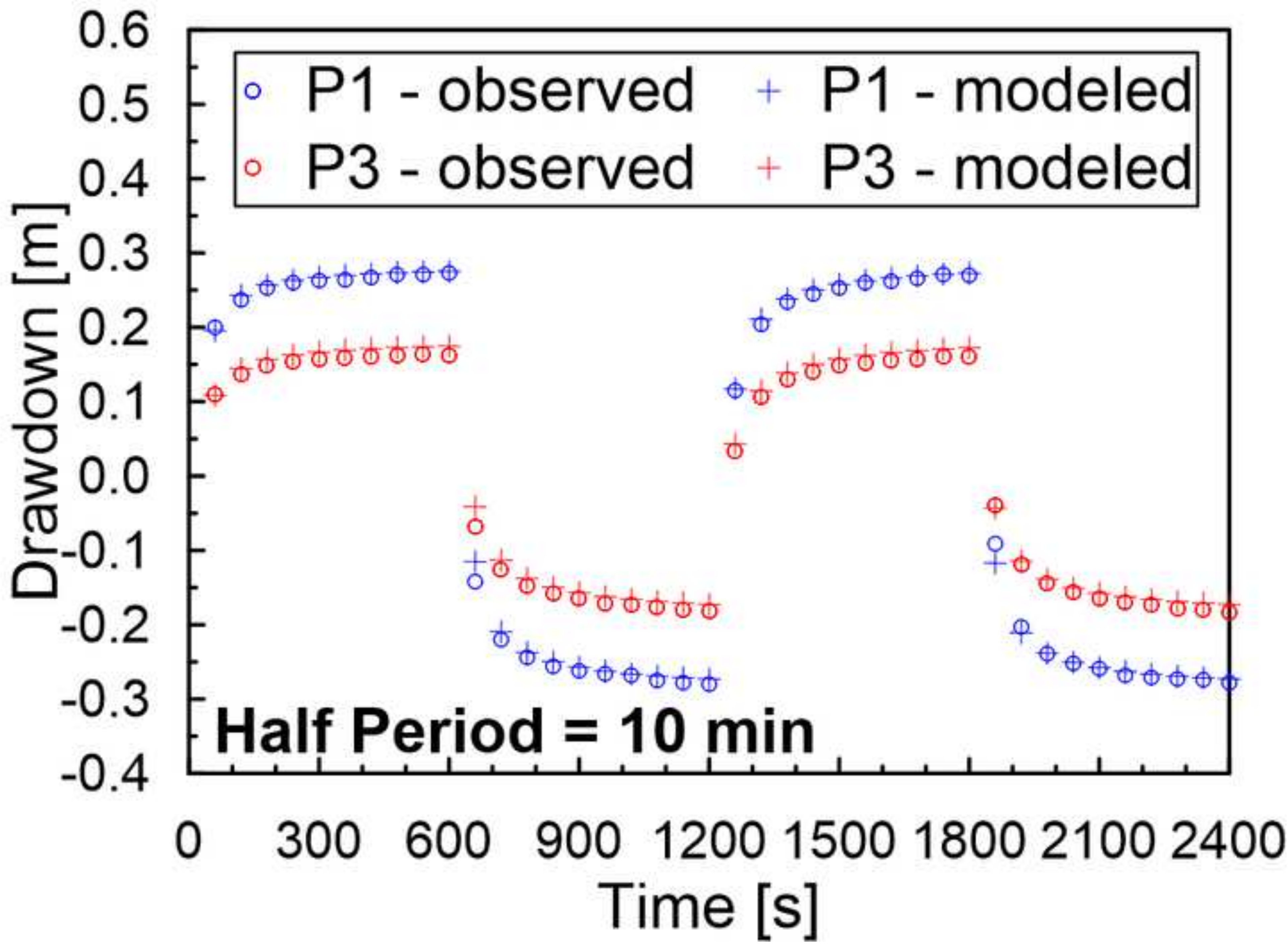


Figure 7





P1 - observed
P1 - modeled
P3 - observed
P3 - modeled
Drawdown [m]
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
Half Period = 10 min
0
300
600
900
1200
1500
1800
2100
2400
Time [s]

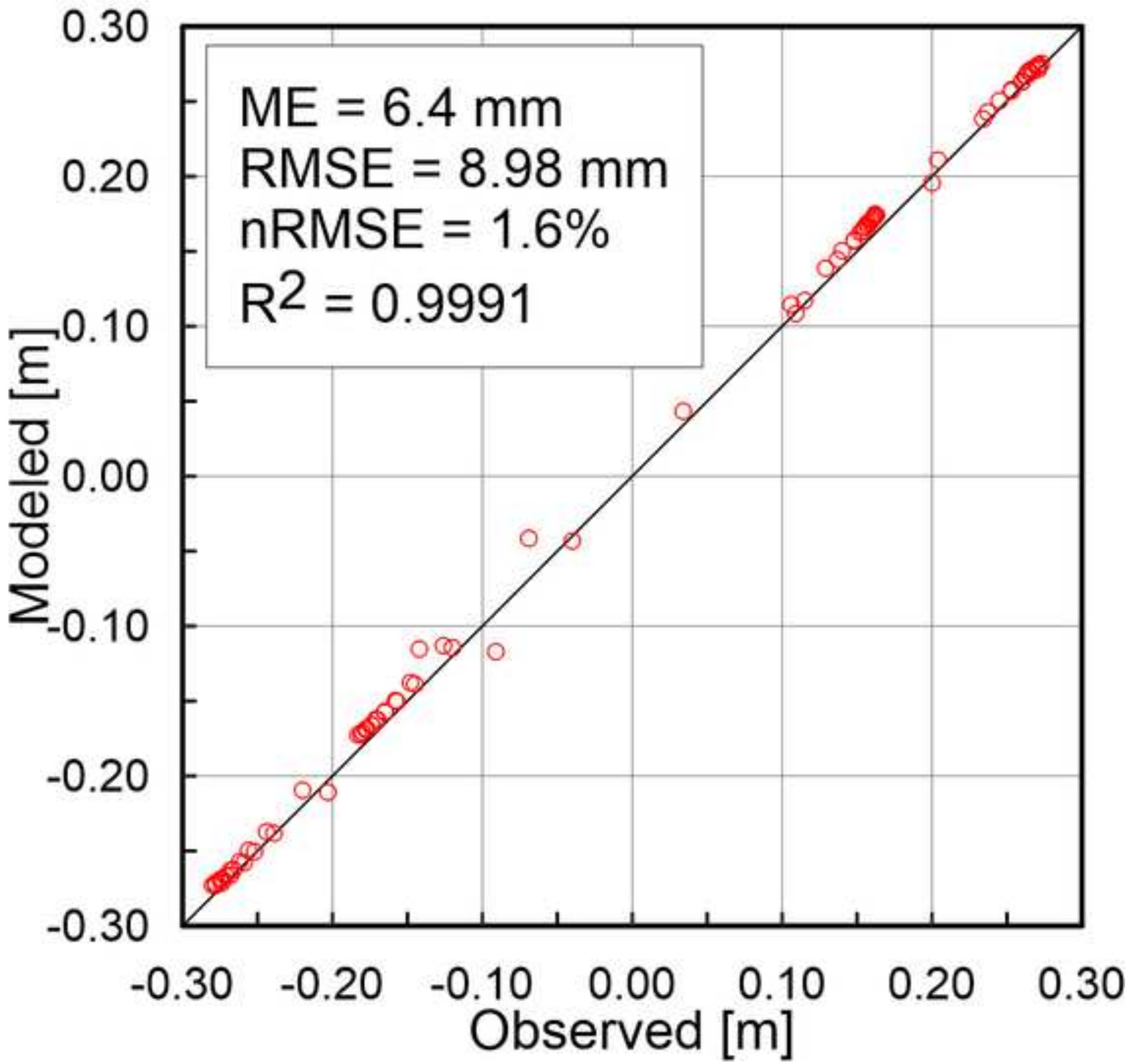

ME = 6.4 mm
RMSE = 8.98 mm
nRMSE = 1.6%
R2 = 0.9991
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
Modeled [m]
-0.30
-0.20
-0.10
0.00
0.10
0.20
0.30
Observed [m]

Figure 9

Click here to download Figure Fig9.tif

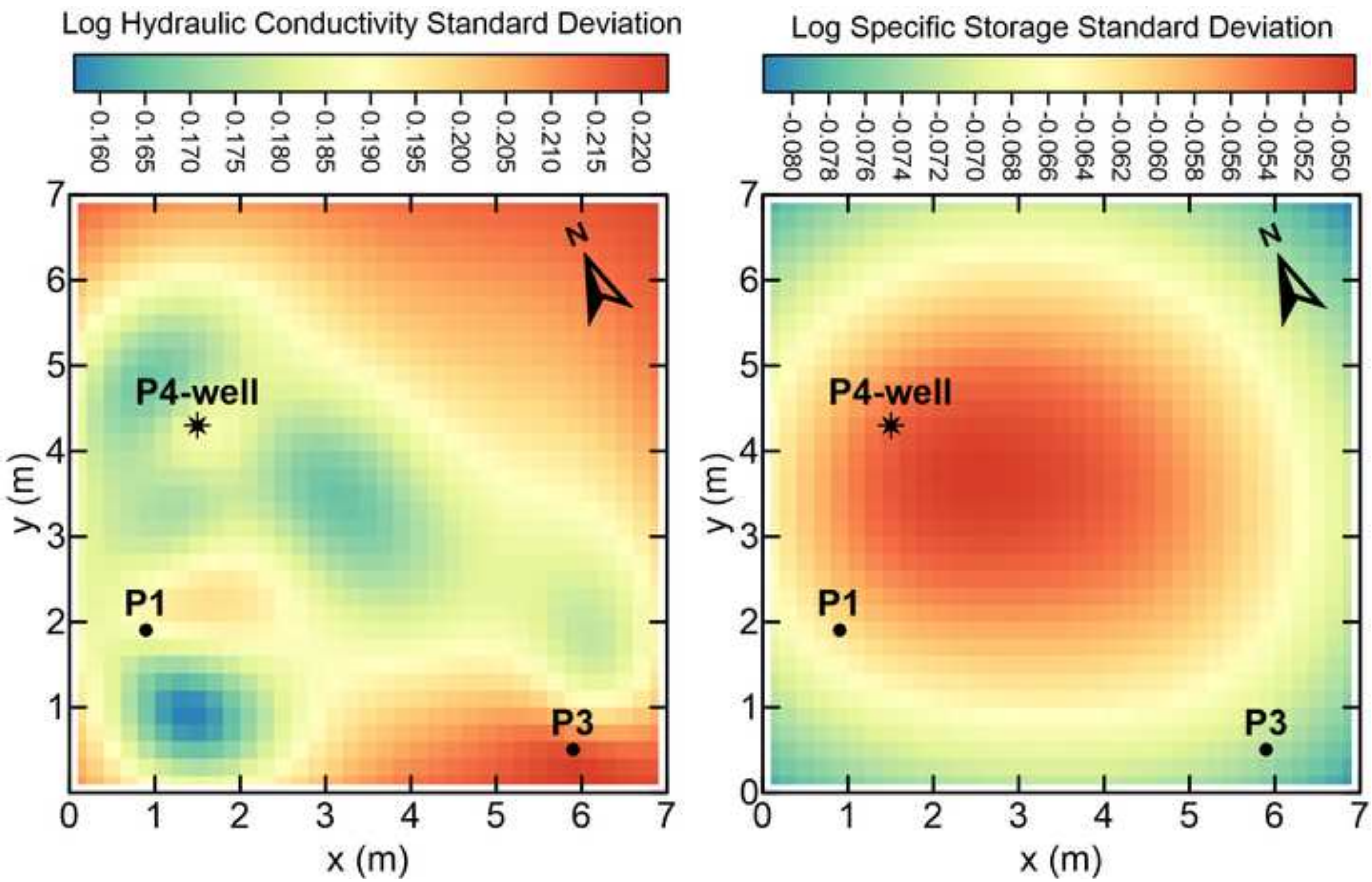
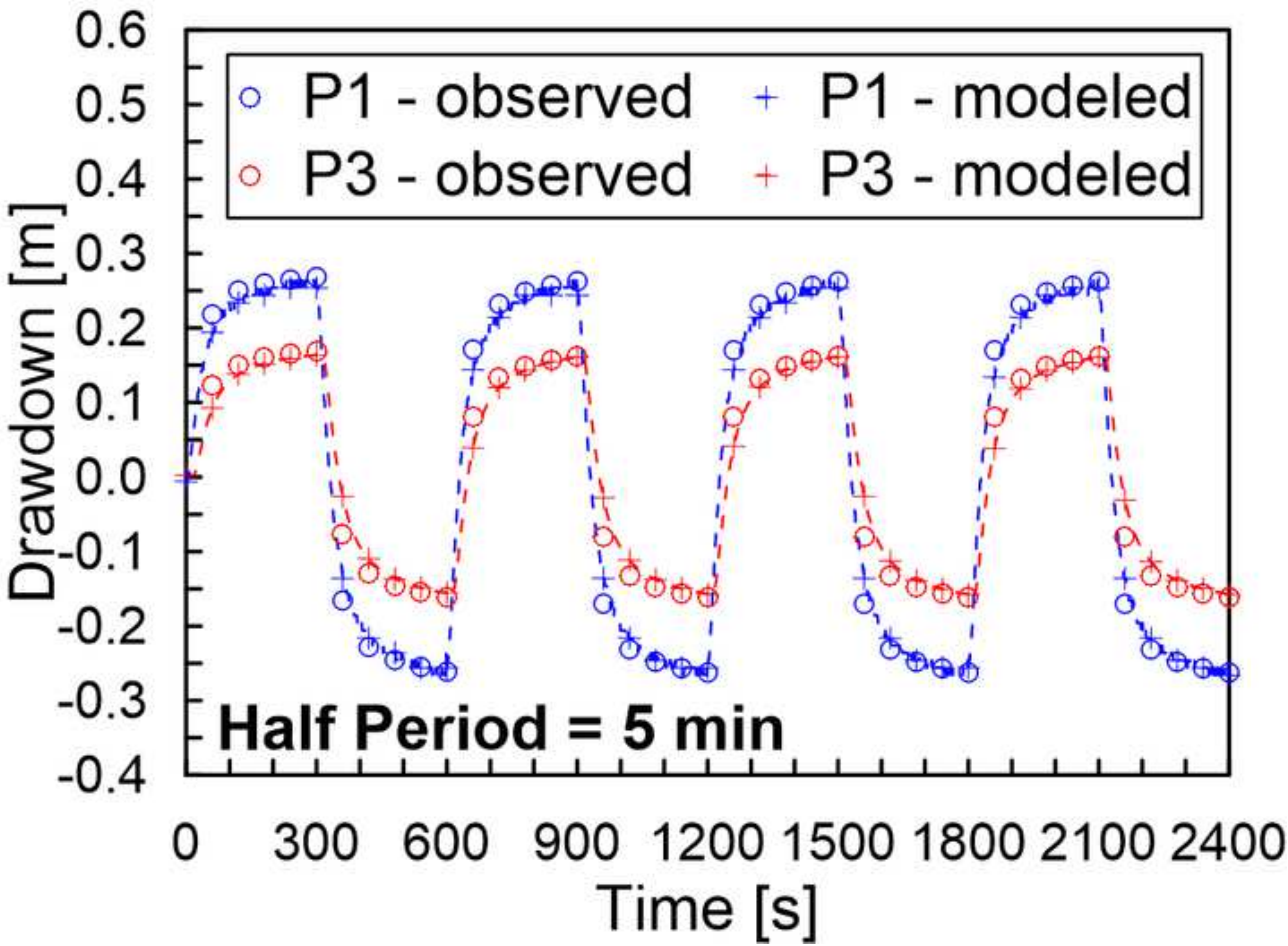


Figure 10





0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
Drawdown [m]
P1 - observed
P1 - modeled
P3 - observed
P3 - modeled
Half Period = 5 min
0
300
600
900
1200
1500
1800
2100
2400
Time [s]