Estimating aquifer properties using groundwater hydrograph modeling

V. Shapoori [1*], T.J. Peterson[1], A.W. Western[1] and J.F. Costelloe[1]

[1]Department of Infrastructure Engineering, the University of Melbourne, Parkville, Vic 3010, Australia

*corresponding author

Accepted Article

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: 10.1002/hyp.10583

This article is protected by copyright. All rights reserved.

Abstract:

Estimation of aquifer hydraulic properties is essential for predicting the response of an aquifer to extractions and hence estimating the availability of the groundwater resources. Aquifer tests are commonly used for the estimation of aquifer properties; however they can be expensive and often only characterize the short-term response of the aquifer. This paper presents a time series modeling approach to estimating aquifer hydraulic properties. It is applied to 42 bores monitoring an unconfined aquifer within an irrigation region of south-eastern Australia and the resulting probabilistic estimate of hydraulic properties are evaluated against pumping test estimates. It is demonstrated that the time-series modeling can provide a reliable estimate of the hydraulic properties that are typical of a very long-term pumping test. Furthermore, the application of the time-series modeling to 42 bores provided novel insights into the aquifer heterogeneity. We encourage others to further test the approach and the source-code is available from:

http://www.mathworks.com/matlabcentral/fileexchange/48546-peterson-tim-j-groundwater-statistics-toolbox

Keywords: groundwater, aquifer properties, time series modeling.

Accepted Article

This article is protected by copyright. All rights reserved.

1. Introduction

The flow and storage properties of an aquifer are the key hydrogeological properties required for predicting aquifer responses to extraction and hence to estimate the availability of the groundwater resources for consumptive use (Wu *et al.*, 2005; Yeh and Lee, 2007; Butler 2009). Recently, time series modeling has been introduced as a new approach for quantifying such hydrogeological properties (Harp and Vesselinov, 2011; Obergfell *et al.*, 2013). However, to date, the aquifer hydraulic property estimates have not been derived probabilistically nor have they been evaluated against other independent data. This paper extends the current literature by presenting probabilistic estimates of aquifer hydraulic properties using a time-series model and then demonstrating the robustness of this approach by evaluating the aquifer property estimates against multiple pumping tests results.

In common practice, aquifer tests are used for the estimation of aquifer properties and hence predicting the aquifer response over different pumping regimes and scenarios (Yeh and Lee, 2007; Butler 2009; Obergfell *et al.*, 2013). This test usually involves constant pumping at the production bore and measuring the drawdown at observation bores nearby. The analytical models, mainly based on steady state or transient solutions, are then used to model the drawdown and estimate the aquifer properties by matching the model simulation and observed drawdown (Kruseman and De Ridder, 1994). Applying such a test at only one location usually provides average estimates of aquifer properties, representing the overall response of the aquifer to pumping (Butler 2009). For more detailed investigation into spatial variation of aquifer properties, one option is to undertake sequentially several pumping tests at a number of bores in a systematic manner and use methods such as tomography to analyze the head responses and characterize the aquifer properties at finer scales (Yeh and Liu, 2000;

Accepted Article

This article is protected by copyright. All rights reserved.

Bohling *et al.*, 2002; Bohling *et al.*, 2007; Yeh and Lee, 2007; Butler 2009). While the pumping tests are extensively used and the effectiveness of this approach in providing valuable information about the aquifer characteristics has been proven, there remain several challenges when applying the aquifer test (Wu *et al.*, 2005; Yeh and Lee, 2007; Butler 2009).

First, pumping tests are quite expensive and involve a significant labor cost (Yeh and Liu, 2000; Obergfell *et al.*, 2013), in particular when several pumping tests need to be undertaken to quantify the aquifer property variation over a large area. Secondly, these tests are often applied over short periods of time (Kruseman and De Ridder, 1994) (e.g. between 24 to 72 hours) and hence characterize the short-term response of the aquifer. However, in many cases, the response of the aquifer to pumping can be different if pumping continues for a long period (Butler 2009). For example, the aquifer may respond consistently during a short period of pumping (e.g. 72 hour) but at longer periods of pumping (e.g. weeks or several months), other additional factors (e.g. delayed gravity response) can affect the drawdown and change the way in which the groundwater responds to pumping (Kruseman and De Ridder, 1994; Butler 2009). In such circumstances, undertaking pumping tests over a short duration would provide aquifer parameters appropriate for estimations of the short-term response of the aquifer. When used to estimate the impact of pumping over several months, these parameters can lead to substantially under or over-estimation of the long-term drawdown (Kruseman and De Ridder, 1994; Moench *et al.*, 2001).

An alternative to conducting aquifer tests is to use monitoring network potentiometric data (Yeh and Lee, 2007; Harp and Vesselinov, 2011). These monitoring data often reflect the long term influence of pumping; similar to that obtained from a long-term pumping test. Hence, the monitoring data can potentially provide insights into the long-term responsiveness of the aquifer to pumping. Furthermore, insights into aquifer heterogeneity could potentially

This article is protected by copyright. All rights reserved.
Accepted Article

be quantified if pumping occurred at several bores. Realizing these potential outcomes would address the prior mentioned challenges of pumping tests and derive greater value from monitoring networks.

Unlike pumping test drawdown data, long-term groundwater hydrographs are likely to show the influence of many natural stressors (e.g. recharge, groundwater evapotranspiration). Therefore, for pumping investigation using long-term observation records, detailed modeling of these stressors is required. One option is to build a numerical groundwater flow model based on the understood local processes and hydrogeology and then estimate the model parameters (i.e. aquifer hydraulic properties) by calibration. When the monitoring network is sufficiently dense, calibration of numerical groundwater models to the monitoring bores can potentially provide a reasonable estimate of the aquifer hydraulic properties and their heterogeneity. More commonly the monitoring network is too sparse for calibration to adequately capture the aquifer heterogeneity. This problem of a lack of observation data relative to the number of model parameters (i.e. the hydraulic properties for each model grid cell) causes the problem to be ill-posed (Zhou *et al.*, 2014). To address this challenge, and to adequately simulate the possible aquifer heterogeneity, a number of calibration approaches have been proposed. These range from the delineating zones of constant hydraulic properties (Medina and Carrera, 1996; Zhou *et al.*, 2014) to the use of geostatistical simulations of hydraulic properties that are iteratively perturbed to identify, not one final map of aquifer hydraulic properties, but a set of maps, each of which results in the groundwater model showing adequate agreement with the monitoring data (Gómez-Hernánez *et al.*, 1997; Franssen *et al.*, 2003; Li *et al.*, 2012; Zhou *et al.*, 2014). On the latter, a number of approaches have been developed for the sequential refinement of the geostatistical model (Caers 2007; Hu, 2008) and, while they have considerable computational requirements and a spatial field of observations, they can produce highly heterogeneous stochastic estimates of

This article is protected by copyright. All rights reserved.

aquifer hydraulic properties. Similarly, when pump tests have been conducted at multiple sufficiently proximate sites, instead of long-term water level observations, various tomography approaches exist for the stochastic estimation of the heterogeneity of aquifer hydraulic properties (Liu and Kitanidis, 2011; Alzraiee *et al.*, 2014). However, like stochastic inverse modeling, the computational demands are also considerable.

An alternative approach for estimating the aquifer hydraulic properties is the use of simpler models, such as time series modeling. Similar to inverse groundwater modeling, a time-series model that accounts for pumping drawdown could be calibrated to long-term water level observations and then the aquifer hydraulic property estimates could be derived from the model parameters. However, time series model are more applicable to areas with relatively simple hydrogeological conditions than to complex situations. This is mainly because of simplified assumptions often made during their construction. For areas with relatively simple configuration, either numerical or time series models can be adopted with similar assumptions and the same results can be achieved. An advantage of time series models is that they are easier to formulate than a numerical model. In addition, by their nature, time-series models are very oriented towards the observation data, which make them particularly useful for bore-by-bore investigations. Therefore, in the right circumstances, time series models are a valuable alternative option warranting further investigation.

Recently, time series models have been developed to quantify pumping drawdown (Von Asmuth *et al.*, 2008; Shapoori *et al.*, 2015) and to estimate aquifer hydraulic properties (Obergfell *et al.*, 2013). They used a continuous form of a transfer function noise approach (Box and Jenkins, 1970), whereby historic input forcing data (i.e. pumping and climate data) were effectively weighted and summed to estimate the groundwater level displacement from a datum groundwater level. In doing so, the drawdown from pumping was modeled using the

Accepted Article

This article is protected by copyright. All rights reserved.

instantaneous form an analytical pumping equation (e.g. Theis or Hantush). Obergfell *et al.* (2013) used this approach to quantify the aquifer properties at eleven observation bores near a well field consisting of seven production bores. Harp and Vesselinov (2011) used a similar approach to Von Asmuth *et al*. (2008) and Obergfell *et al*. (2013) to quantify aquifer properties. The model was applied to an area with three observation bores and seven pumping bores. While it was shown in both Obergfell *et al*. (2013) and Harp and Vesselinov (2011) studies that it is possible to use the time series model and extract aquifer properties from observed groundwater level variation, neither studies evaluated the aquifer property estimations against other independent data. Such evaluation is essential to provide evidence that reliable estimates of the aquifer properties have been achieved. In addition, in these two studies, a deterministic calibration method (e.g. Levenberg-Marquardt algorithm (Marquardt, 1963)) was implemented to derive a single best set of aquifer hydraulic property estimates and simple uncertainty measures (linear prediction intervals) were adopted to quantify the uncertainty in the hydraulic parameter estimates. A more rigorous approach would be to use more advanced methods, such as Bayesian approaches (Kuczera and Parent, 1998; Vrugt *et al.*, 2003; Vrugt *et al.*, 2009), which produce an ensemble of hydraulic parameter sets capable of the plausible representation of the observed data. Prediction uncertainties from these parameter set ensembles tend to be more plausible since they avoid reliance on linearity assumptions.

The objective of this paper is to derive probabilistic estimates of transmissivity and storage from the time-series models, use these estimates to explore the aquifer heterogeneity and then evaluate the estimates against pumping test estimates. The investigation is undertaken for an agricultural region in south-eastern Australia using the time-series model structures developed by Shapoori *et al*. (2015).

Accepted Article

This article is protected by copyright. All rights reserved.

2. Study area

The Clydebank groundwater sub-region of the Gippsland Basin in south-eastern Australia (Figure 1) was used as the study area. This area is located to the east of Lake Wellington and surrounded by the Avon and Thompson Rivers to the north and south, respectively. The region is classified as temperate climate with mild winter in the Köppen–Geiger climate classification (Peel *et al.*, 2007) and used mainly for dryland grazing.

2.1. Hydrogeology

The top 20-50 meters in depth consists of two layers of sedimentary deposits formed from the early Pliocene to recent times. The upper layer is a sequence of clays (dominant), sands and gravels (Jenkin, 1966; Hocking, 1976) and is known as the 'prior stream deposits'. This layer can be divided into upper and lower sub-units. In Figure 2, the extent of prior stream deposits layer within the study area is shown. Overall, the upper sub-unit is composed of fine materials (e.g. sandy clay and silty clay) with an average thickness of 8 meters. Within the west of the study area, this upper sub-unit is dominated mainly by sandy clay, while it changes gradually to silty clay (e.g. finer materials) to the east. The lower sub-unit contains mainly well-developed beds of coarse sand and gravel (yellow colored layer in Figure 2). This sub-unit has an average thickness of 10 meters and is thicker in the east (i.e. near Lake Kakydra) than in the west. The observed hydrographs in eight nested bores indicate that these two subunits are highly connected and respond similarly to climate or pumping signal. The 'prior stream deposits' layer is considered as the shallow unconfined aquifer and mainly used for private irrigation and public wells. The second layer is the Haunted Hills Formation, which has a low permeability and acts as an aquitard, separating the shallow unconfined aquifer from deeper layers (Shapoori *et al.*, 2015). Further details of the geology of study area can be found in Shapoori *et al*. (2015).

This article is protected by copyright. All rights reserved.

Accepted Article

2.2. Groundwater pumping and observed data

Between 1994 and 2005, nine groundwater extraction bores were constructed and screened within the ‘prior stream deposits’ layer to lower the water table and reduce land salinization in this area. The pumps operate during winter and spring and the pumping rate is usually set at the beginning of each month by the local resource management authority (Southern Rural Water). On average, the pumps have operated 14 years and monthly extraction rates have been recorded since operations commenced.

The water level is monitored monthly at 54 observation wells (Figure 1) screened within the ‘prior streams deposit’ layer. The observations are recorded monthly and observation record length is within the range of 7 to 23 years. At 46 out of 54 bores, the observation records start before the commencement of pumping at the closest pumping bore. The detailed information for all observation and pumping bores is available as electronic supplementary material (ESM) in Shapoori *et al*. (2015). To provide climatic forcing data for the time series model, following Shapoori *et al*. (2015), daily precipitation was obtained from East Sale Airport weather station (Lat 38.12 °S, Long 14.13 °E) and daily potential evapotranspiration was estimated from the Morton complementary relationship areal model (Morton, 1983) using daily values of maximum and minimum temperature, vapor pressure and net radiation from the same weather station.

During the installation of nine public pumping bores, pumping tests were also undertaken to investigate the drawdown characteristics at the pumping bores. These tests usually took between 24 to 48 hours and during the test, the water level changes were recorded at observation bores nearby. However, the details of all pumping tests are not currently available. So, in this study, those bores with detailed pumping test datasets were obtained from previous pumping tests reports by SKM (1999), (1995) and (2003) and used for

This article is protected by copyright. All rights reserved.

Accepted Article

pumping test analysis. In addition to those short pumping tests, two additional longer pumping tests (195 and 267 hours) were conducted for pumping bore 25 by the local management authority (Toogood 2000). These two long pumping tests were also used to estimate the aquifer properties. The details of pumping tests and observation bores are available as electronic supplementary material (ESM).

3. Methods

In this section, the transfer function noise model with pumping is first outlined following by details of the method used to derive the aquifer properties from the models. The parameter uncertainty estimation approach is then presented followed by details of the pumping test analysis and the approach by which these estimates were compared against those from the time-series model.

3.1. The transfer function noise model

Overall, the groundwater transfer function noise model describes the groundwater level elevation (h_t [L]) as the combined effects of forcing data (h_t^* [L]) plus the local fixed drainage level (d [L]) and the remaining residual series (e_t [L]), which consist of any other unknown influences not accounted for (Eq. (1)):

$$h_t = h_t^* + e_t + d \tag{1}$$

h_t^* consists of past historic time series of forcing data (R) and an impulse response function (θ [-]) for each stressor (e.g. rainfall, evapotranspiration (ET) or pumping), which weights all of the historic forcing. Any change in groundwater level can then be derived by summing the integral of each weighted forcing (Eq. (2)).

$$h_t^* = \sum_{i=1}^{m} \left(\int_{-\infty}^{t} R_i\,(\tau)\theta_i(\tau)d\tau \right) \tag{2}$$

This article is protected by copyright. All rights reserved.
Accepted Article

To identify the statistically significant stressors throughout the study area, Shapoori *et al.* (2015) undertook a top-down evaluation of possible drivers influential at each observation bore; which included single and multiple pumping bores with and without recharge boundaries, phreatic ET and free drainage. It was shown that phreatic ET was not significant at any bore and that groundwater pumping was influential at 42 of the 54 observation bores. For this study, hydraulic properties were estimated for these 42 observation bores using the optimal model structure form from Shapoori *et al*. (2015). In doing so, the head at each bore was simulated as follows (Eq. 3):

$$h_t = \int_{-\infty}^{t} R_\tau \,\theta_p(t-\tau)d\tau - \int_{-\infty}^{t} Q_\tau \theta_F\,(t-\tau)d\tau + e_t \,+ d \tag{3}$$

where the integrals simulate the impact of current and past recharge, R_τ [LT^{-1}], and pumping, Q_τ [L^3T^{-1}], on the groundwater level at the current time step and θ_p [-] and θ_F [-] are the response functions for recharge and pumping respectively.

With regard to the recharge estimations, following Shapoori *et al*. (2015), the one layer soil moisture model from Peterson and Western (2014) was adopted to transform daily precipitation and potential ET (PET) into a free-drainage recharge estimate. A modified version of Pearson type III distribution from Peterson and Western (2014) was also adopted for θ_p.

For the pumping response function (θ_F), Shapoori *et al.* (2015) adopted the Ferris and Knowles well formula (Ferris and Knowles, 1963) (Eq. 4). It is an instantaneous or impulse form of the Theis equation. The original Theis analytical equation is a step response function and it assumes the pumping rate is constant. Considering that in our case, pumping varies, the equation used for pumping response function (θ_F in Eq. 3) needs to be an impulse response type and hence Ferris and Knowles well formula is preferred over the use of the original

Accepted Article

This article is protected by copyright. All rights reserved.

Theis equation. It should be mentioned that second integral in Eq. 3 essentially superimposes the results from each time step to obtain the long-term pumping response. The Ferris and Knowles equation estimates the change in groundwater head (X [L]) from an instantaneous extraction of water (V [L3]).

$$X = \frac{V}{4\pi Tt} exp(-\frac{r^2 S}{4Tt}) \quad (4)$$

where , T [L^2T^{-1}] and S [-] are the transmissivity and storativity of the aquifer around the well respectively; r is the distance from the pumping well to the observation bores. Following Shapoori *et al*. (2015), the pumping impulse function is derived from Ferris and Knowles' well formula by replacing the terms $\frac{1}{4\pi T}$ and $\frac{S}{4T}$ with parameters α and β respectively (Eq. (5)).

$$\theta_F(t) = \frac{\alpha}{t} exp(-\frac{r^2 \beta}{t}) \quad (5)$$

In general, Ferris and Knowles' well formula has several assumptions. That is, the aquifer needs to be homogenous, of infinite extent, the saturated thickness should be constant, storage changes should occur instantaneously in response to head changes, and the aquifer should be affected by only one pumping bore. In our study area, some of these assumptions are not valid. The aquifer is unconfined and delayed gravity yield response is likely to be important. The aquifer is generally thin (10-20 m) and given the observed drawdown of more than 5 m, the saturated thickness is not constant. In addition, there are areas impacted by multiple pumping bores. In Shapoori *et al.* (2015), three extensions to Eq. 5 were adopted to account for time-varying saturated thickness, local lake recharge and multiple pumping bores in this area. These extensions were included in the model structure for bores in which there was statistical evidence for their inclusion (i.e. model fits improved). Details of those inclusions can be found in Shapoori *et al.* (2015). We used the final pumping structure for each bore derived from Shapoori *et al.* (2015) to estimate the aquifer properties. The key

Accepted Article

This article is protected by copyright. All rights reserved.

remaining invalid assumption not considered by Shapoori *et al.* (2015) is the delayed gravity yield. At this stage, we accept this limitation given the computational difficulty in the inclusion of such an extra process in the pumping component of the model. Further discussion relating to this problem is provided in section 5.2.

To estimate the aquifer hydraulic properties, the pumping parameters from the fitted time series model (i.e. calibrated α and β parameters) were used to estimate the aquifer properties. This was achieved by first estimating transmissivity (*T*, Eq. (6)) from the parameter in Eq. (5) and then using that estimation, as well as β, to get the storativity (*S*) (Eq. (7)):

$$T = \frac{1}{4\pi\alpha} \quad (6)$$

$$S = 4T\beta \quad (7)$$

In simulating pumping drawdown, Shapoori *et al*. (2015) evaluated three extensions to Eq. 5 to account for time-varying saturated thickness, lake recharge and multiple pumping bores. Those extensions were included in bores in which there was statistical evidence for their inclusion (i.e. model fits improved). Details of those inclusions can be found in Shapoori *et al*. (2015). It should be noted that those extensions will not change the way in which the transmissivity and storativity are estimated in Eq. (6) and (7) and hence the same method was used to estimate the aquifer properties at those bores in which those extensions were applied. The time series model presented in this section is hence referred as '*soil moisture transfer function noise model (SMS-TFN)*'.

3.2. Parameter uncertainty analysis

In this study, a formal Bayesian approach, the DREAM method (Vrugt *et al.*, 2009), was adopted to determine the parameter uncertainty. In summary, DREAM runs multiple chains with initial parameter sets derived from prior distribution and then updates continuously the

Accepted Article

This article is protected by copyright. All rights reserved.

scale and orientation of the prior distribution to globally explore the parameter space and estimate the posterior distribution of each parameter. During the evolution process, it uses a genetic algorithm to generate new candidate parameter sets and a metropolis rule to decide whether to choose the new parameter sets (Vrugt *et al.*, 2008; Vrugt *et al.*, 2009). The reasoning for choosing this method is that the proposed time series model is sometimes non-linear around the optima and the effectiveness of the DREAM approach has been demonstrated for complex and highly non-linear target distributions (Vrugt *et al.*, 2008; Vrugt *et al.*, 2009).

In this study, the number of chains for DREAM was set to the number of model parameters. Prior parameter distributions were assumed to be uniform. The prior distribution ranges are detailed in Shapoori *et al*. (2015). In total, 500,000 model runs were undertaken using the DREAM approach. The weighted least square objective function presented in Peterson and Western (2014) and Von Asmuth *et al*. (2002) was used to measure the difference between the model prediction and observation data. The last 10% of samples in each chain were used to derive the posterior parameter distributions. To summarize the parameter uncertainty ranges at each bore, the cumulative distribution functions (cdf) were calculated and then 5th and 95th percentiles were estimated. The 50th percentile was also estimated from the cdf to show the median parameter values.

3.3. Pumping test analysis

Pumping test drawdown data were available from 8 observation bores. In undertaking the pump-test analysis, the Theis method (Theis, 1935) was initially applied to all observation bores using AQTESOLVE software and a least-squares objective function (Hydrosolve, 2014). Overall, the short-term (i.e. less than 48 hours) pumping tests represent the rapid response of the aquifer to pumping. The Theis simulations fitted very well to observed

This article is protected by copyright. All rights reserved.

pumping test data for the short-term (i.e. less than 48 hours) pumping tests (Coefficient of Efficiency (COE) > 0.95); however, for longer pumping tests (i.e. more than 195 hours), the COE declined to 0.75. For the long-term pumping tests, delayed gravity yield is likely to be a significant factor contributing to this decline in the Theis modeling performance. Therefore, we used the Neuman method (Neuman 1975) (i.e. method for unconfined aquifer with delayed gravity yield impact) using AQTESOLVE to analyze the long-term pumping tests. Considering that the AQTESOLVE software provides only the optimum aquifer property values, the DREAM method detailed in Section 3.3 was adopted to obtain the uncertainties of the estimates. This uncertainty analysis was only undertaken for the short pumping tests. In doing so, the Theis equation (Eq. 5) was implemented separately and coupled with DREAM method to obtain the uncertainty of the aquifer property estimates.

4. Results:

4.1. Aquifer property estimates derived from time series analysis

4.1.1. Overall estimation of aquifer properties

Figure 3 presents maps of the median transmissivity and storativity obtained from the time-series modeling posterior parameter distribution at each observation bore. The overall pattern of aquifer properties at this scale indicates that the area can be divided hydrogeologically into two regions. One is the area around pumping bores 21, 23, 25, 27 and 29 (the west side of the region) in which the transmissivity is lower than 300 m^2/day and storativity changes from 0.01 to 0.2. The other area is in the east, near pumping bores 20, 22, and 24, where larger transmissivity values are modeled, and for most bores, the storativity is smaller than in the west region. Figure 3 also shows that at the smaller scale, the aquifer property estimates tend to be reasonably consistent around each pumping bore. Given that the time series model was applied separately at each observation bore, the aquifer property estimates at different bores

This article is protected by copyright. All rights reserved.

can be considered as an independent estimate. This spatial consistency in the aquifer property estimates provides more confidence that reliable estimates of aquifer properties have been achieved.

To explore the uncertainty in the estimates and to more rigorously investigate the existence of the two hydrogeological regions, the cumulative distribution functions (CDF) were derived and are summarized in Figure 4. The CDFs for observation bores located on the west side of the area are colored green and for the east side are colored red. In addition, for those in which the pumping signal is weak (i.e. pumping drawdown is <0.7 m), the CDF is dashed. Figure 4 shows that for many bores, the uncertainty (i.e. spread of the CDF) is less than the spatial variability between the two regions (east and west), and that there is no clear difference in the uncertainty between the two regions.

Figure 4 also shows greater uncertainty, as a ratio to the median estimate, in storativity than the transmissivity estimates. This is primarily due to the parameter sensitivity. That is, the transmissivity is a function of only the alpha parameter (Eq. 6) while the storativity is a function of both the alpha and beta parameters (Eq. 7). Figure 4 also shows that when the drawdown is not significant (e.g. CDFs with dashed line in Figure 4), the uncertainty often increases. This can be particularly seen in the uncertainty range for those bores with higher transmissivities (e.g. > 800 m^2/day) in the east region that correspond to instances where the pumping signals are small. Between those with transmissivity > 800 m^2/day, there is only one case where the pumping signal is considerable but the uncertainty is quite large. This is due to poor model performance (Coefficient of Efficiency (COE) < 0.5).

To further explore how the uncertainty varies spatially, the ratio of the 90 % prediction uncertainty to the median is shown in Figure 5. It indicates that the uncertainty ratio is spatially correlated but dependent upon the model performance; that is, the fit to the observed

This article is protected by copyright. All rights reserved.

hydrograph. For example, for bores around pumping bore 24 and between pumping bore 22 and Lake Kakydra (east side of the study area) good model performances were achieved (Shapoori *et al.*, 2015) and the uncertainty is consistently low. For bores around pumping bores 21, 23 and 29, where poorer model performances were achieved (Shapoori *et al.*, 2015), uncertainty is consistently high.

4.2. Evaluation of estimated aquifer properties

4.2.1. Comparison with short term pumping test results

This section presents an evaluation of the time-series estimates of transmissivity (T) and storativity (S) against pumping test estimates at eight observation bores. Figure 6 presents scatter plots of T and S from the posterior distribution for both pumping test analysis and SMS-TFN model (note, the y axis in Figure 6 is a log scale). It shows that the estimates of T and S from the SMS-TFN models do not match the estimates from the short-term pumping tests. Overall, the T values from the pumping tests are greater than that from SMS-TFN modeling. Conversely, the S values from pumping tests are at least one or two orders of magnitude smaller than that estimates from the SMS-TFN models. The uncertainty estimates in Figure 6 show that T and S are highly correlated in both the SMS-TFN and pumping test analysis. It also shows that uncertainty in T values from SMS-TFN model and pumping test analysis intersect at four of the eight bores (Figure 6a-c,h) while there is no intersection for the S. However, at the 90% uncertainties range of T and S (not shown here), there is no intersection and hence the estimates appear significantly different.

4.2.2. Comparison with long term pumping test

Figure 6 showed that the estimates from the SMS-TFN modeling and the short term pumping tests are significantly different. While the differences in T estimates were modest, the S

This article is protected by copyright. All rights reserved.

estimates could differ by an order of magnitude or more. This inconsistency may be due to the time scale of the hydrographs used as the input to each analysis. That is, the SMS-TFN model used hydrographs experiencing several years of pumping while the hydrographs used for pumping test analysis were of one or two days duration.

To explore the hypothesis that the T and S estimates (from both SMS-TFN and pumping tests) are sensitive to the duration of pumping, pumping test results from three durations (33, 195 and 265 hours) at two observation bores (pumping at bore 25) were analyzed. Figure 7a-d shows that during the first few hours of pumping, the drawdown is rapid for both short and long-term pumping test. However, after approximately 20 hours, the water level either stabilizes or declines at a slower rate in comparison to the early drawdown (i.e. <20 hours) in the long-term (195 hours) pumping test (see Figure 7c-d). This indicates that the aquifer responds differently for short and long periods of pumping. Figure 7e-f also shows similar response to pumping at the early time stage. However, at the later times, the pumping impact is obscured by rainfall.

Figure 7a-b illustrates the estimated modeled drawdown from using the aquifer parameters obtained from the SMS-TFN model and short term pumping test. Similarly, Figure 7c-f shows the pumping drawdown for the longer tests from the aquifer properties obtained from either the SMS-TFN model or the short term pumping test. While the pumping simulations from SMS-TFN model parameters considerably underestimate the pumping impact during the early pumping phase (see Figure 7a-b), it converges to a similar drawdown in the later phase of pumping. In contrast, the pumping simulation from short term pumping tests can acceptably estimate the early drawdown but overestimates the drawdown in the late pumping phase.

This article is protected by copyright. All rights reserved.

4.2.3. Aquifer property estimations from long-term drawdown

In the previous section, it was shown that the aquifer responds differently to pumping during the mid to late phases of drawdown. This response often happens within unconfined and semi-confined aquifer and is known as the delayed yield response. To investigate, the Neuman (1975) method was applied to account for this complex drawdown and to re-estimate the aquifer properties when accounting for a delayed yield response. In the Neuman method, the transmissivity and two storativity values are estimated (i.e. elastic storativity and specific yield) with the latter representing the response of the aquifer to short and long periods of pumping. The AQTSOLVE implementation of Neuman's method was applied to the 195 hour pump test at bores 136529 and 52874. A comparison of the estimates with that from the SMS-TFN model is shown in Figure 8 and Table 1.

Figure 8 shows that the Neuman method accurately simulates the early and late stage of drawdown. With regard to aquifer property estimation, Table 1 shows that for both bores 136529 and 52874, T from Neuman's method is comparable to that from the SMS-TFN modeling and the pumping test analysis. However, S from Neuman's method is only consistent with that from the SMS-TFN modeling. Interestingly, the elastic storativity (S_s) estimation from Neuman's method, obtained from the first segments of drawdown (early rapid decline of groundwater level), is in good agreement with S from the 33 hours pumping test. This indicates that the SMS-TFN model provides estimates of T and S that are most representative of the long term drawdown.

This article is protected by copyright. All rights reserved.

5. Discussion and Conclusion

5.1. Insights into spatial variations in hydrogeology

It has been demonstrated that the SMS-TFN time-series model can identify spatial variations in aquifer properties. For the study area, the eastern region was found to have a mean transmissivity 3.5 times higher, and storativity 1.5 orders of magnitude lower, than the western region. The production and observation bores are screened within a sandy gravel aquifer overlain by hydraulically connected layers of sandy and/or silty clay. Hence, the estimated hydraulic properties represent the bulk transmissivity and storativity of the two layers rather than just that of the screened layer. Therefore, any general pattern in the aquifer properties would be an indication of change in the overall aquifer characteristics and/or aquifer geometry conditions. From the lithology data, the eastern region around pumping bore 20 and 22 has a thicker layer of gravel and sand, which is in agreement with the result of this study indicating a more transmissive aquifer in the east. But the aquifer thickness is unlikely to be proportionally 3.5 times thicker in the east compared to the western region, suggesting that there is some contribution from the hydraulic conductivity as well as the thickness influences in an increase of transmissivity values. This highlights the potential for the SMS-TFN modeling to provide insights into the large scale aquifer heterogeneity.

5.2. Time series analysis and pumping test

It has long been recognized that the duration of a pumping test can have a significant impact on the hydraulic property estimation (Kruseman and De Ridder, 1994; Walton, 2007). This is particularly important in unconfined or semi-confined aquifer where the phenomenon of delayed gravity yield response usually occurs (Boulton 1954; Neuman, 1975; Kruseman and De Ridder, 1994). In this study, the delayed gravity yield effect seems to be an important factor, resulting in different aquifer responses to short and long periods of pumping. That is,

This article is protected by copyright. All rights reserved.

Accepted Article

the same analytical equation (i.e. Theis equation) was used in both the SMS-TFN model and short term pumping test analysis, but the delayed gravity yield led to differing estimates of T and S. The SMS-TFN model used a much longer time series of drawdown and thus the hydraulic aquifer parameters from the SMS-TFN model are considered to represent the long-term response to pumping (see Figure 7). Figure 7c-d shows that the observed long-term drawdown approaches the SMS-TFN defined drawdown. However, a clear bias remains between the long-term drawdown and the SMS-TFN drawdown. This is likely to be due to the duration of the long-term pumping test since it is still only a fraction of the time-span represented by the SMS-TFN model. Further analysis and evaluation of the SMS-TFN estimated aquifer properties would require comparison against even longer-term pumping tests (i.e. >1000 hours). In addition, the aquifer hydraulic estimates from SMS-TFN were compared against two long term pumping tests available in this area. Overall, this supported the feasibility of obtaining the plausible estimates of aquifer properties from available hydrograph data using the SMS-TFN model. For more rigorous testing of the approach, more analyses with long term pumping tests need to be undertaken for other study areas. We strongly support such further evaluation of the method presented here.

In general, delayed gravity yield can affect the groundwater hydrograph early in the pumping (Neuman, 1975; Kruseman and De Ridder, 1994). This phenomenon usually takes from a few minutes to several days (Kruseman and De Ridder, 1994; Walton, 2007). This suggests that the SMS-TFN model may be sensitive to when the observed groundwater level is recorded relative to the time of the pumping. If the observed groundwater data are collected long after the pumping started, the delayed gravity yield effect is unlikely to be an important factor and hence the proposed time series model can potentially provide results with high level of accuracy. For example, in the Clydebank region, the observations are recorded mostly between 15th to 25th of each month but the pumping rate is set at the beginning of each

This article is protected by copyright. All rights reserved.

month. Therefore, the groundwater hydrograph mainly contains late drawdown information and hence the influence of delayed gravity yield is minimal, and the SMS-TFN model structure is sufficient to simulate the late response of the pumping for this study area.

In situations where the pumping occurs over small periods of time and/or observed groundwater level data are collected at a high frequency (e.g. hourly) similar to the pumping test, the influence of gravity yield may be an important factor within the groundwater level records. In such situations, it might be beneficial to account for delayed gravity yield in the pumping component of the time series model (Obergfell *et al.*, 2013). However, such an inclusion is challenging in the transfer function model. This is mainly because the pumping equations considering delayed gravity yield (e.g. Neuman equation) are complex and require numerical iteration to be solved. In addition, those solutions assume a constant pumping rate and instantaneous forms of these equations are not available. For the typical case of varying pumping, use of an instantaneous form is important as it considerably simplifies the numerical estimation within the transfer function model by providing the pumping impact with one iteration. Without the instantaneous form, many nested equations need to be solved at every time step and then superimposed to obtain the pumping impact. Furthermore, at every time step the implementation of the Neuman delayed gravity yield equation would require the iterative solution of a number of equations that do not have an analytical solution. Hence, the implementation of more advanced drawdown approaches is possible but the computations burden would be very high and was therefore considered premature when significantly more efficient methods exist.

In summary, the behavior of aquifers over long time periods (i.e. months) is often more significant to resource management than the short-term behavior analyzed by pumping tests with a few days duration. This study highlights the effectiveness of the time series modeling

Accepted Article

This article is protected by copyright. All rights reserved.

approach to maximize the use of available multi-annual groundwater data and to provide valuable insight into aquifer properties without the expense of pumping tests. In regions with available pumping test information, it can be also be used to complement pumping tests by providing long-term bounds to the aquifer parameter estimations. We encourage others to further investigate this powerful approach using the source-code available from http://www.mathworks.com/matlabcentral/fileexchange/48546-peterson-tim-j-groundwater-statistics-toolbox .

Acknowledgments

The authors are grateful for the financial support received from the Australian Research Council (grant numbers: LP0991280 and LP130100958), the Department of Sustainability and Environment, Victoria, Australia; the Department of Primary Industries, Victoria, Australia; and the Bureau of Meteorology, Australia. The authors also thank Mr. Andrew Harrison and Mr. Terry Flynn for providing the pumping data and information on previous work.

References

Alzraiee AH, Baú D, Elhaddad A. 2014. Estimation of heterogeneous aquifer parameters using centralized and decentralized fusion of hydraulic tomography data. Hydrology and Earth System Sciences, **18**: 3207-3223. DOI: 10.5194/hess-18-3207-2014.

Bohling GC, Butler JJ, Jr., Zhan X, Knoll MD. 2007. Field assessment of the value of steady shape hydraulic tomography for characterization of aquifer heterogeneities. Water Resources Research. DOI: 10.1029/2006WR004932.

Bohling GC, Zhan XY, Butler JJ, Zheng L. 2002. Steady shape analysis of tomographic pumping tests for characterization of aquifer heterogeneities. Water Resources Research, **38**. DOI: 10.1029/2001WR001176.

This article is protected by copyright. All rights reserved.

Accepted Article

Boulton NS. 1954. Unsteady radial flow to a pumped well allowing for delayed yield from storage. International Association of Scientific Hydrology, **37**: 472-477.

Box GEP, Jenkins GM. 1970. Time series analysis: forcasting and control. Holden-Day.

Butler J. 2009. Pumping tests for aquifer evaluation—time for a change? Groundwater, **47**: 615-617. DOI: 10.1111/j.1745-6584.2008.00488.x.

Caers J. 2007. Comparing the gradual deformation with the probability perturbation method for solving inverse problems. Mathematical Geology, **39**: 27-52. DOI: 10.1007/s11004-006-9064-6.

Ferris JG, Knowles DB. 1963. The slug-injection test for estimating the coefficient of transmissibility of an aquifer. In: Methods of determining permeability, transmissibility, and drawdown, Bentall R (ed.) U.S.Geological Survey pp: 299-304.

Franssen HH, Gómez-Hernández J, Sahuquillo A. 2003. Coupled inverse modelling of groundwater flow and mass transport and the worth of concentration data. Journal of Hydrology, **281**: 281-295.

Gómez-Hernánez JJ, Sahuquillo A, Capilla J. 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data—I. Theory. Journal of Hydrology, **203**: 162-174.

Harp D, Vesselinov V. 2011. Identification of pumping Influences in long-term water level fluctuations. Groundwater, **49**: 403-414. DOI: 10.1111/j.1745-6584.2010.00725.x.

Hocking JB. 1976. Gippsland Basin. In: Geology of Victoria, Douglas JA, Ferguson JA (eds.) Geological Society of Australia, pp: 248-274.

Hu LY. 2008. Extended probability perturbation method for calibrating stochastic reservoir models. Mathematical geosciences, **40**: 875-885. DOI: 10.1007/s11004-008-9158-4.

Hydrosolve. 2014. Aqtesolv Version 4.0 Professional. Hydrosolve.

Accepted Article

This article is protected by copyright. All rights reserved.

Jenkin JJ. 1966. The geomorphology and upper cainozoic geology of South-East Gippsland. PhD Thesis , University of Melbourne.

Kruseman GP, De Ridder NA. 1994. Analysis and evaluation of pumping test data. 2 Edn., International Institute for Land Reclamation and Improvement.

Kuczera G, Parent E. 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. Journal of Hydrology, **211**: 17.

Li L, Zhou H, Gómez-Hernández JJ, Franssen HH. 2012. Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble Kalman filter. Journal of Hydrology, **428**: 152-169. DOI: 10.1016/j.jhydrol.2012.01.037.

Liu X, Kitanidis PK. 2011. Large-scale inverse modeling with an application in hydraulic tomography. Water Resources Research, **47**. DOI: 10.1029/2010WR009144.

Marquardt DW. 1963. An algorithm for least-squares estimation of nonlinear parameters. Journal of the Society for Industrial and Applied Mathematics, **11**: 431-441.

Medina A, Carrera J. 1996. Coupled estimation of flow and solute transport parameters. Water Resources Research, **32**: 3063-3076.

Moench AF, Garabedian SP, LeBlanc DR. 2001. Estimation of hydraulic parameters from an unconfined aquifer test conducted in a glacial outwash deposit. U.S. Geological Survey , Professional paper: 1629.

Morton FI. 1983. Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology. Journal of Hydrology, **66**: 1-76.

Neuman SP. 1975. Analysis of pumping test data from anisotropic unconfined aquifers considering delayed gravity response. Water Resources Research, **11**: 329-342.

Obergfell C, Bakker M, Zaadnoordijk WJ, Maas K. 2013. Deriving hydrogeological parameters through time series analysis of groundwater head fluctuations around well fields. Hydrogeology Journal **21**: 987-999. DOI: 10.1007/s10040-013-0973-4.

Accepted Article

This article is protected by copyright. All rights reserved.

Peel MC, Finlayson BL, McMahon TA. 2007. Updated world map of the Köppen-Geiger climate classification. Hydrology & Earth System Sciences Discussions, **4**: 439-473.

Peterson TJ, Western AW. 2014. Nonlinear time series modeling of unconfined groundwater head. Water Resources Research, **50**: 8330–8355. DOI: 10.1002/2013WR014800.

Shapoori V, Peterson TJ, Western AW, Costelloe JF. 2015. Top-down groundwater hydrograph time series modeling for climate-pumping decomposition. Hydrogeology Journal. DOI: 10.1007/s10040-014-1223-0 (Published online).

SKM. 1995. Groundwater investigation for Public pump 23 at Yuill Road Bundalaguah. Victorian Department of Environment and Primary Industries pp: 1-15.

SKM. 1999. Drilling and pumping test of 3 public pumping bores in the clydebank subregion. Vitorian Department of Environment and Primary Industries pp: 1-33.

SKM. 2003. Construction and pump testing of production bores for public pump 29. Victorian Department of Environment and Primary Industries pp: 1-32.

Theis C. 1935. The relation between the lowering of the piezometric surface and the rate and duration of discharge of a well using ground-water storage. Eos Transactions American Geophysical Union, **16**: 519-524.

Toogood G. 2000. Aquifer characteristics of pumping bore 25. Sinclair Knight Merz (SKM), pp: 1-29.

Von Asmuth JR, Bierkens MFP, Maas K. 2002. Transfer function-noise modeling in continuous time using predefined impulse response functions. Water Resources Research, **38**: 23. DOI: 10.1029/2001 WR001136.

Von Asmuth JR, Maas K, Bakker M, Petersen J. 2008. Modeling time series of ground water head fluctuations subjected to multiple stresses. Groundwater, **46**: 30-40. DOI: 10.1111/j.1745-6584.2007.00382.x.

Accepted Article

This article is protected by copyright. All rights reserved.

Vrugt JA, Braak CFJt, Clark MP, Hyman JM, Robinson BA. 2008. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research, **44**: W00B09. DOI: 10.1029/2007WR006720.

Vrugt JA, Gupta HV, Bouten W, Sorooshian S. 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water resources research, **39**: 1201. DOI: doi:10.1029/2002WR001642.

Vrugt JA, Ter Braak CJF, Diks CGH, Robinson BA, Hyman JM, Higdon D. 2009. Accelerating markov chain monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling. International Journal of Nonlinear Sciences and Numerical Simulation, **10**: 273-290.

Walton WC. 2007. Aquifer test modeling. CRC Press.

Wu CM, Yeh TCJ, Zhu J, Lee TH, Hsu NS, Chen CH, Sancho AF. 2005. Traditional analysis of aquifer tests: Comparing apples to oranges? . Water Resources Research, **41**: W09402. DOI: 10.1029/2004WR003717.

Yeh TCJ, Liu SY. 2000. Hydraulic tomography: Development of a new aquifer test method. Water Resources Research, **36**: 2095-2105. DOI: 10.1029/2000WR900114.

Yeh TJ, Lee C. 2007. Time to change the way we collect and analyze data for aquifer characterization. Groundwater, **45**: 116-118. DOI: 10.1111/j.1745-6584.2006.00292.x.

Zhou H, Gómez-Hernández JJ, Li L. 2014. Inverse methods in hydrogeology: evolution and recent trends. Advances in Water Resources, **63**: 22-37. DOI: 10.1016/j.advwatres.2013.10.014.

Accepted Article

This article is protected by copyright. All rights reserved.

Table 1. Estimated transmissivity (T) and storativity (S) obtained from AQTESOLVE software using the Neuman (1975) method for 195 hour pumping test and Theis method for 33 hours pumping test for the observation bores 136529 and 52874 and SMS-TFN model. The 90 % confidence intervals are provided inside the bracket. Note that the DREAM uncertainty analysis is applied only to SMS-TFN model and the Theis method.

Bore ID	T (m^2/day)			S			S_s
	SMS-TFN	Pumping test (33 Hours)	Pumping test (195 Hours)	SMS-TFN	Pumping test (33 Hours)	Pumping test (195 Hours)	Pumping test (195 Hours)
136529	248 (227-271)	285 (281-288)	279	8.76E-2 (3.71E-2-2.01E-1)	2.34E-3 (2.11E-2.63E-3)	2.41E-2	7.84E-4
52874	232 (186-305)	540 (491-584)	402	6.40E-2 (2.0E-2-1.7E-1)	2.33E-3 (2.06E-3-2.68E-3)	2.40E-2	1.06E-3

Accepted Article

This article is protected by copyright. All rights reserved.

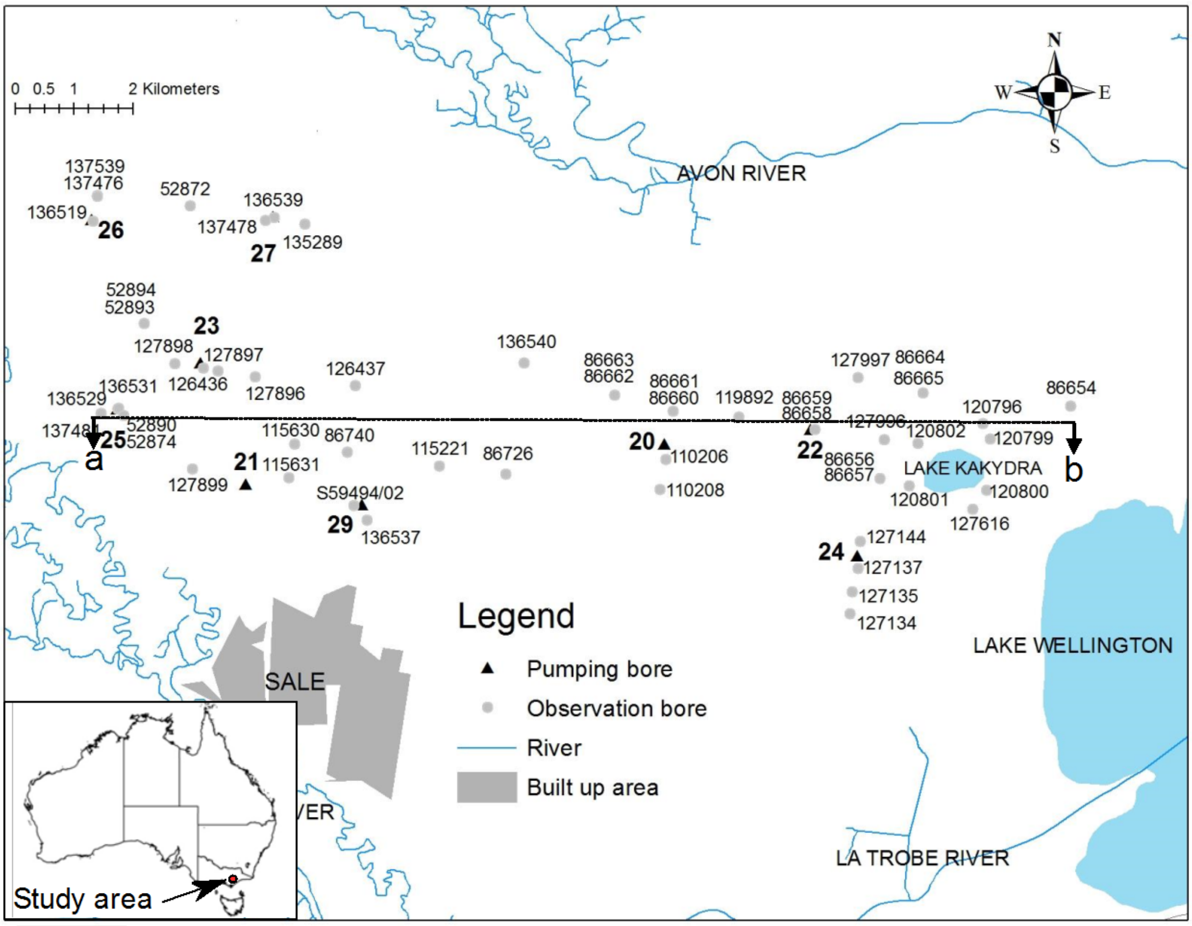

Figure 1: The study area showing the locations of observation and pumping bores.

This article is protected by copyright. All rights reserved.

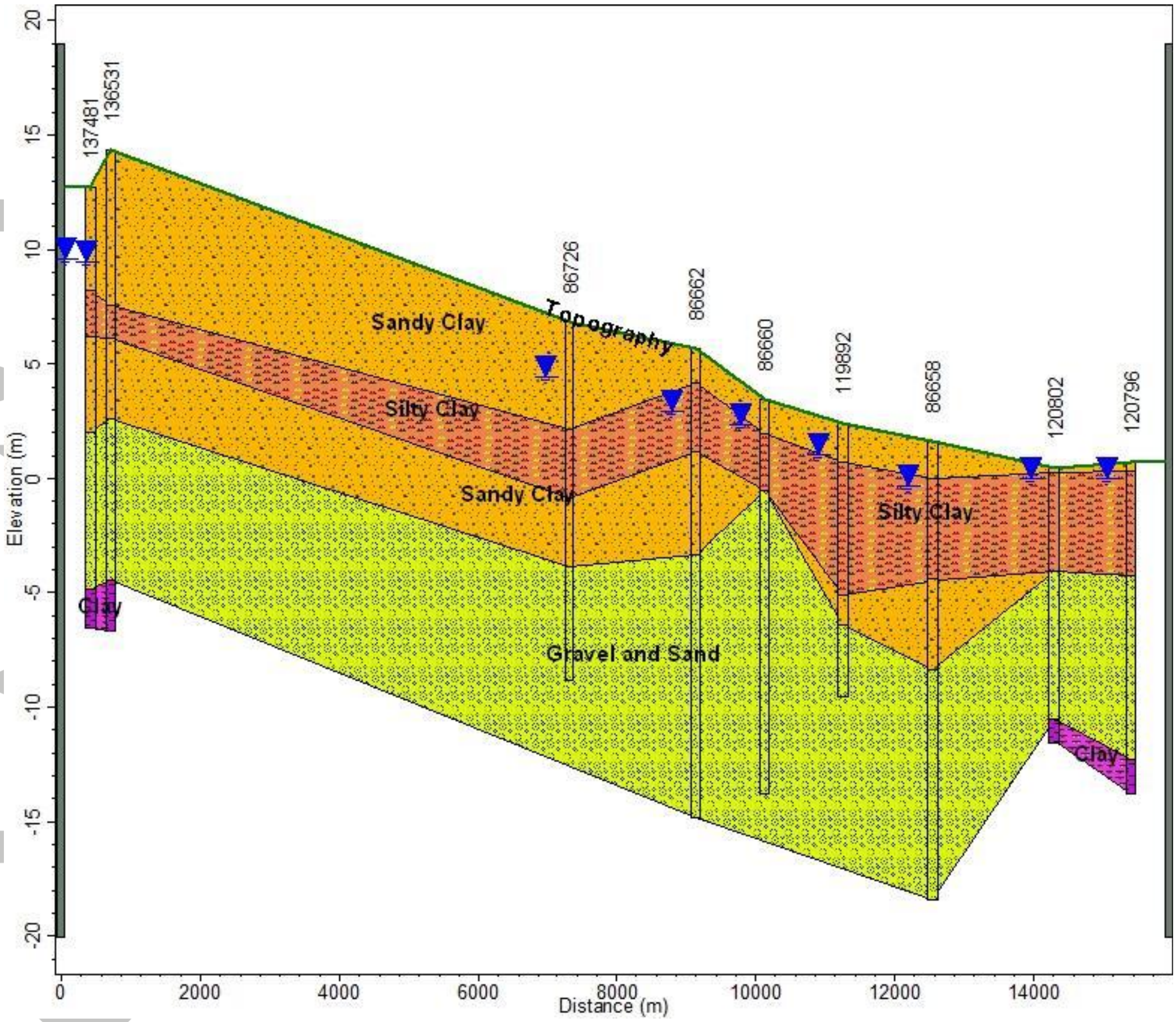

Figure 2: Cross section a-b (see Figure 1) showing the hydrogeological layers..

Accepted

This article is protected by copyright. All rights reserved.

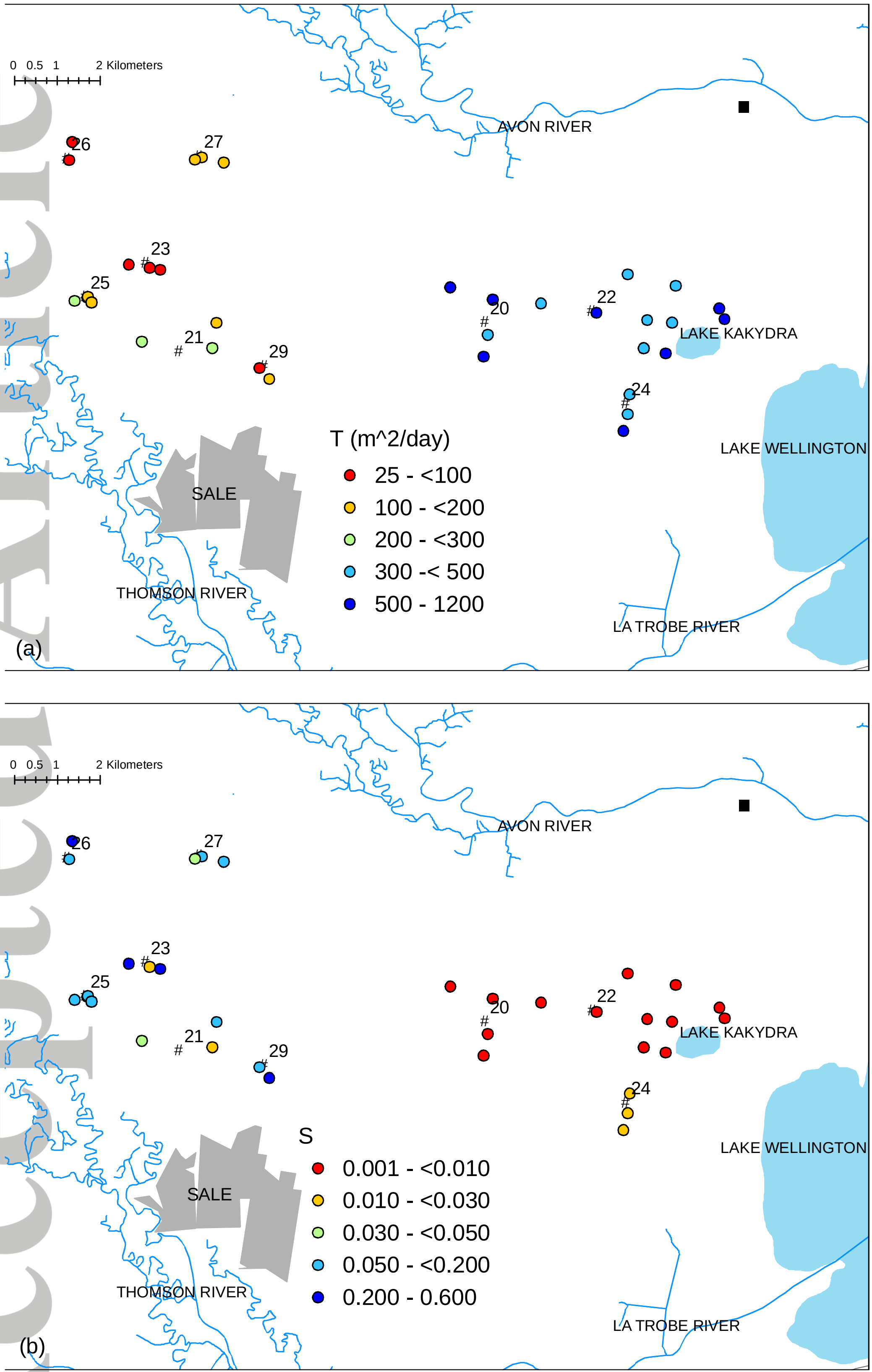

Figure 3: Aquifer property estimates for (a) transmissivity, and (b) storativity. The colored circles are observation bores and the numbered triangles are pumping bores.

Accepted Article

This article is protected by copyright. All rights reserved.

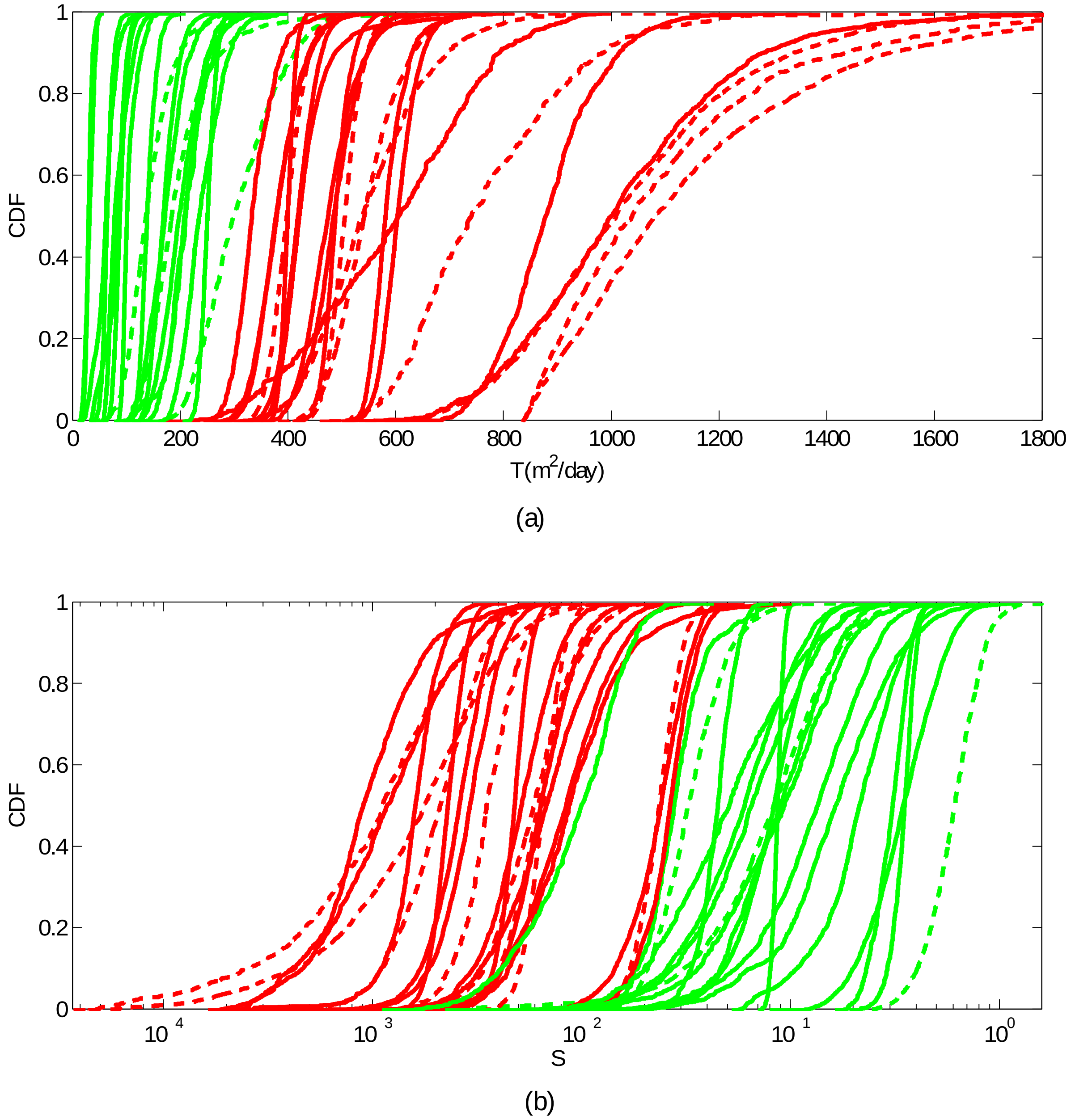

Figure 4: The Cumulative Distribution Function (CDF) for (a) transmissivity and (b) storativity uncertainty estimates from DREAM. Note that the CDFs for observation bores on the west side of the area are colored green and for the east side are colored red. In addition, for those where the pumping signal is weak (i.e. pumping drawdown is <0.7 m), the CDF is dashed.

This article is protected by copyright. All rights reserved.

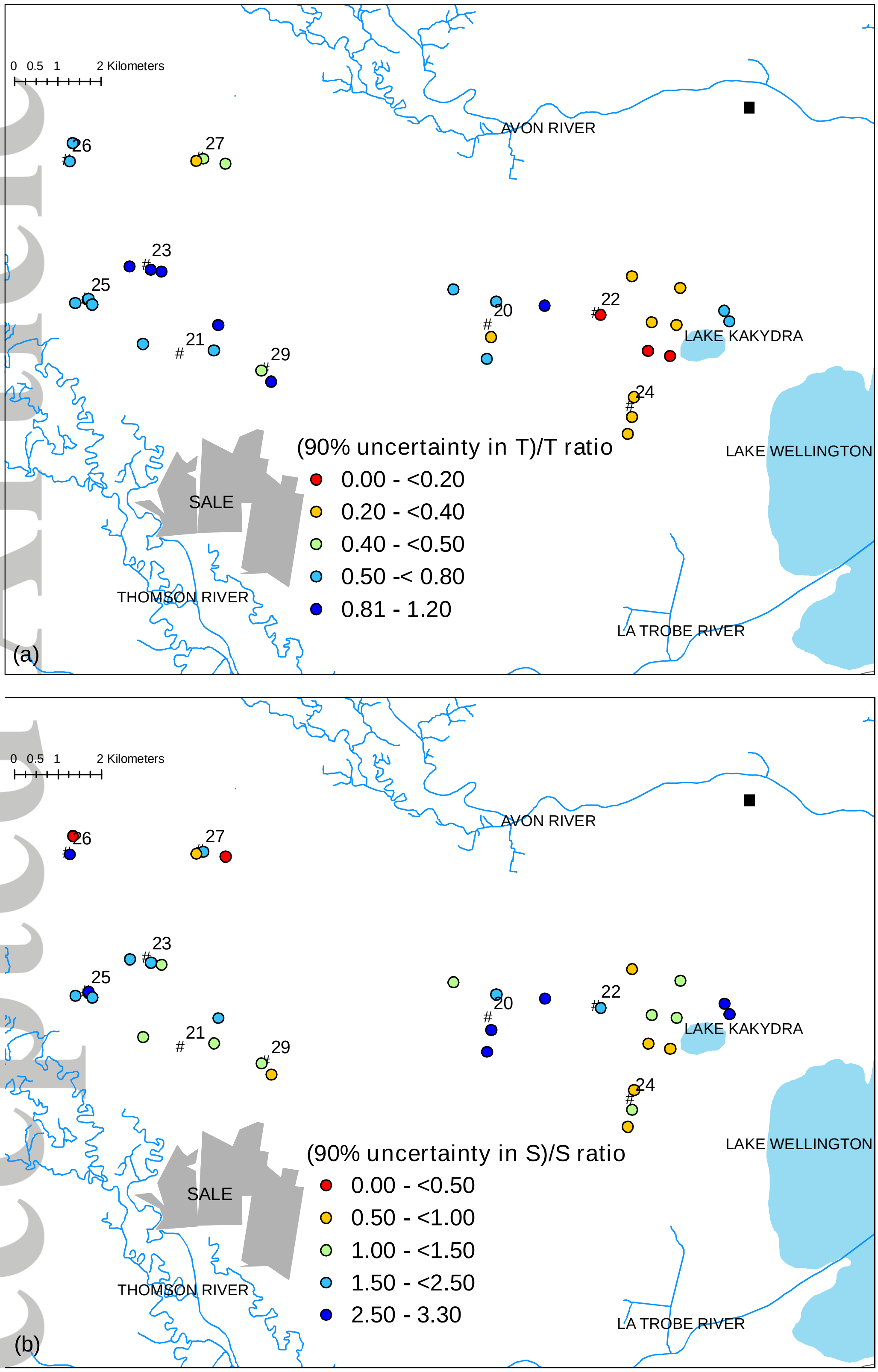

Figure 5: (a) The ratio of the 90% uncertainty range of transmissivity to the median transmissivity value at each bore, (b) The ratio of the 90% uncertainty range of storativity to the median storativity value at each bore.

Accepted Article

This article is protected by copyright. All rights reserved.

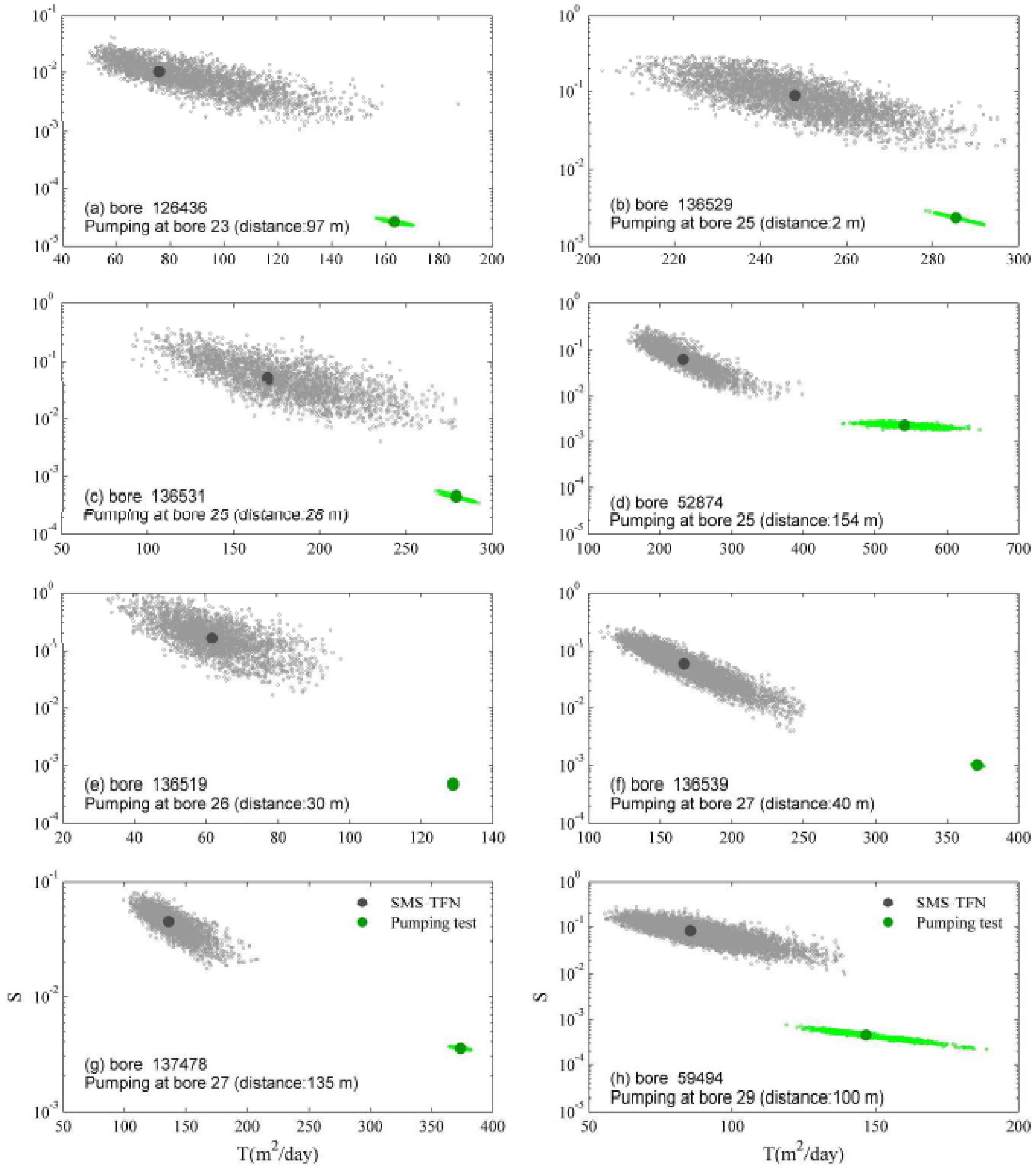

Figure 6: Scatter plots showing the distributions of T and S values from the SMS-TFN model and the short term pumping test results derived using DREAM. The grey and green open circles denote the SMS-TFN and pump test posterior distribution estimates respectively. The median is denoted by solid large circle.

This article is protected by copyright. All rights reserved.

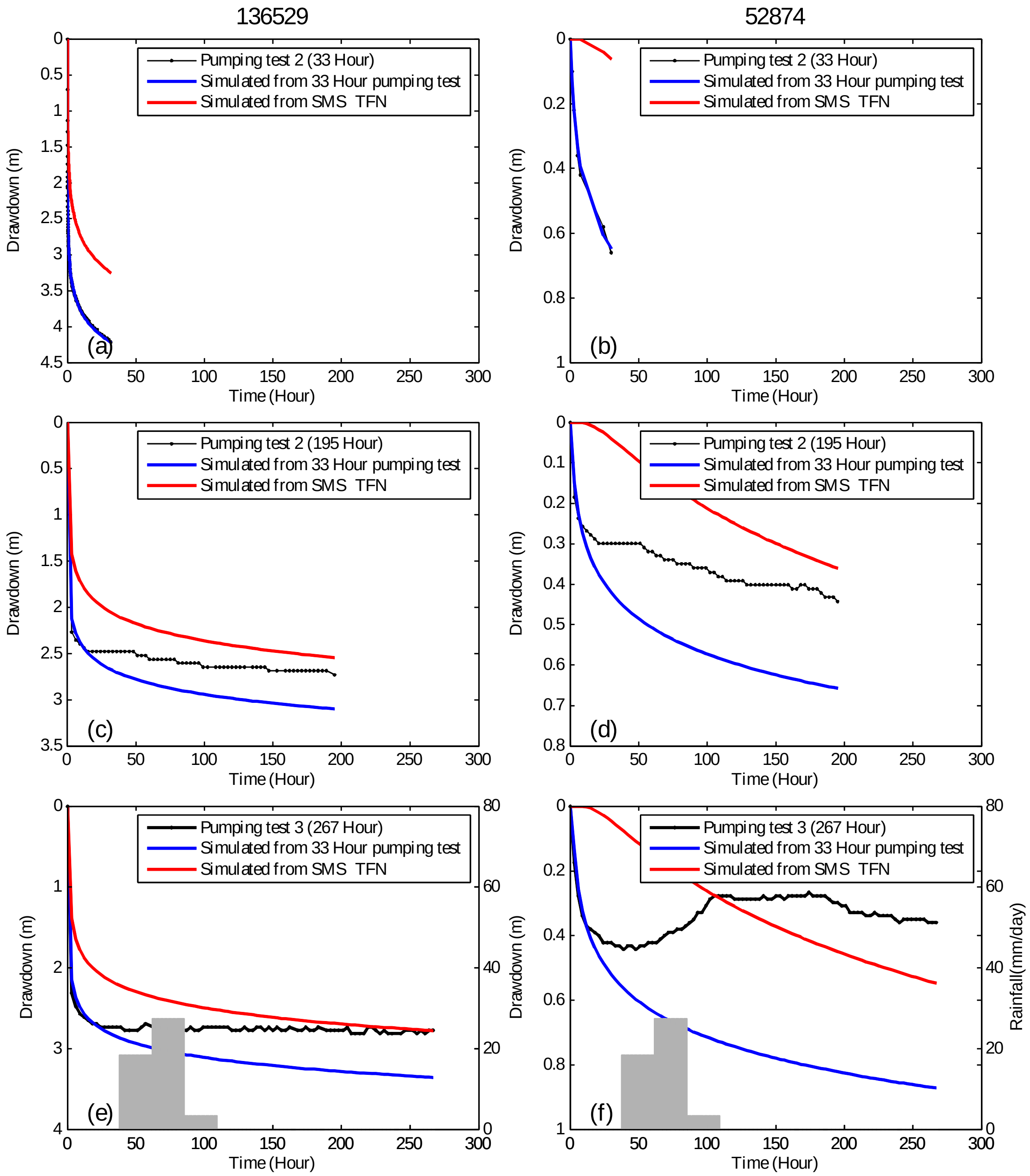

Figure 7: Pumping test data (black line) for bores 136529 and 52874. The red line denotes the simulated pumping influence with the aquifer properties obtained from SMS-TFN model. The blue line is the simulated pumping influence with the aquifer properties obtained from the 33 hour pumping test. Rainfall during the pumping test is shown as the grey columns in (e) and (f).

This article is protected by copyright. All rights reserved.

Accepted Article

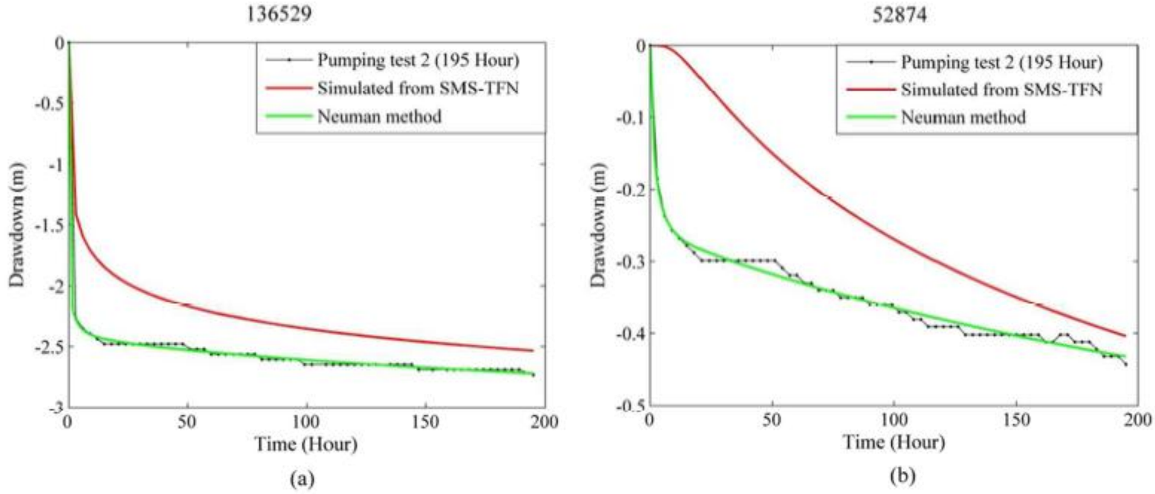

Figure 8: Pumping simulations using the Neuman (1975) method and the SMS-TFN model for the 195 hour pumping tests.

This article is protected by copyright. All rights reserved.