Hydrogeology Journal
DOI 10.1007/s10040-015-1269-7

Decomposing groundwater head variations into meteorological and pumping components: a synthetic study

V. Shapoori · T. J. Peterson · A. W. Western · J. F. Costelloe

Abstract Time-series modeling is often used to decompose groundwater hydrographs into individual drivers such as pumping and meteorological factors. To date, there has been an assumption that a simulation fitting the total hydrograph produces reliable estimates of the impact from each driver. That is, assessment of the decomposition has not used an independent estimate of each decomposition result. To begin to address this, a synthetic study is undertaken so that the impact of each driver is known. In this study, 500 MODFLOW groundwater models of a one-layer unconfined aquifer were constructed. For each model, three hydrogeological properties (saturated hydraulic conductivity, storativity and depth to aquifer basement), the distance between observation and pumping bores, and extraction rate were set randomly and synthetic groundwater hydrographs were derived. For each hydrograph, the influence of individual drivers was estimated using six different time-series models. These estimates were then compared to the known meteorological and pumping influences derived from the MODFLOW models. The results demonstrate that hydrograph separations obtained from time-series models do not always result in reliable estimation of pumping and meteorological influences even when the overall hydrograph fit is good. However, when the time-series model represents the important processes (e.g. phreatic evaporation is included for shallow water tables) and the (head) variance of the pumping signal to the meteorological signal is between 0.1 and 10, the time-series model has the potential to adequately separate the influence of pumping and climate.

Keywords Statistical modeling · Groundwater pumping · Time series modeling · Australia

Received: 29 October 2014 / Accepted: 6 May 2015

V. Shapoori (✉) · T. J. Peterson · A. W. Western · J. F. Costelloe
Department of Infrastructure Engineering,
The University of Melbourne, Parkville, Vic 3010, Australia
e-mail: shapoori@student.unimelb.edu.au

Introduction

The dynamics of unconfined groundwater levels are usually the result of numerous and interacting factors, including those arising from climate variability (e.g. changes in recharge and discharge, phreatic evapotranspiration) and those arising from anthropogenic causes (e.g. land cover change and groundwater pumping; Doll 2009; Shamsudduha et al. 2011). Over the last 50 years, groundwater levels have declined dramatically in many arid and semi-arid areas as the result of over-extraction and changes in rainfall patterns (Konikow and Kendy 2005; Kundzewicz and Doll 2009; Sophocleous 2003; Tularam and Krishna 2009; Zektser et al. 2005). To support water-resource-management plans in such areas, estimating the impact from pumping and separating its influence on groundwater head variation from other factors (i.e. mainly meteorological influences) is highly significant. Recently, time-series modeling has been introduced as a new tool to simulate those fluctuations and further quantify the influence of individual drivers (Obergfell et al. 2013; Peterson and Western 2011; Von Asmuth et al. 2002, 2008; Yihdego and Webb 2011). This involves first fitting the time-series model to a groundwater hydrograph and then using the calibrated model to decompose the hydrograph and quantify the influence of each driver; however, to date, there has been an assumption that a simulation fitting the total hydrograph produces reliable estimates of the impact from each driver. This is a weak falsification test for the time-series decomposition, whereas a more rigorous method would use an independent estimate of the impacts from each individual driver. Considering that in the field such data are almost never available, in this study a synthetic approach is adopted, whereby hydrographs are generated with known impacts from pumping and meteorological factors and then they are analyzed and decomposed using time-series models. The robustness of this approach is demonstrated by using synthetically derived groundwater hydrographs and a wide range of possible time-series models.

Time-series analysis is a convenient and strongly data driven approach that can play a substantial role in investigating the effects of climate and human interventions in groundwater head fluctuations. To date, a variety

Published online: 22 May 2015

of studies have used a time-series approach to quantify recharge estimates (Andreu et al. 2011; Crosbie et al. 2005; Cuthbert 2010; Scanlon et al. 2002; Viswanathan 1984) and vegetation consumption of groundwater (Butler et al. 2007; Gerla 1992; Loheide et al. 2005; Rosenberry and Winter 1997; White 1932). The time-series approach has been also adopted specifically for modeling groundwater levels and predicting the groundwater level hydrograph under different scenarios. With respect to this application, the specific time-series technique employed ranges from simple—such as the linear regression hydrograph analysis methodology (HARTT; Ferdowsian et al. 2002)—to the more sophisticated such as transfer function noise models (Bakker et al. 2008; Peterson and Western 2011, 2014; Von Asmuth et al. 2002, 2008; Yi and Lee 2004; Yihdego and Webb 2011). Overall, a transfer function noise (TFN) model simulates an observed output at a given point in time as a weighting of recent input forcing data (i.e. the transfer function) plus a correlation term for the observed output not explained by the forcing data (i.e. the noise). Von Asmuth et al. (2002) proposed a continuous form of TFN model which basically simulates the groundwater level as the linear combination of weighted past rainfall, potential evaporation and pumping. Peterson and Western (2014) extended the TFN model developed by Von Asmuth et al. (2002) to account for nonlinear vadose zone processes. This was achieved by inclusion of a parsimonious, vertically lumped, soil moisture model in the Von Asmuth et al. (2002) model structure.

While in some studies, time-series models were used to simulate the groundwater level influenced by climatic drivers only (Lehsten et al. 2011; Manzione et al. 2010; Siriwardena et al. 2011), other studies have also applied these models specifically to quantify or separate the influence of individual drivers, including pumping and meteorological factors (Obergfell et al. 2013; Shapoori et al. 2015; Von Asmuth et al. 2008) and weather and land use (Peterson and Western 2011; Yihdego and Webb 2011). In particular, Obergfell et al. (2013) applied the model initially proposed by Von Asmuth et al. (2002) to eleven observation bores around a well field consisting of seven pumping bores. Given that the model performed very well in hydrograph simulation for all bores in that study, the modeled hydrograph separations (meteorological influences as well as influence of seven pumping bores) were assumed to represent the true influences and no further assessment was undertaken to evaluate the reliability of those separations. In other studies—e.g. Peterson and Western (2011) and Yihdego and Webb (2011)—a similar approach was also undertaken to decompose groundwater hydrographs without any evaluation of subsequent hydrograph separations. Although good simulation of an observed groundwater hydrograph is a prerequisite for plausible decomposition, it is a weak falsification test for adequate decomposition. In other words, simply fitting an observed hydrograph provides no guarantee that the separation into major driver impacts is accurate.

This challenge is similar to that in rainfall-runoff modeling, whereby a calibrated model may accurately simulate total streamflow but inaccurately estimate subcomponents of the streamflow such as baseflow, or internal catchments processes such as soil moisture (Jakeman and Hornberger 1993; Sivapalan and Young 2005; Young 1978). To assess the adequacy of these subfluxes and processes, synthetic studies have been undertaken, whereby a 'true' system is simulated using a complex, physically based, distributed, numerical model. A simpler model is then calibrated to the synthetic total streamflow and its internal fluxes and dynamics are evaluated against that from the synthetic model (Ferket et al. 2010; Li et al. 2013, 2014; Partington et al. 2012, 2011; Szilagyi 2004). For instance, in Li et al. (2013) and (2014), the performances of recursive digital filter (RDF) methods (simpler models) commonly used to estimate baseflow were assessed using a fully integrated surface water/groundwater model. In doing so, the synthetic (i.e. known) outflow and baseflow hydrographs were generated from the fully integrated model with different types of hydrogeology and forcing conditions. The synthetic outflow hydrograph was then used as the input to the RDF techniques to simulate the filtered baseflow hydrograph and provide a basis for evaluating the accuracy of the RDF technique.

In this study, a similar synthetic approach was adopted and the main goal is to assess whether groundwater-time-series models can reliably decompose a groundwater hydrograph to the major drivers. Ideally, to investigate the adequacy of groundwater-time-series decomposition, many synthetic catchment models with different types of structure and/or boundary conditions are required. However, as a proof of concept, this paper focuses on one type of catchment—an upland valley unconfined aquifer. As demonstrated, the rigorous analysis of this synthetic experiment is not a trivial exercise, even for one catchment type. In constructing the synthetic environment, 500 aquifer models were created using the MODFLOW code. All models simulated an irrigation region supplied exclusively by time-varying groundwater extraction from a single layer unconfined aquifer. Each model differed in that randomly chosen aquifer hydraulic properties, depth to aquifer basement, pumping rate, and distance between observation and pumping bore were adopted. These randomly chosen catchment variables provide a wide range of scenarios (e.g. different magnitude of pumping signal relative to climatic signal within shallow or deep groundwater level situations). Next, transient MODFLOW simulations were undertaken with and without groundwater pumping, and a synthetic hydrograph was extracted for each simulation, with the subtraction of the two providing the 'true' impact from the pumping (and associated irrigation). The forcing data and synthetic hydrographs with groundwater pumping impact were then used as the input to the time-series models to calibrate time-series-model parameters and simulate the hydrographs. The simulated hydrographs from the time-series model were then decomposed into the impacts from meteorological

DOI 10.1007/s10040-015-1269-7

factors and pumping and evaluated against the 'true' impacts from MODFLOW.

Methods

In this section, first, the MODFLOW models used to generate the synthetic hydrographs are described. Next, the linear and nonlinear time-series models are described, followed by the implementation of the time series models with the synthetically derived groundwater-level time series. The head uncertainty and decomposition uncertainty estimates are also explored for examples of shallow and deep groundwater situations using DREAM uncertainty analysis (Vrugt et al. 2008, 2009). The DREAM analysis provides only the parameter uncertainty; to estimate the total uncertainty, the remaining error needs to be added to parameter uncertainty bounds (for further detail of the application of the DREAM method, see Shapoori et al. 2015).

Synthetic MODFLOW models

To assess the time-series models, 500 MODFLOW groundwater models were constructed with a single-layer unconfined aquifer. Figure 1 shows the geometry of the MODFLOW models. All models consisted of a simple rectangular catchment of 5 km×10 km, cell resolution of 100×100 m (which reduced to 50×50 m for the 10 rows and columns closest to the centre of catchment), 0.4 % valley side slope and a 0.2 % slope along the valley, a constant head boundary condition at the downstream end of the catchment, and no flow boundaries on the sides and top of the catchment. The evapotranspiration extinction depth was set at 1.5 m. Within each synthetic MODFLOW model, there was one groundwater pump in the middle of the catchment servicing an irrigated pasture with the maximum extent of 1.2 km × 1.2 km (see Fig. 1). Groundwater supplied irrigation occurred for 6-month periods, from October up to March of the following year (i.e. southern hemisphere summer) and the extracted volume for each month was calculated as a function of the irrigation demand deficit, D_m [LT^{-1}].

$$D_m = \max(0, E_m - P_m) \quad (1)$$

where P_m and E_m [LT^{-1}] are monthly rainfall and potential evapotranspiration. The extracted volume was then calculated as follows:

$$Q_m = I_c A_{max} D_m \quad (2)$$

Here I_c [−] is the fraction of irrigated land from a maximum area, A_{max} [L^2], of 1.2 km ×1.2 km and Q_m [L^3T^{-1}] is the monthly pumping rate. Estimating the pumping rate based on Eq. (1) allows the pumping to be a function of climatic demand, which can result in correlation between the input drivers (e.g. pumping and recharge). Considering that this correlation between the drivers is realistic and would occur often in agricultural

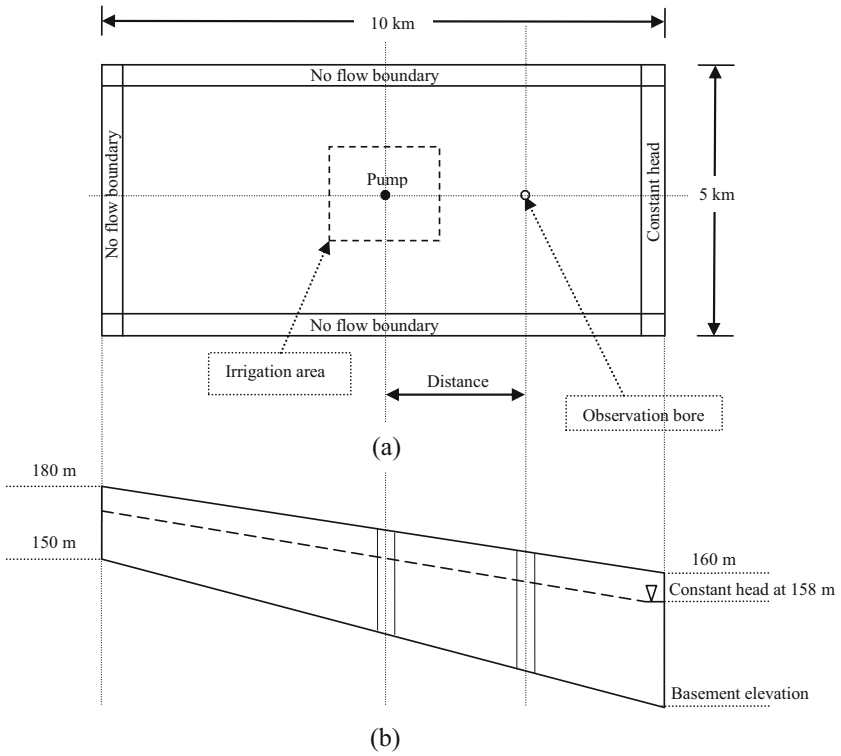

Fig. 1 The conceptual catchment dimensions: **a** plan view and **b** side view. Elevations in meters are relative to an arbitrary base level of 0 m

 DOI 10.1007/s10040-015-1269-7

catchments, such a highly correlated environment provides a promising basis to test the time-series model in a generally realistic situation.

Within the irrigation area, rainfall and irrigation depths were summed and, therefore, irrigation effects were included in the recharge from the irrigation area. In this case, pumping causes two anthropogenic influences—pumping and irrigation recharge. Considering that on average, 20 % of irrigation becomes recharge, between the pumping and irrigation recharge, pumping is the dominant factor; thus, hereafter, this driver is referred as simply 'pumping'.

In generating the 500 models, the following model properties were randomly chosen from a uniform distribution: saturated horizontal hydraulic conductivity, specific yield, distance of observation from pumping bore, elevation of the aquifer basement at the downstream end of the catchment and the fraction of irrigated land. The sampling of the five random properties was undertaken using Latin Hypercube sampling from uniform distributions within the ranges shown in Table 1. The aquifer parameters ranges were chosen to simulate unconsolidated gravel to sandy aquifers, as specified by Freeze and Cherry (1979).

In this study, the recharge was simulated as a function of soil moisture by adoption of the unsaturated-zone-flow MODFLOW code (UZF1; Niswonger et al. 2006). To simulate unsaturated flow, UZF1 simplifies Richard's equation to vertical flow through a homogeneous unsaturated zone driven only by gravitational potential gradients and not suction gradients. In using UZF1, three additional parameters were introduced, specifically, the saturated water content, the Brooks-Corey exponent and the soil vertical saturated hydraulic conductivity. The saturated water content and the Brooks-Corey exponent parameters were fixed at 0.3 and 3.5 respectively to simulate unconsolidated sand (Brooks and Corey 1964). A sandy soil profile, while unusual for an irrigated region, was adopted because the soil vertical saturated hydraulic conductivity was set to a fixed fraction (20 %) of the randomly sampled aquifer saturated hydraulic conductivity. One consequence of using a sandy profile was that most rainfall infiltrated and runoff was minimal.

Table 1 The range of physical characteristics of the catchment and aquifer properties in the MODFLOW model. *Min* minimum; *Max* maximum

Parameter	Unit	Min value	Max value
Saturated hydraulic conductivity	m/day	1	30
Specific yield	–	0.05	0.4
Bottom level of aquifer at the end of the catchment	m	0	130
Distance between pumping and observation[a]	m	0	1,500
Fraction of land used for irrigation	–	0	1

[a] The location of the observation bore was discretized to the centre of the MODFLOW grid cell

With regard to the meteorological forcing, observed data from the Australian Bureau of Meteorology rainfall station 081049 at the town of Tatura was adopted. This area is located in an irrigation district of south-eastern Australia (longitude = 145.27 °E and latitude=36.44 °S) and the climate is classified as Cfb (i.e. temperate climate without dry season and warm summer) under the Köppen–Geiger climate classification system (Peel et al. 2007). The potential evapotranspiration was derived from Morton's complementary relationship areal model (Morton 1983). The input data for the Morton model are daily maximum and minimum temperature, vapor pressure and net solar radiation. The daily maximum and minimum temperature, and vapor pressure were obtained from the Australian Water Availability Project (AWAP) database and net solar radiation was estimated following empirical equations provided in FAO56 (Allen et al. 1998) for the closest grid cell to the rainfall station. These data were then used as the input to the Morton model and the potential evapotranspiration was estimated accordingly. The same potential evapotranspiration rates were applied across the entire model domain as the vegetation within and outside the irrigated area was assumed to be pasture. Figure 2 presents the monthly precipitation and potential evapotranspiration based on the data from 1942 to 2000. Average annual rainfall and potential evapotranspiration is 485 and 1,094 mm/year respectively. The rainfall is almost uniformly distributed throughout the year with a small rise during the winter and early spring. Potential evapotranspiration is summer dominant and the monthly average potential evapotranspiration varies from 28 mm/month in June up to 168 mm/month during January.

Finally, the daily precipitation and potential evapotranspiration were aggregated into monthly rates, and then specified as the infiltration (with any irrigation added to rainfall) and evapotranspiration rates to the top of the unsaturated zone in the UZF1 package. Monthly meteorological forcing data were adopted for computational efficiency and may result in a lower recharge rate than if daily forcing data were adopted. In running each model, the steady-state heads were first derived using the long-term mean precipitation and evapotranspiration rates (no pumping was simulated). Next, the steady-state heads were used as initial conditions for a 20-year transient run that extended from January 1980 to December 1999. The pumping started in October 1989 and continued to the end of the simulation. In order to quantify the drawdown from the extraction bore and the meteorological influence, the pump was removed and MODFLOW transient solutions were rederived using the same steady-state solution for the initial conditions. It should be mentioned that in the model simulation without pumping, there is no irrigation within the irrigation area. The derived hydrograph with no pumping represents the meteorological influence (i.e. only using meteorological forcing data) and the difference between it and the hydrograph with pumping gives the drawdown from pumping.

DOI 10.1007/s10040-015-1269-7

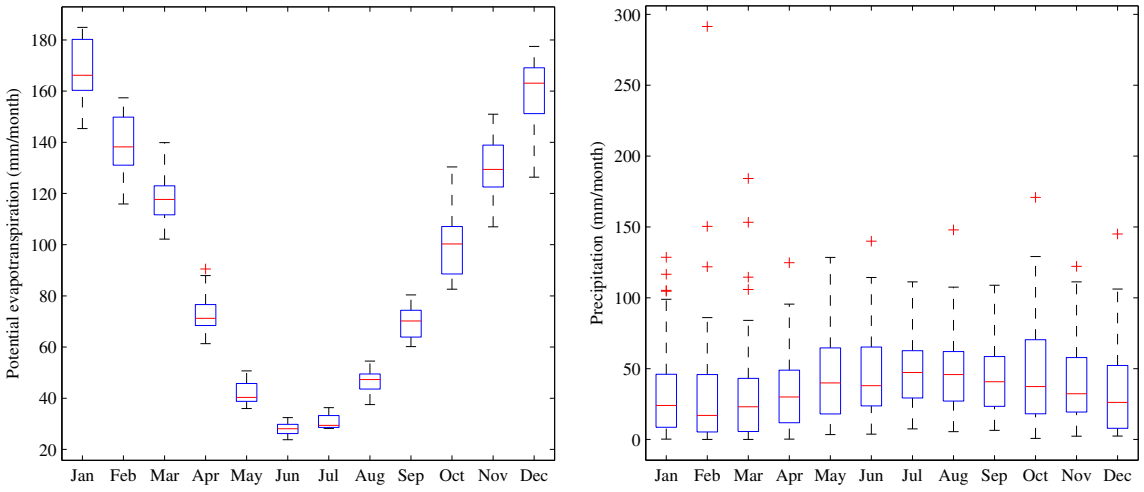

Fig. 2 Box plots of monthly **a** potential evapotranspiration and **b** precipitation. The *whiskers* extend a maximum of 1.5 times the inter-quantile range. Values beyond that are assumed as outliers

Synthetic groundwater hydrographs

For each MODFLOW transient simulation, a groundwater hydrograph was extracted at a randomly generated distance from the extraction bore; however, to provide a realistic assessment of the time-series models, forcing and observation errors were introduced. Past studies have adopted a log-normal multiplicative errors model for precipitation (Kavetski et al. 2006; McMillan et al. 2011; Renard et al. 2010). Following such studies, the daily rainfall P_t [LT^{-1}], for the time-series models was generated by corrupting the rainfall input, $\widetilde{P}_t$ [LT^{-1}], to the MODFLOW model with temporally independent, unbiased, normal distribution of errors, e_1 [−], having a standard deviation of 20 %:

$$P_t = \frac{\widetilde{P}_t}{\exp(e_1)} \tag{3}$$

$$e_1 \sim N\left(0,\ 0.2^2\right) \tag{4}$$

While the evapotranspiration was assumed free of errors, the monthly pumping rate input to the MODLFOW model, $\widetilde{Q}_t$ [L^3T^{-1}], was corrupted using an unbiased normal multiplicative error, e_2 [−], with standard deviation of 0.1 to give the extraction rate input, Q_t [L^3T^{-1}], to the time-series models:

$$Q_t = \widetilde{Q}_t(1 + e_2) \tag{5}$$

$$e_2 \sim N\left(0,\ 0.1^2\right) \tag{6}$$

The error in the groundwater hydrograph was a function of the depth to the groundwater. By replacing the depth to water level with the head, the observed groundwater head ,$\widetilde{h}_t$ [L], was corrupted using an unbiased normal distribution, e_3 [−], with standard deviation of 0.02 times the depth below natural surface,B_t [L]:

$$h_t = \widetilde{h}_t - B_t e_3 \tag{7}$$

$$e_3 \sim N\left(0,\ 0.02^2\right) \tag{8}$$

Time-series model

Linear transfer function noise model

In the original TFN model, an observed univariate time series is simulated as the linear sum of three main components. The three components are a deterministic transformation of an input time series, a noise term and local drainage level (Eq. 6). Von Asmuth et al. (2002) adopted this approach to simulate an observed groundwater-level elevation as the dynamic relationship between the input contribution to groundwater variation (h_t^*) and input forcing (e.g. meteorological factors) data (Eq. 10):

$$h_t = h_t^* + r_t + d \tag{9}$$

$$h_t^* = \sum_{i=1}^{m} \left(\int_{-\infty}^{t} R_i(\tau)\theta_i(t-\tau)d\tau \right) \tag{10}$$

where h_t [L] is the simulated groundwater level at time step t; h_t^* [L] is the contribution to the groundwater level at time step t attributed to the combination of all m stressors; r_t [L] is the residual series; d [L] is the local

DOI 10.1007/s10040-015-1269-7

drainage level; i [−] indexes individual stressors; θ_i [−] is the weighting or impulse response function of stress i; and R_i [L] is the value of stress i at time step t.

The most important part of the transfer noise function is the impulse response function (θ). It specifies the response of groundwater head to an instantaneous change in the stressor such as precipitation. The Pearson type III distribution function was adopted by Von Asmuth et al. (2002) for the precipitation response function; however, Peterson and Western (2014) have identified five weaknesses in the use of the original Pearson type III distribution in time-series models. Following Peterson and Western (2014), a robust version of Pearson type III distribution was selected in this study (Eqs. 7 and 12) to represent the impulse response function of climatic stresses (θ_p).

When exp (z) – 1≥1:

$$\theta_p(t) = A\frac{t^{\exp(z)-2}\exp(-bt)}{\left(\frac{\exp(z)-2}{b}\right)^{\exp(z)-2}\exp(2-\exp(z))} \tag{11}$$

When exp (z) – 1<1:

$$\theta_p(t) = A\frac{t^{\exp(z)-2}\exp(-bt)-f_{\text{limit}}}{1-f_{\text{limit}}} \tag{12}$$

$$f_{\text{limit}} = t_{\text{limit}}{}^{\exp(z)-2}\exp(-bt_{\text{limit}}) \tag{13}$$

$$t_{\text{limit}} = \min(t)-100\times 365 \tag{14}$$

where A, b, z are parameters and min(t) is the first meteorological observation date; t_{limit} is 100 years prior to the first meteorological observation date and f_{limit} is the distribution value at 100 years prior to the first meteorological observation. The parameter A acts as a scalar to transform the stress into the groundwater level time series. This version of Pearson type III produces the same response shape as the original one but minimizes the covariance between the parameters and improves the calibration efficiency (for more details see Peterson and Western 2014).

For pumping stresses, Von Asmuth et al. (2008) adopted Hantush's well formula (Hantush 1956) for an aquifer underlying a storage-free aquitard (i.e. a leaky aquifer). The response function (θ_H) was defined by the time derivative of the step response function (Eq. 8):

$$\theta_H(t) = \frac{\alpha}{t}\exp\left(-\frac{\beta^2}{t}-\gamma^2 t\right) \tag{15}$$

where α, β and γ are parameters that will only have physical meaning if the basic Hantush assumptions are satisfied. In this paper, Hantush's response function was one option adopted for simulating the pumping drawdown. The reasoning for simulating leakage within an unconfined aquifer is that, like pumping within a leaky confined aquifer causing inflow, preliminary analysis indicated that within the MODFLOW model, increased pumping can cause a similar response by reducing groundwater evaporation and, hence, increasing net recharge. However, to rigorously assess this mechanism, the leakage term from Eq. (8) was omitted and the simplest version of Hanstush's well formula, known as Ferris and Knowles' well formula (Ferris and Knowles 1963), was also investigated (Eq. 9).

$$\theta_F(t) = \frac{\alpha}{t}\exp\left(-\frac{\beta^2}{t}\right) \tag{16}$$

Finally, Eq. (10) details a linear TFN model comprising of both meteorological forcing and groundwater pumping. The first and second integrals are the effect of precipitation and evapotranspiration and the third integral is the pumping effect, in which the transfer function, $\theta_{H\|F}$, can be set to either Hantush's (Eq. 8) or Ferris and Knowles' (Eq. 9) well formulas; henceforth, it is referred to as the '*Von Asmuth 2008 model (VA) with Hantush's (H) well formula*' or the ' *Von Asmuth 2008 (VA) model with Ferris and Knowles' (FK) well formula* '.

$$h_t = \int_{-\infty}^{t} P_\tau\theta_p(t-\tau)d\tau - \int_{-\infty}^{t} f_E E_\tau\theta_p(t-\tau)d\tau - \int_{-\infty}^{t} Q_\tau\theta_{H\|F}(t-\tau)d\tau + r_t + d \tag{17}$$

where P_τ [LT^{-1}] is the daily precipitation; E_τ [LT^{-1}] is the daily potential evapotranspiration; f_E [−] is a dimensionless parameter scaling the transfer function for application to the evapotranspiration signal; and Q_τ [L^3T^{-1}] is the daily rate of pumping. Note that the evapotranspiration (ET) component (i.e. the second integral in Eq. 10) represents the head loss due to soil and plant evapotranspiration of groundwater; however, this ET is simulated as independent of catchment wetness. Peterson and Western (2014) found that to adequately simulate long-term head decline, the groundwater ET (GET) should be inversely proportional to the catchment wetness; that is, GET increases when the soil moisture is low. Further detail pertaining to Peterson and Western (2014) study is provided in the next section.

Nonlinear transfer function noise model

The linear TFN model adopts a linear relationship between precipitation and head; hence, if rainfall doubles from 1 day to the next then the head response will also double. However, because large rainfall events produce runoff and groundwater recharge often occurs only when the soil is wet, the relationship between precipitation and

DOI 10.1007/s10040-015-1269-7

head is nonlinear. To parsimoniously capture these nonlinearities, Peterson and Western (2014) added a nonlinear filter to the Von Asmuth 2008 model (see section 'Linear transfer function noise model') consisting of a vertically lumped soil moisture model. Depending upon the dynamic sought, the soil moisture model can have between one and five parameters. Siriwardena et al. (2011) tested all of the models on 620 bores from across Victoria, Australia and found the following soil model was the most parsimonious:

$$\frac{dS}{dt} = P\left(1-\frac{S}{S_{\text{cap}}}\right)-E\left(\frac{S}{S_{\text{cap}}}\right) \tag{18}$$

In Eq. (11), S_{cap} [L] is a parameter for soil-moisture-storage capacity; S [L] is a state variable for the soil moisture at time t; P [LT^{-1}] is the rate of precipitation and E [LT^{-1}] is the potential evapotranspiration rate. Following Peterson and Western (2014), to incorporate the soil moisture component into the continuous TFN model, two modifications were made to the transfer function model of Eq. (10). Firstly, the precipitation term, P_τ, was replaced by $\left(\frac{S_\tau}{S_{\text{cap}}}\right)^\beta$ where β is a parameter controlling the responsiveness of recharge to soil moisture. This parameter (β) is similar to the Campbell or Brooks and Corey pore index. The influence of soil evapotranspiration is included in Eq. (11); however, in areas with a shallow water table, groundwater evapotranspiration can also occur. To simulate this effect, Peterson and Western (2014) estimated the groundwater evapotranspiration as $E_\tau\left(1-\frac{S_\tau}{S_{\text{cap}}}\right)$, which is the residual potential evapotranspiration remaining after evapotranspiration from the soil. This flux estimate replaced the potential evapotranspiration, E_τ, in Eq. (10) and a unique evapotranspiration transfer function, θ_{E}, based on Eqs. (7) and (12) was adopted. Peterson and Western (2014) tested this model for a dry-land area in Victoria, Australia, where groundwater pumping does not occur. They concluded that this groundwater evapotranspiration term is essential for the simulation of periods of long-term groundwater level decline; however, by coupling the model with a pumping component, preliminary analysis in this study showed that the groundwater evapotranspiration term can compensate for other influences (e.g. pumping effect), making the hydrograph decomposition very challenging. Thus, to further identify any possible internal compensation, the nonlinear time-series model was tested both with and without the groundwater evapotranspiration term.

In summary, Eqs. (12) and (13) detail the nonlinear TFN models with and without groundwater evapotranspiration respectively. The two left-most integrals in Eq. (12) show the free drainage (recharge estimate) and groundwater evapotranspiration. The third integral is the pumping component, in which similarly the transfer function, $\theta_{\text{H||F}}$, can be set to either Hantush's (Eq. 8) or Ferris and Knowles' (Eq. 9) well formulas; henceforth this model is referred to as the '*Soil moisture transfer function noise model (SMS-TFN) with groundwater evapotranspiration (GET) and Hantush's (H) well formula*' or the '*Soil moisture transfer function noise model (SMS-TFN) with groundwater evapotranspiration (GET) and Ferris and Knowles' (FK) well formula* '.

$$h_t = \int_{-\infty}^{t}\left(\frac{s_\tau}{s_{\text{cap}}}\right)^\beta \theta_{\text{p}}(t-\tau)d\tau - \int_{-\infty}^{t} E_\tau\left(1-\frac{s_\tau}{s_{\text{cap}}}\right)\theta_{\text{E}}(t-\tau)d\tau - \int_{-\infty}^{t} Q_\tau \theta_{\text{H}||\text{F}}(t-\tau)d\tau + r_t + d \tag{19}$$

In Eq. (13), all components are the same and the only difference is the omission of the groundwater evapotranspiration component; henceforth this model is referred to as the '*Soil moisture transfer function noise model (SMS-TFN) with Hantush's (H) well formula'* or the '*Soil moisture transfer function noise model (SMS-TFN) with Ferris and Knowles' (FK) well formula*'.

$$h_t = \int_{-\infty}^{t}\left(\frac{s_\tau}{s_{\text{cap}}}\right)^\beta \theta_p(t-\tau)d\tau - \int_{-\infty}^{t} Q_\tau \theta_{\text{H}||\text{F}}(t-\tau)d\tau + r_t + d \tag{20}$$

Calibration and implementation of time-series model

In applying each time-series model, a split-sample calibration-evaluation approach was adopted, whereby the first 10 years of data (1980–1989) were removed to minimize the impact of initial conditions from the MODFLOW models, and the time-series model was calibrated to the following 7 years of data (1990–1996) and evaluated on the remaining 3 years data (1997–1999). Following Von Asmuth et al. (2002) in calibrating the time-series model, the residuals at each time point were calculated as:

$$\tilde{r}_t = h_{\text{obs},t} - h_{\text{mod},t} \tag{21}$$

where $\tilde{r}_t$ [L] is the residual at time t; $h_{\text{obs},t}$ [L] is the observed groundwater level at time t; and $h_{\text{mod},t}$ [L] is the modeled groundwater level at time t. Next, the innovation series were calculated by subtraction of the exponential noise components of previous time steps from the current time step residuals:

$$\upsilon_t = \check{r}_t - e^{-\mu\Delta t}\widetilde{r_{t-\Delta t}} \tag{22}$$

Here, v_t [L] is the innovation at time t; $\widetilde{r_{t-\Delta t}}$ [L] is the residual at the previous time; Δt [T] is the time step; and μ is a time-series model parameter defining the decay rate of the noise.

In calibrating the parameters, a weighted least squares objective function was adopted whereby v_t^2 was weighted by the water-level-time-step size. The weighted least

DOI 10.1007/s10040-015-1269-7

squares objective function was then minimized with respect to the parameters using a multi-start, trust-region Levenberg-Marquardt gradient based algorithm (Fan and Pan 2006; Levenberg 1944; Marquardt 1963).

In assessing the performance of the models in simulating the groundwater hydrographs and quantifying the influence of individual drivers (i.e. pumping and climatic effects), the Nash-Sutcliffe efficiency (NSE; Nash and Sutcliffe 1970) was used.

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n} \left(D_{\mathrm{mod}}^{i} - D_{\mathrm{obs}}^{i}\right)^2}{\sum_{i=1}^{n} \left(D_{\mathrm{obs}}^{i} - \overline{D_{\mathrm{obs}}}\right)^2} \tag{23}$$

Here, D_{mod} [L] and D_{obs} [L] denote either predicted and observed groundwater head or predicted and observed driver's influence (pumping or meteorological effect). The range of NSE is from 1 to $-\infty$. An NSE of 1 means a perfect fit between modeled and observed hydrograph, while NSE values of less than zero indicate that the average of the observed hydrograph is better than the model simulation.

Results

In this section, the results of applying six time-series models—i.e. the combinations of the original TFN model and the SMS-TFN models with the H and the FK pumping transfer models—to the 500 synthetic hydrographs are shown. The performance statistics (NSE) were used to compare the results of these six models. The simulated hydrographs were also decomposed into their major drivers and compared with known (i.e. synthetic) responses derived from the MODFLOW models.

Hydrograph-time-series-model performance

The performance of each of the six models was assessed by the NSE and Fig. 3 shows a box plot of NSE values for the models during the calibration and evaluation periods. Figure 3 shows that both linear time-series models (i.e. Von Asmuth 2008 models with Hantush's and Ferris and Knowles' well formulas; VA+H and VA+FK respectively) had comparable performance in terms of the median and inter-quartile range during the calibration period. For this period, inclusion of the soil moisture component significantly improved the median and inter-quartile range of the NSE of all models. Compared with the calibration period, the performance of all models significantly deteriorated during the evaluation period. Considering just the evaluation period and the models with the soil moisture component, the median NSE was higher for the models including the GET, compared with the equivalent models without GET.

To explore the performance of the VA and SMS-TFN models in detail, and to understand which model structure performs best under which conditions, Fig. 4 presents scatter plots of the calibration NSE for VA against the two SMS-TFN models (with GET and without GET). The mean depth to groundwater level is shown by the color. Overall, Fig. 4 shows that when the water table is deeper than 1 m, VA+FK often performs worse than the SMS-TFN+GET+FK and SMS-TFN+FK, which is primarily because of the differences in the simulation of meteorological forcing and is further investigated later.

Simulation of selected hydrographs

To compare the models in detail, and to further explore the change in performance with depth to the water table, Fig. 5 shows simulations from the six models for a hydrograph of 0.59 m mean depth (Fig. 5a–b) and a hydrograph of 6.70 m mean depth (Fig. 5c–d). With regard to the shallow hydrograph (Fig. 5a–b), all models behaved similarly during the calibration and evaluation periods and simulated most of the seasonal peaks and troughs. Figure 5a–b also shows that simulations of the seasonal peaks are poorest for two models without the groundwater ET component (i.e. SMS-TFN+H and SMS-TFN+FK; e.g. see 1996–1998 in Fig. 5a; note that the extinction depth of the MODFLOW model was 1.5 m). With regard to the deeper simulations (Fig. 5c–d), while all models performed reasonably well, the SMS-TFN models better simulated the hydrograph than the VA models (especially between Jan 1992 and Jan 1994 in Fig. 5c–d). The NSE values also improved when the soil moisture model is included in the model structure. For example, the calibration NSE for both VA models was around 0.6, while it increased to more than 0.8 for all four SMS-TFN models. This is consistent with the previous result (see section 'Hydrograph-time-series-model performance') and indicates better performance of SMS-TFN models compared with the VA models for deeper (i.e. >1 m) groundwater levels. Overall, these results showed that while there is a clear difference in performance of the models with depth to water table, all models worked reliably and performed reasonably well.

To explore the uncertainties of the model prediction, the nonlinear DREAM uncertainty analysis (Vrugt et al. 2008, 2009) was implemented. Figure 6 represents the result of uncertainty analysis for the two best models of VA+H and SMS-TFN+FK at the bores shown in Fig. 5. In Fig. 6, the uncertainty bounds are shown with two different colors, one is the 90 % prediction bounds due to parameter uncertainty obtained from DREAM analysis (i.e. dark grey color) and the other one (light grey) represents the prediction bounds due to residual errors. In general, the prediction bounds due to residual (unexplained) errors represent the model structural errors and measurement errors. Figure 6 shows that total uncertainty (i.e. the parameter uncertainty plus the contribution from residual error) captures most of the observation data. The coverage of total uncertainty is 91 and 89 % for shallow and deep groundwater level situations, respectively, which is consistent with the 90 % uncertainty range

DOI 10.1007/s10040-015-1269-7

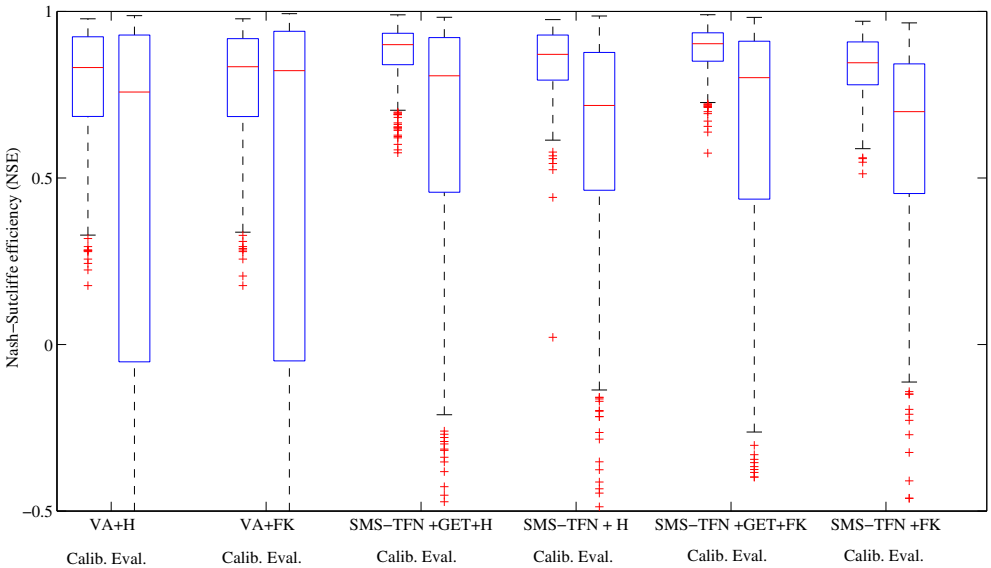

Fig. 3 Box plot of the Nash-Sutcliffe efficiency (*NSE*) during calibration (*Calib.*) and evaluation (*Eval.*) periods for all six proposed time models. *VA* Von Asmuth 2008 model, *SMS-TFN* soil moisture transfer function noise model, *GET* groundwater evapotranspiration component, *H* Hantush's well formula, *FK* Ferris and Knowles' well formula

used; hence, it can be concluded that the total uncertainty derived from DREAM analysis gives plausible estimates of actual uncertainty of total head. However, the parameter uncertainty (dark grey) is relatively small compare to total uncertainty, particularly for the shallow groundwater level situation (Fig. 6a) and, hence, the parameter uncertainty alone seems to underestimate the actual uncertainty.

Decomposition of selected hydrographs

Figure 5 shows that the performance of all six TFN models in simulating a shallow and deeper water level was acceptable to very good. To assess if an acceptable fit to the observed hydrograph translates into comparably reliable decomposition to the individual drivers, Fig. 7 presents the pumping decomposition for each of the six models in Fig. 5 and the known impact from pumping (including the irrigation recharge) derived from the MODFLOW modeling. For the shallow water table, Fig. 7a shows that each model using Hantush's (H) well formula reliably estimated the known pumping drawdown and that differences in the groundwater hydrograph simulations (see Fig. 5a) arose from differing model structure for the meteorological forcing. However, Fig. 7b shows that each model using Ferris and Knowles' (FK) well formula could estimate the drawdown dynamics but had a significant negative bias of between −0.56 and −0.85 m. For a deeper water table, Fig. 7c shows that the inclusion of the GET for each model using the H well formula resulted in very poor estimation of the drawdown and a considerable bias of 0.37 m. Lastly, Fig. 7d shows that each model using the FK well formula resulted in comparable performance to the pumping simulation with SMS-TFN+H (blue line in Fig. 7c), with a relatively small bias but acceptable estimation of the dynamics. Overall, Fig. 7 illustrates that reliable simulation of the total

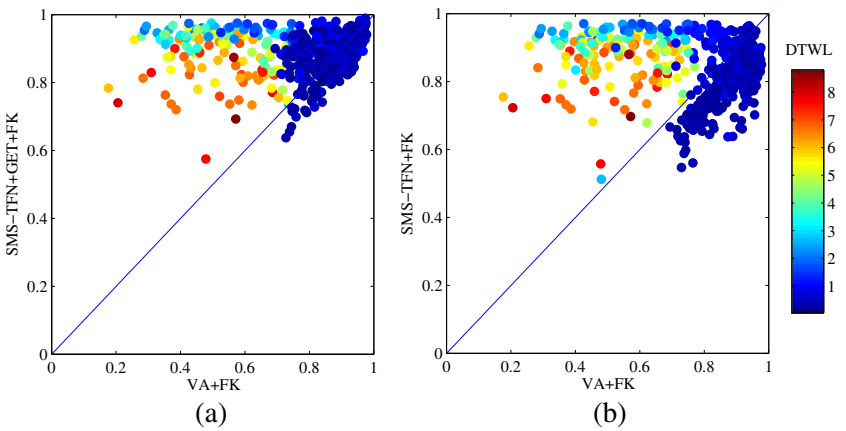

Fig. 4 Scatter plots of the soil moisture transfer function noise model (*SMS-TFN*) and Von Asmuth 2008 (*VA*) models with Ferris and Knowles' *(FK)* well formula: **a** SMS-TFN with groundwater ET component (*GET*) and VA models, **b** SMS-TFN and VA models. *DTWL* depth to water level (m)

DOI 10.1007/s10040-015-1269-7

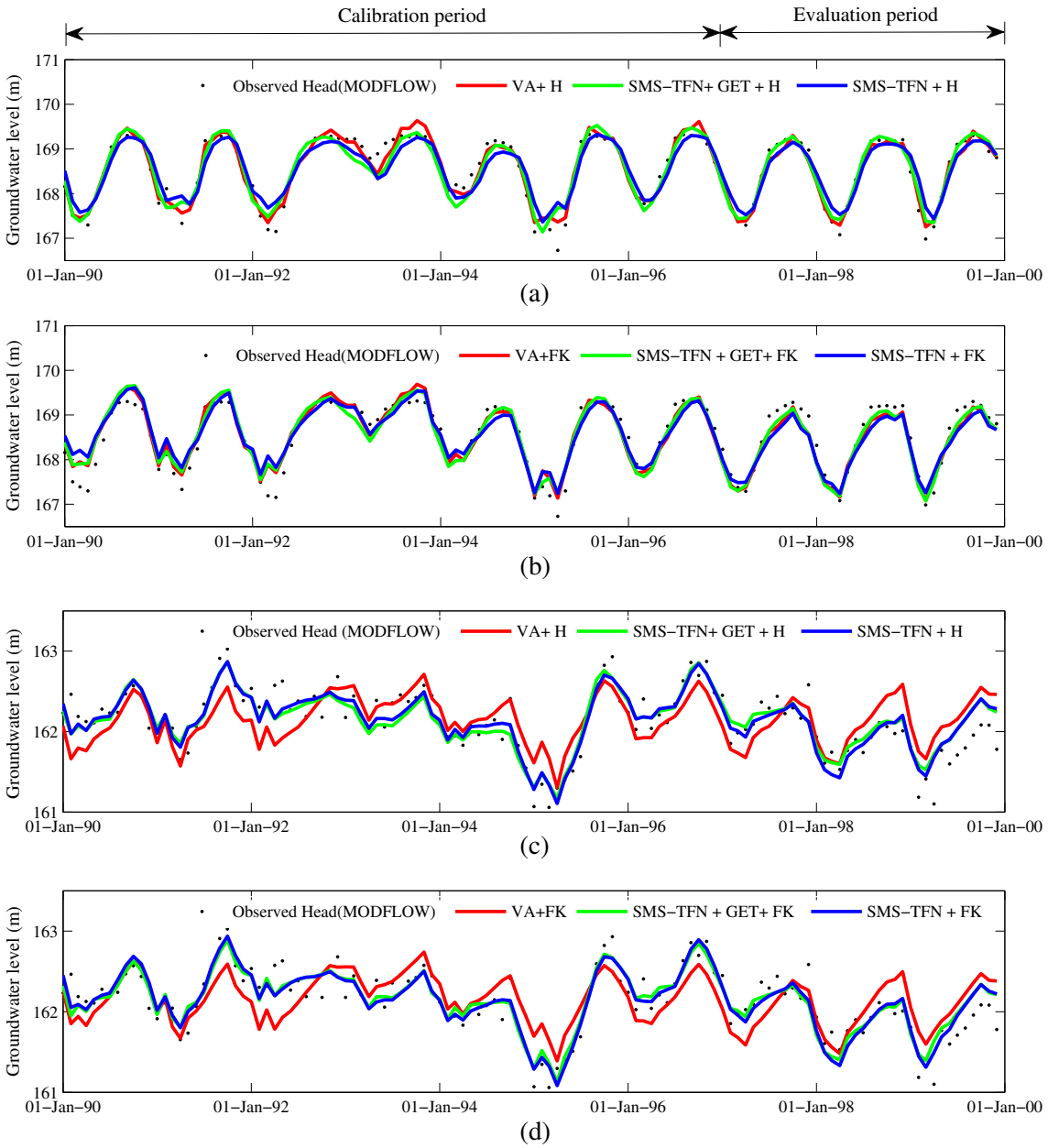

Fig. 5 Time series of observed and modeled groundwater head for two different bores. **a–b** The shallow groundwater level (mean depth to water level: 0.59 m, distance between observation bore and pumping bore: 300 m, specific yield: 0.11, saturated hydraulic conductivity: 5.46 m/day, mean thickness of aquifer: 35 m). **c–d** The deep groundwater level (mean depth to water level: 6.70 m, distance between observation bore and pumping bore: 400 m, specific yield: 0.12, saturated hydraulic conductivity: 17.98 m/day, mean thickness of aquifer: 88 m)

hydrograph (see Fig. 5), does not guarantee unbiased estimation of the drawdown and reliable estimation of the drawdown dynamics.

In an attempt to simulate the uncertainty bounds for the pumping contribution, the pumping parameter uncertainty derived from DREAM analysis for the head prediction in the prior section was used to obtain the 90 % parameter uncertainty bounds for the pumping contribution. The result indicates that the 90 % parameter uncertainty bounds do not demonstrate adequately the uncertainty within the pumping decomposition, which is mainly because, as described in the prior section, the majority of the total uncertainty in the head is represented by the residual unexplained error, which cannot be split in the decomposition analysis.

Pumping simulations

To further explore the reliability of drawdown estimation, this section quantifies the drawdown from all six TFN models applied to the 500 synthetic hydrographs. To assess the model performances statistically in detecting the pumping signal, the known drawdown from pumping and that estimated from time-series models was used to

DOI 10.1007/s10040-015-1269-7

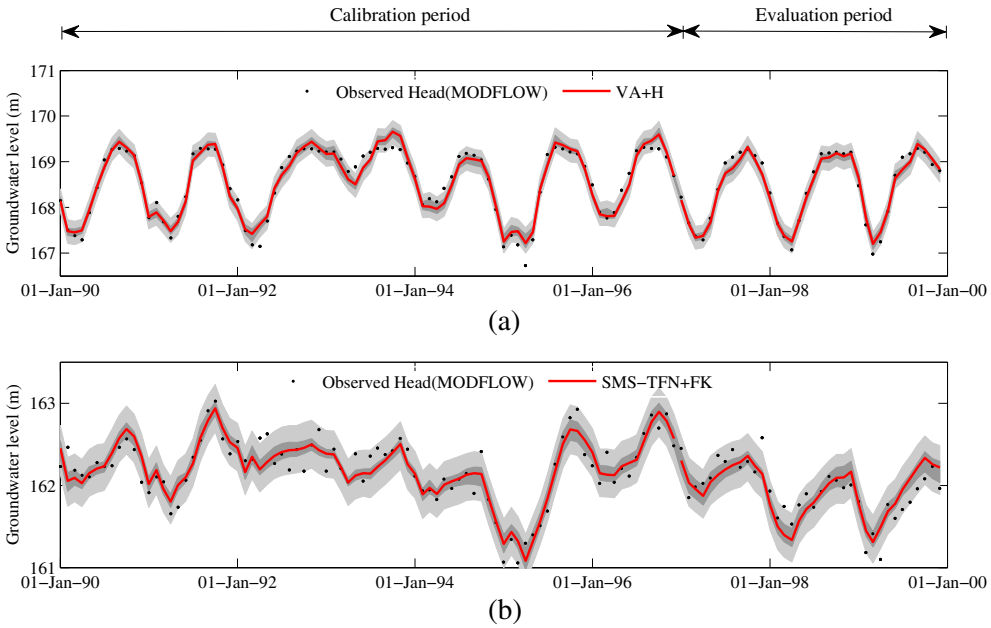

Fig. 6 Observed and best modeled groundwater levels at bores provided in Fig. 5 for **a** shallow groundwater level **b** deep groundwater level. The *dark grey area* represents the 90 % uncertainty interval due to parameter uncertainty. The *light grey region* shows the additional spread due to remaining residual error

calculate the NSE of the pumping signal estimate. In addition, it might be expected that the ability of the time series to detect the pumping signal would be a function of its variance relative to other drivers. Typically, a larger variance of the pumping signal indicates a stronger pumping impact on groundwater level variation, and hence, a potentially higher chance of being detected by the time-series model; therefore, the ratio between the variance of pumping signal and variance of the recharge–discharge signal (Q variance/R variance) has been also calculated for each of the 500 synthetic hydrographs. Figure 8 presents the scatter plot of the performance of each model in estimating the pumping (NSE for pumping signal in the evaluation period) as the color and the ratio of the variance of pumping to recharge–discharge signals (Q variance/R variance) and mean depth to groundwater level as two important factors on the x and y axes respectively (note that values of NSE≤0 are denoted as grey dots in Fig. 8). In addition, since significant bias arose between the true and model pumping simulation (Fig. 7b–c), the average of bias was also estimated for each model within shallow (i.e. <1 m) and deeper groundwater level (i.e. ≥ 1 m) categories and is shown in Table 2.

Figure 8 shows that when the pumping signal is weak (i.e. Q variance/R variance is<0.1), the pumping NSE from all model structures is less than or equal to zero, which indicates that the small variance of pumping signal (i.e. small pumping impacts) prevents the proper detection of any impact from pumping. For those hydrographs having a higher pumping signal and shallow groundwater level (e.g. <1 m), the H well formula models performed reasonably well (see the high NSE values in Fig. 8a,c,d) while for similar hydrographs, the FK well formula performed poorly (see Fig. 8b,e,f). The low values of NSEs were also associated with a considerable negative bias (i.e. mean biases of≤−0.35 m for the models with the FK equation in shallow groundwater level in Table 2). This poor performance of the FK equation is mainly due to the change in the net recharge during pumping and recovery. Overall, in the shallow groundwater situation, the aquifer is nearly full and GET is the main process of removing water from the aquifer. Any impact from pumping decreases the groundwater level and, hence, reduces the groundwater ET, which causes an increase in the net recharge. In addition, the aquifer receives extra recharge from irrigation within the irrigation area. These two processes attenuate the pumping impact and cause the full recovery after each period of pumping—e.g. known as pumping drawdown (black dots) in Fig. 7a. The FK equation is unable to simulate the impact of that extra recharge, while the leakage term in the Hantush equation enables the model to simulate extra recharge from other sources. Interestingly, the simulated total hydrograph from the FK models was relatively unbiased and well calibrated since the meteorological contribution internally compensates for the pumping bias.

For deeper groundwater levels (e.g. > 1 m) with a comparable pumping signal (e.g. Q variance/R variance>0.1), it can be seen that for SMS-TFN+GET+H and VA+H models, whereby the evapotranspiration component (i.e. second integral in either Eqs. 10 or 12) is incorporated explicitly in the model structure, the NSE degrades to less than or equal to zero. The mean bias was also high (>0.35 m) for those two models simulating deeper groundwater levels (Table 2). Again, this is in agreement with Fig. 7c, which similarly shows poor pumping simulation with significant bias for those two

DOI 10.1007/s10040-015-1269-7

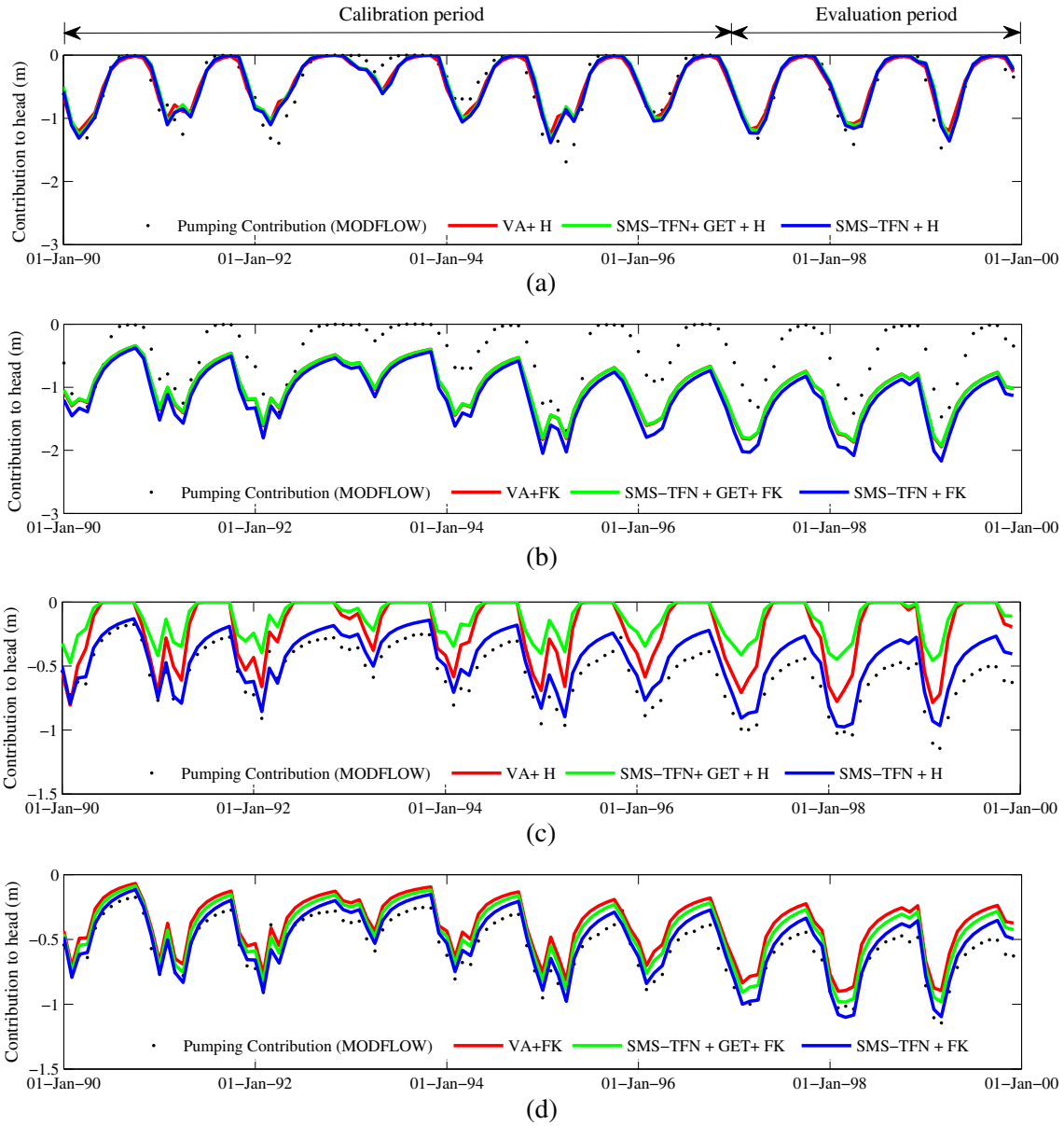

Fig. 7 Time series of observed and modeled pumping influence for two different bores. **a–b** The shallow groundwater level (mean depth to water level: 0.59 m, distance between observation bore and pumping bore: 300 m, specific yield: 0.11, saturated hydraulic conductivity: 5.46 m/day, mean thickness of aquifer: 35 m). **c–d** The deep groundwater level (mean depth to water level: 6.70 m, distance between observation bore and pumping bore: 400 m, specific yield: 0.12, saturated hydraulic conductivity: 17.98 m/day, mean thickness of aquifer: 88 m)

models (VA+H and SMS-TFN+GET+H models). Given that the influence of groundwater evapotranspiration and irrigation is minimal at deep groundwater levels, the inclusion of the evapotranspiration component explicitly in the time-series-model structure and the leakage term in the Hantush well formula seems to be redundant and this poor performance of pumping simulation is one of the consequences of over-parameterization in the model structure. Overall, these results are consistent with previous findings (see section 'Decomposition of selected hydrographs') and indicate that the decomposition of the groundwater hydrograph does not necessarily result in reliable impact of pumping, even if the overall head prediction is good.

With regard to the model structure providing the best overall drawdown, for deep groundwater levels (i.e.>1 m), it was found that the SMS-TFN model with FK's well formula gives the most reliable drawdown. However, when the water level is shallow (i.e. <1 m), the pumping drawdown can be reliably estimated using the Hantush well formula with either the VA or SMS-TFN model for the meteorological structure.

Meteorological simulations

Figure 8 showed that for quite modest to large pumping signals (e.g. Q variance/R variance>0.1), the time-series model has the ability to simulate the drawdown reliably.

DOI 10.1007/s10040-015-1269-7

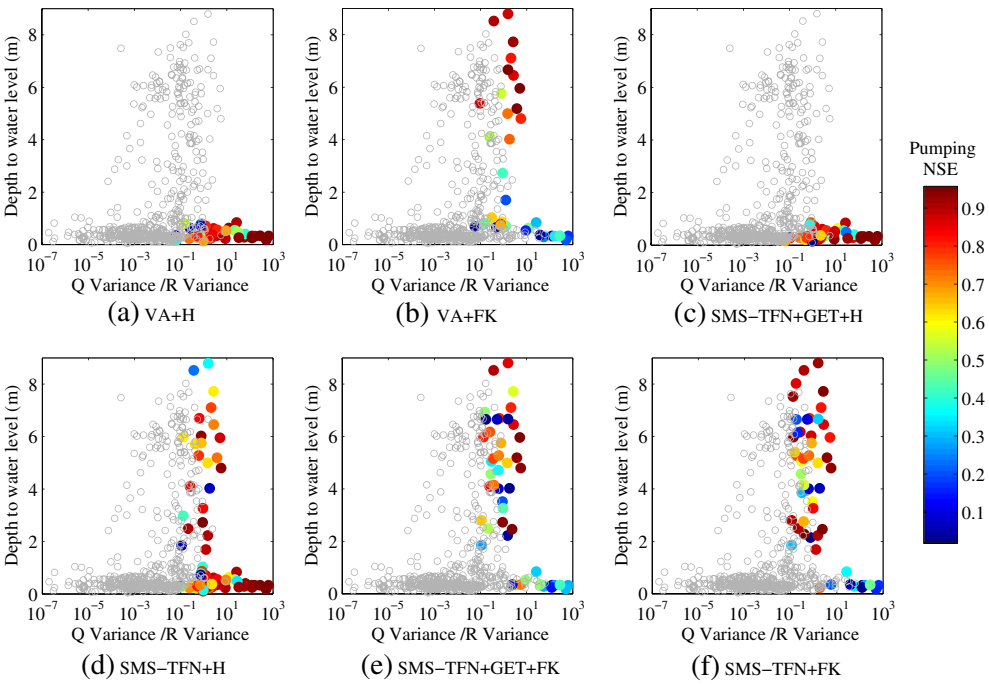

Fig. 8 Plot of depth to water level and the ratio between variance of pumping signal *(Q variance)* to the variance of recharge-evapotranspiration signal (*R variance*). The *dot* points are colored based on NSE for pumping signal and the *x axis* has the base 10 logarithmic scale. Note that the unbiased NSE values equal or below zero are shown as dark grey color. As previously, *VA* Von Asmuth 2008 model, *SMS-TFN* soil moisture transfer function noise model, *GET* groundwater evapotranspiration component, *H* Hantush's well formula, *FK* Ferris and Knowles' well formula

To assess if this reliable estimation of pumping drawdown also results in reliable estimation of climatic influences, Fig. 9 represents the NSE for the meteorological contribution during the evaluation period for all six time-series models in color, with the *x* and *y* axes showing the ratio of the variance of pumping to recharge–discharge variance and the depth to groundwater level, respectively. Figure 9 shows that poor NSEs resulted for all models when the meteorological signal is small (e.g. those with Q variance/R variance>10). This is similar to previous results for pumping simulations in section 'Pumping simulations', which showed that small variance in the pumping signal inhibits the reliable estimation of pumping (e.g. those with Q variance/R variance<0.1 in Fig. 8). These results suggest that the accurate estimation of one driver (pumping or meteorological signal) does not necessarily result in adequate estimation of the other one and clearly questions the validity of hydrograph decomposition when the groundwater is dominated by one of the main drivers.

Comparison of model performances in Fig. 9 also indicates that all SMS-TFN models perform acceptably for deep groundwater levels (e.g. >1 m); however, the SMS-TFN model without GET performs worse when the groundwater level is shallow (e.g. <1 m). The lower model performance for those models (SMS-TFN+FK and SMS-TFN+H) in shallow aquifers is due to the lack of a groundwater evapotranspiration component when the groundwater evapotranspiration is one of the important processes. In addition, both VA models cannot simulate the meteorological contribution properly for deeper groundwater levels (e.g. >1 m), which is consistent with the previous findings for the overall model simulation in section 'Hydrograph-time-series-model performance' and indicates that poor estimation of the meteorological contribution is the main reason for lower performance of VA compare to SMS-TFN in Fig. 4a–b.

Recharge patterns for shallow and deep groundwater levels

To explore the often poor meteorological simulation from the VA models, Fig. 10 summarizes two dominate

Table 2 Mean bias in the simulated drawdown from the 500 hydrograph simulations for each of the six model structures. Note a positive bias denotes the simulated pumping underestimated drawdown and vice versa for a negative bias. *Italic values* denote significant negative or positive bias

Water table	VA+H	VA+FK	SMS-TFN+GET+H	SMS-TFN+H	SMS-TFN+GET+FK	SMS-TFN+FK
Shallow (DTWL <1 m)	0.08	*−0.35*	0.08	0.07	*−0.42*	*−0.58*
Deep (DTWL ≥1 m)	*0.49*	−0.05	*0.50*	0.15	0.05	0.06

VA Von Asmuth 2008 model, *SMS-TFN* soil moisture transfer function noise model, *GET* groundwater evapotranspiration component, *H* Hantush's well formula, *FK* Ferris and Knowles' well formula, *DTWL* depth to water level (m)

DOI 10.1007/s10040-015-1269-7

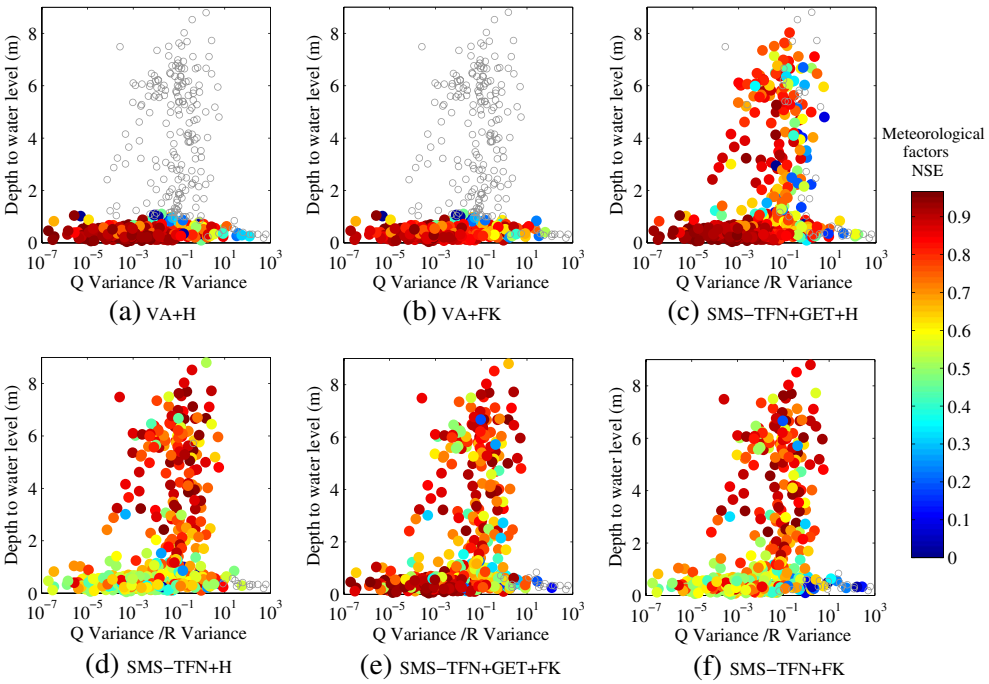

Fig. 9 Plot of depth to water level and the logarithm of the ratio between variance of pumping signal (*Q variance*) to the variance of recharge-evapotranspiration signal (*R variance*). The *dot points* are colored based on NSE for the meteorological signal. Note that the meteorological contribution from the time-series model can be at a different level from the true meteorological contribution; to make those comparable, the average of the time series were subtracted from both the model and true meteorological contributions and then the remaining were used for NSE estimation

recharge mechanisms from the MODFLOW models for a shallow (Fig. 10a) and a deeper (Fig. 10b) water table. Figure 10a shows that there is no timing difference between infiltration (the amount of water which is added to the top of the unsaturated zone) and recharge (the amount of water percolating to groundwater) for shallow groundwater situations. In addition, most rainfall (about 60 %) infiltrates through the unsaturated zone and the recharge is relatively linear with input infiltration. In contrast, Fig. 10b shows that significant nonlinearity occurs between the precipitation input and the recharge output for a given month for deeper groundwater levels

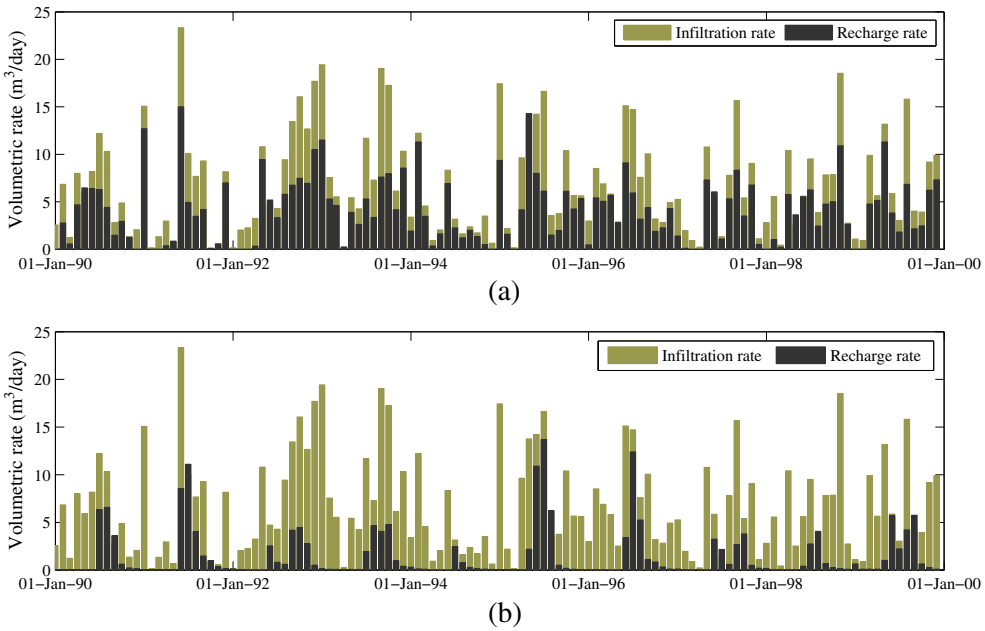

Fig. 10 The infiltration rate (m^3/day) and recharge rate (m^3/day) for **a** shallow groundwater and **b** deep groundwater example mentioned in Fig. 5. Note the infiltration rate is the amount of volumetric water which is applied to the top of the unsaturated zone and the recharge rate is the volumetric water that reaches to groundwater level after passing through the unsaturated zone flow model (UZF)

 DOI 10.1007/s10040-015-1269-7

and only small amounts percolate to the groundwater (around 20 % of total volumetric infiltration). This indicates that the VA model performs poorly because it lacks the ability to simulate the nonlinearity between the infiltrated water and recharge.

Discussion

Importance of hydrograph decomposition evaluation

One of the important aspects of a time-series model is its ability to model the groundwater level and decompose the groundwater hydrograph into its contributions; however, the results of this study clearly demonstrate that simulating the groundwater hydrograph properly does not necessary result in reliable hydrograph decomposition. It was shown that internal model compensation, over-parameterization and strong dominance of one driver over the other inhibit accurate estimation of one or both of the pumping and climatic influences, clearly highlighting the need for model structure evaluation when accurate hydrograph decomposition is desired. This has often not been addressed in previous studies (Obergfell et al. 2013; Von Asmuth et al. 2008; Yihdego and Webb 2011) where successful simulation of groundwater hydrographs was assumed to imply accurate decomposition. The inherent lack of independent data for responses to major drivers at field sites is the probable reason why this hydrograph decomposition assessment has not been undertaken previously.

Identifying the most reliable time-series model

The results of this study showed that reliable decomposition of drawdown and pumping could be achieved but differing model structures were required for shallow and deep water-table conditions. In summary, for a shallow groundwater level, recharge responds linearly to infiltration, and GET and irrigation are essential drivers; thus, Hantush's well formula and either VA or SMS-TFN with GET model structures were required to simulate the pumping and climatic impacts properly. For deeper groundwater levels, the recharge responds nonlinearly to infiltration, and the influence of GET and irrigation are minimal; therefore, the SMS-TFN model structure, which accounts for nonlinearity in recharge combined with Ferris and Knowles' well formula (i.e. the well formula with no leakage parameter), best simulated both pumping and meteorological influences.

This work highlights the complexity of groundwater level responses to different drivers, even within a synthetic framework, and emphasizes the importance of good knowledge about the catchments and the main processes contributing to groundwater level variation prior to adopting any particular time-series model. For instance, if there is clear evidence indicating no pumping induced recharge or leakage from other sources, adopting the simplest pumping response function (i.e. the FK well formula) is preferred over more complex response functions such as the H well formula, since it removes any redundant component and reduces the chance of internal compensation. Similarly, if water tables are sufficiently deep to prevent the local vegetation accessing groundwater, the GET component should be excluded.

Guidance for reliable estimation of pumping drawdown

In section 'Simulation of selected hydrographs', the ratio of pumping variance to recharge–discharge variance was found to be a useful guide to determine the range of pumping signals in which the time series approach might be expected to detect the pumping drawdown with an acceptable level of accuracy; however, for the application of the time-series model to field sites, the ratio of the pumping variance to recharge–discharge variance is unknown. One alternative is to estimate the true ratio using the simulated ratio. To trial such an approach, the SMS-TFN+FK and SMS-TFN+GET+H models were chosen for deep (e.g. >1 m) and shallow situations (e.g. ≤1 m), respectively, and the simulated ratio was derived for the two models and compared with the known ratio from the synthetic MODFLOW models. Figure 11 presents the scatter plot of fitted and known pumping to recharge–discharge variance ratio. Given that there is still a modest bias occurring in the pumping simulation for these two groups of models (mean bias of 0.06 and 0.08 m for SMS-TFN+FK and SMS-TFN+GET+H for deep and shallow groundwater respectively in Table 2), the evaluation NSE for pumping was rederived for the two models by removing the bias from the NSE estimation. In doing so, the bias was estimated from Eq. (17) and then added to Eq. (18) to reestimate NSE.

$$\text{Bias} = \frac{\sum_{i=1}^{n} \left(D^{i}_{\text{mod}} - D^{i}_{\text{obs}}\right)}{n} \tag{24}$$

$$\text{Unbiased\_NSE} = 1 - \frac{\sum_{i=1}^{n} \left(D^{\text{i}}_{\text{mod}} - D^{i}_{\text{obs}} - \text{Bias}\right)^{2}}{\sum_{i=1}^{n} \left(D^{i}_{\text{obs}} - \overline{D_{\text{obs}}}\right)^{2}} \tag{25}$$

This rederived NSE is further labeled 'unbiased_NSE' and is represented by the color spectrum in Fig. 11, which shows that for a fitted pumping to recharge–discharge variance ratio of ≥1 (i.e. the size of pumping signal is at least the same or bigger than the meteorological signal), the fitted ratio is generally unbiased and good pumping NSE is achieved (e.g. unbiased_NSE>0.65), and the fitted variance ratio represents, to some extent, the true variance ratio. For a fitted variance ratio of<1, an acceptable pumping NSE might be achieved but the grey dots indicate that there is a risk that the fitted ratio could be a poor representation of the true ratio, which suggests that, overall, the model most likely represents the true influence

DOI 10.1007/s10040-015-1269-7

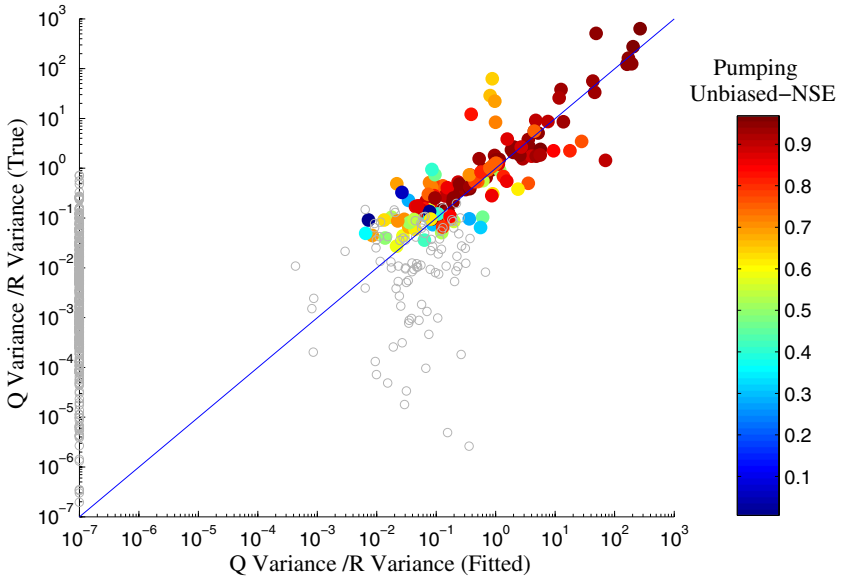

Fig. 11 Scatter plot of ratio of variance of pumping to recharge–discharge signal (*Q variance/R variance*) resulted from fitted and true estimation. The unbiased_NSE is the NSE whereby the bias is removed from the NSE estimation. Note that the unbiased_NSE values equal or below zero are shown as *dark grey color*, and for better presentation, the fitted variance ratio for those where the fitted variance is equal or less than 10^{-7}, is assumed to be equal to 10^{-7}

of pumping well if the simulated pumping impact is significant relative to the other drivers.

Impact of additional errors to forcing and input data

To explore any possible impact of the forcing and head measurement error introduced in section 'Synthetic groundwater hydrographs' on the ability of the time series model to detect the pumping and meteorological signals, the two models (SMS-TFN+GET+H and SMS-TFN+FK), used in the prior section for shallow and deep situations, were selected and recalibrated with 'true' data (i.e. no error was added to input forcing and groundwater heads). The result indicates that removing measurement errors does not change the 0.1 and 10 thresholds for the pumping to meteorological variance ratio (identified in sections 'Pumping simulations' and 'Meteorological simulations'), which provide plausible estimation of pumping and meteorological impacts. In addition, for pumping to meteorological variance ratios between 0.1 and 10, the median NSE for pumping decomposition improves significantly from 0.13 to 0.59 when perfect input data are used. Similarly, the median COE for meteorological decomposition also indicates an improvement from 0.73 to 0.86 when using error free data. It should be noted that the magnitude of error introduced in the paper is within typical observation error range and consistent with errors used in literature (e.g. Renard et al. 2010 and Hill and Tiedeman 2006).

Conclusion

The results of this study demonstrate that hydrograph separations obtained from time-series models do not always result in reliable estimation of pumping and climatic influences even when the overall hydrograph fit is good. However, when the time-series model represents the important processes (e.g. phreatic evaporation is included for shallow water tables) and the (head) variance of the pumping signal to the meteorological signal is between 0.1 and 10, the model has the potential to adequately separate the influence of pumping and meteorological factors.

In this study, Hantush's and Ferris and Knowles' well formulas were adopted to account for the impact of pumping. The underlying assumptions of these equations are that the aquifer is confined, horizontal, of infinite extent and homogeneous and that the pumping bore fully penetrates the aquifer. In this case, the main violations from aforementioned assumptions are that the aquifer is unconfined, finite and sloping. With regard to the adoption of a confined aquifer well formula for an unconfined situation, it has been demonstrated that the confined pumping response function can be used in unconfined situations if the phenomenon of delayed water-level response is minimal and the drawdown due to pumping remains small relative to the total aquifer thickness (Kruseman and De Ridder 1994). In this synthetic study, this assumption appears to be satisfied, because in all simulations, there is no delayed response influence and the maximum decline due to pumping is less than 10 % of the total thickness of the aquifer. With regard to the assumption of infinite extent, obviously, the catchment here is bounded but considering that in 96 % of total model simulations, the pumping impact does not reach to any boundary around the catchment, this assumption appears to be satisfied in most of model simulations. It is acknowledged that for those 4 % of total simulation where pumping influence reaches to the boundaries, there

DOI 10.1007/s10040-015-1269-7

might be some degradation in the simulated pumping. Some field trials have found that fixed head boundary conditions, caused by surface-water bodies, can degrade the performance of the time-series model (Shapoori et al. 2015).

It should also be noted that such a synthetic study of model performance as presented here is rarely undertaken in hydrogeological studies. It is not clear why few such studies have been undertaken; however, in light of the complexity of groundwater systems (e.g. boundary conditions, aquifer structure, aquifer parameter values, recharge dynamics) and the variability between systems, only a subset of groundwater systems can ever be investigated and investigation of all types is impossible. Hence, the falsification and evaluation of the various time-series models for decomposition will always be based on a subset of possible synthetic model. Overall, the development herein of 500 synthetic models of an unconfined aquifer is felt to be a sufficient subset of models to identify the applicability of the groundwater-time-series method for reliable decomposition of the impacts of pumping and meteorological factors.

While the results of this study can be used as a guide for aquifers (i.e. upland valley) with a similar range of properties, the details should not be generalized to all situations. Any major changes could potentially lead to different results—for example, if the pumping bore is used for different purposes (e.g. for drinking usage), the temporal pattern of pumping would be different (e.g. constant extraction over the entire year instead of only during the irrigation period). In addition, different assumptions for the vegetation (e.g. deep-rooted vegetation) could change the evapotranspiration extinction depth and affect the rate of evapotranspiration, which would lead to different thresholds for the mean depth to groundwater level when differentiating between shallow and deeper groundwater levels. Having said that, the result that pumping impact estimates improve as the pumping signal gets larger compared with the meteorological signal is likely to be general, even if the relationship differs in detail between situations. To provide guidance for other cases, the time-series model performance needs to be assessed in a synthetic environment simulating each particular case with its specifics with regard to meteorological and pumping regimes.

Acknowledgements The authors are grateful for the financial support received from the Australian Research Council (grant numbers: LP0991280 and LP130100958), the Department of Environment and Primary Industries (Victoria, Australia), and the Bureau of Meteorology (Australia). The authors thank the editors and anonymous reviewers for their constructive comments.

References

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration-guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56, FAO, Rome

Andreu JM, Alcala FJ, Vallejos A, Pulido-Bosch A (2011) Recharge to mountainous carbonated aquifers in SE Spain: different approaches and new challenges. J Arid Environ 75:1262–1270. doi:10.1016/j.jaridenv.2011.01.011

Bakker M, Maas K, Von Asmuth JR (2008) Calibration of transient groundwater models using time series analysis and moment matching. Water Resour Res 44:W04420. doi:10.1029/2007wr006239

Brooks RH, Corey AT (1964) Hydraulic properties of porous media. Hydrology Papers, Colorado State University, Fort Collins, CO, 24 pp

Butler JJ, Kluitenberg GJ, Whittemore DO, Loheide SP, Jin W, Billinger MA, Zhan XY (2007) A field investigation of phreatophyte-induced fluctuations in the water table. Water Resour Res 43:W02404. doi:10.1029/2005WR004627

Crosbie RS, Binning P, Kalma JD (2005) A time series approach to inferring groundwater recharge using the water table fluctuation method. Water Resour Res 41:W01008. doi:10.1029/2004WR003077

Cuthbert MO (2010) An improved time series approach for estimating groundwater recharge from groundwater level fluctuations. Water Resour Res 46:W09515. doi:10.1029/2009WR008572

Doll P (2009) Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment. Environ Res Lett 4:035006. doi:10.1088/1748-9326/4/3/035006

Fan J, Pan J (2006) Convergence properties of a self-adaptive Levenberg-Marquardt algorithm under local error bound condition. Comput Opt Appl 34:47–62

Ferdowsian R, George A, Bee G, Smart R (2002) Groundwater level reductions under Lucerne depend on the landform and groundwater flow systems (local and intermediate). Aust J Soil Res 40:381–396

Ferket BVA, Samain B, Pauwels VRN (2010) Internal validation of conceptual rainfall–runoff models using baseflow separation. J Hydrol 381:158–175. doi:10.1016/j.jhydrol.2009.11.038

Ferris JG, Knowles DB (1963) The slug-injection test for estimating the coefficient of transmissibility of an aquifer. In: Bentall R (ed) Methods of determining permeability, transmissibility, and drawdown. US Geol Surv Water Suppl Pap 1536-I

Freeze RA, Cherry JA (1979) Groundwater. Prentice Hall, Englewood Cliffs, NJ

Gerla PJ (1992) The relationship of water-table changes to the capillary-fringe, evapotranspiration, and precipitation in intermittent wetlands. Wetlands 12:91–98

Hantush MS (1956) Analysis of data from pumping tests in leaky aquifers. Trans Am Geophys Union 37:702–714

Hill MC, Tiedeman CR (2006) Effective groundwater model calibration: with analysis of data, sensitivities, predictions, and uncertainty. Wiley, Hoboken, NJ

Jakeman AJJ, Hornberger GM (1993) How much complexity is warranted in a rainfall-runoff model? Water Resour Res 29:2637–2649

Kavetski D, Kuczera G, Franks S (2006) Bayesian analysis of input uncertainty in hydrological modeling, 1. Theor Water Resour Res 42:W03407. doi:10.1029/2005WR004368

Konikow LF, Kendy E (2005) Groundwater depletion: a global problem. Hydrogeol J 13:317–320. doi:10.1007/s10040-004-0411-8

Kruseman GP, De Ridder NA (1994) Analysis and evaluation of pumping test data. International Institute for Land Reclamation and Improvement, Wageningen, The Netherlands

Kundzewicz ZW, Doll P (2009) Will groundwater ease freshwater stress under climate change? Hydrol Sci J 54:665–675. doi:10.1623/hysj.54.4.665

Lehsten D, Von Asmuth JR, Kleyer M (2011) Simulation of water level fluctuations in kettle holes using a time series model. Wetlands 31:511–520. doi:10.1007/s13157-011-0174-7

Levenberg K (1944) A method for the solution of certain non-linear problems in least squares. Quart Appl Math 2:164–168

Li L, Maier HR, Lambert MF, Simmons CT, Partington D (2013) Framework for assessing and improving the performance of

recursive digital filters for baseflow estimation with application to the Lyne and Hollick filter. Environ Model Softw 41:163–175. doi:10.1016/j.envsoft.2012.11.009
Li L, Maier HR, Partington D, Lambert MF, Simmons CT (2014) Performance assessment and improvement of recursive digital baseflow filters for catchments with different physical characteristics and hydrological inputs. Environ Model Softw 54:39–52. doi:10.1016/j.envsoft.2013.12.011
Loheide SP, Butler JJ, Gorelick SM (2005) Estimation of groundwater consumption by phreatophytes using diurnal water table fluctuations: a saturated–unsaturated flow assessment. Water Resour Res 41:W07030. doi:10.1029/2005WR003942
Manzione RL, Knotters M, Heuvelink GBM, Von Asmuth JR, Camara G (2010) Transfer function-noise modeling and spatial interpolation to evaluate the risk of extreme (shallow) water-table levels in the Brazilian Cerrados. Hydrogeol J 18:1927–1937. doi:10.1007/s10040-010-0654-5
Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. J Soc Ind Appl Math 11:431–441
McMillan H, Jackson B, Clark M, Kavetski D, Woods R (2011) Rainfall uncertainty in hydrological modeling: an evaluation of multiplicative error models. J Hydrol 400:83–94. doi:10.1016/j.jhydrol.2011.01.026
Morton FI (1983) Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology. J Hydrol 66:1–76
Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I: a discussion of principles. J Hydrol 10:282–290. doi:10.1016/0022-1694(70)90255-6
Niswonger RG, Prudic DE, Regan RS (2006) Documentation of the Unsaturated-Zone Flow (UZF1) Package for modeling unsaturated flow between the land surface and the water table with MODFLOW-2005. US Geological Survey, Reston, VA
Obergfell C, Bakker M, Zaadnoordijk WJ, Maas K (2013) Deriving hydrogeological parameters through time series analysis of groundwater head fluctuations around well fields. Hydrogeol J 21:987–999. doi:10.1007/s10040-013-0973-4
Partington D, Brunner P, Simmons CT, Therrien R, Werner AD, Dandy GC, Maier HR (2011) A hydraulic mixing-cell method to quantify the groundwater component of streamflow within spatially distributed fully integrated surface water–groundwater flow models. Environ Model Softw 26:886–898
Partington D, Brunner P, Simmons C, Werner A, Therrien R, Maier H, Dandy G (2012) Evaluation of outputs from automated baseflow separation methods against simulated baseflow from a physically based, surface water-groundwater flow model. J Hydrol 458:28–39
Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen-Geiger climate classification. Hydrol Earth Syst Sci Discuss 4:439–473
Peterson TJ, Western AW (2011) Time-series modeling of groundwater head and its de-composition to historic climate periods. Paper presented at the 34th IAHR World Congress, Brisbane, Australia, 26 June–1 July 2011
Peterson TJ, Western AW (2014) Nonlinear time series modeling of unconfined groundwater head. Water Resour Res 50:8330–8355. doi:10.1002/2013WR014800
Renard B, Kavetski D, Kuczera G, Thyer M, Franks SW (2010) Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. Water Resour Res 46:W05521. doi:10.1029/2009wr008328
Rosenberry DO, Winter TC (1997) Dynamics of water-table fluctuations in an upland between two prairie-pothole wetlands in North Dakota. J Hydrol 191:266–289. doi:10.1016/S0022-1694(96)03050-8
Scanlon BR, Healy RW, Cook PG (2002) Choosing appropriate techniques for quantifying groundwater recharge. Hydrogeol J 10:19–39
Shamsudduha M, Taylor R, Ahmed K, Zahid A (2011) The impact of intensive groundwater abstraction on recharge to a shallow regional aquifer system: evidence from Bangladesh. Hydrogeol J 19:901–916. doi:10.1007/s10040-011-0723-4
Shapoori V, Peterson TJ, Western AW, Costelloe JF (2015) Top-down groundwater hydrograph time series modeling for climate-pumping decomposition. Hydorgeol J. doi:10.1007/s10040-014-1223-0 (Published online)
Siriwardena L, Peterson TJ, Western AW (2011) A state-wide assessment of optimal groundwater hydrograph time series models. Paper presented at the International Congress on Modeling and Simulation, Perth, Australia, 12–16 December 2011
Sivapalan M, Young PC (2005) Downward approach to hydrological model development. Encycl Hydrol Sci 3:2081–2098
Sophocleous M (2003) Environmental implications of intensive groundwater use with special regard to streams and wetlands. In: Llamas MR, Custodio E (eds) Groundwater intensive use: challenges and opportunities. Balkema, Dordrecht, The Netherlands
Szilagyi J (2004) Heuristic continuous base flow separation. J Hydrol Eng 9:311–318. doi:10.1061/(asce)1084-0699(2004)9:4(311)
Tularam GA, Krishna M (2009) Long term consequences of groundwater pumping in Australia a review of impacts around the globe. J Appl Sci Environ Sanit 4:151–166
Viswanathan MN (1984) Recharge characteristics of an unconfined aquifer from the rainfall–water table relationship mathematical models, Newcastle, Australia, depends upon the intensity. J Hydrol 70:233–250. doi:10.1016/0022-1694(84)90124-0
Von Asmuth JR, Bierkens MFP, Maas K (2002) Transfer function-noise modeling in continuous time using predefined impulse response functions. Water Resour Res 38:23. doi:10.1029/2001WR001136
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
Vrugt JA, Braak CFJ, Clark MP, Hyman JM, Robinson BA (2008) Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. Water Resour Res 44:W00B09. doi:10.1029/2007WR006720
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. Int J Nonlinear Sci Numeric Simul 10:273–290
White WN (1932) A method of estimating ground-water supplies based on discharge by plants and evaporation from soil: results of investigations in Escalante Valley. US Geol Surv Water Suppl Pap 659-A
Yi MJ, Lee KK (2004) Transfer function-noise modeling of irregularly observed groundwater heads using precipitation data. J Hydrol 288:272–287. doi:10.1016/j.jhydrol.2003.10.020
Yihdego Y, Webb JA (2011) Modeling of bore hydrographs to determine the impact of climate and land-use change in a temperate subhumid region of southeastern Australia. Hydrogeol J 19:877–887. doi:10.1007/s10040-011-0726-1
Young PC (1978) A general theory of modeling for badly defined dynamic systems. In: Vansteenkiste GC (ed) Modeling, identification and control in environmental systems. North Holland, Amsterdam, pp 103–135
Zektser S, Loaiciga HA, Wolf JT (2005) Environmental impacts of groundwater overdraft: selected case studies in the southwestern United States. Environ Geol 47:396–404. doi:10.1007/s00254-004-1164-3