Check for updates

Joint Estimation of Gross Recharge, Groundwater Usage, and Hydraulic Properties within HydroSight

by Tim J. Peterson[1,2] and Simon Fulton[3]

Abstract

Groundwater management decisions are often founded upon estimates of aquifer hydraulic properties, recharge and the rate of groundwater usage. Too often hydraulic properties are unavailable, recharge estimates are very uncertain, and usage is unmetered or infrequently metered over only recent years or estimated using numerical groundwater models decoupled from the drivers of drawdown. This paper extends the *HydroSight* groundwater time-series package (http://peterson-tim-j.github.io/HydroSight/) to allow the joint estimation of gross recharge, transmissivity, storativity, and daily usage at multiple production bores. A genetic evolutionary scheme was extended from estimating time-series model parameters to also estimating time series of usage that honor metered volumes at each production bore and produces (1) the best fit with the observed hydrograph and (2) plausible estimates of actual evapotranspiration and hence recharge. The reliability of the approach was rigorously tested. Repeated calibration of models for four bores produced estimates of transmissivity, storativity, and mean recharge that varied by a factor of 0.22-0.32, 0.13-0.2, and 0.03-0.48, respectively, when recharge boundary effects were low and the error in monthly, quarterly, and biannual metered usage was generally <10%. Application to the 30 observation bores within the Warrion groundwater management area (Australia), produced a coefficient of efficiency of ≥0.80 at 22 bores and ≥0.90 at 12 bores. The aquifer transmissivity and storativity were reasonably estimated, and were consistent with independent estimates, while mean gross recharge may be slightly overestimated. Overall, the approach allows greater insights from the available data and provides opportunity for the exploration of usage and climatic scenarios.

Introduction

Estimates of aquifer hydraulic properties are fundamental for managing groundwater usage. Ideally, long-term pumping tests are conducted and appropriate drawdown equations are fit to the observed drawdown using the known extraction rate(s). However, too often pumping test estimates are not available or are biased due to insufficient duration (Shapoori et al. 2015c). Time-series analysis of groundwater hydrographs that account for time-varying groundwater extraction has been shown to be a reliable alternative to pumping tests (Bakker et al. 2008; Shapoori et al. 2015c). It requires only existing data and is comparable to a very long-term pumping test with multiple production bores, each having a differing and time-varying

extraction rate. When the complete history of groundwater usage is frequently metered (i.e., monthly), time-series analysis allows the efficient estimation of aquifer properties and, when undertaken at many observation bores, a data-driven estimation of aquifer heterogeneity (Shapoori et al. 2015c) and the cumulative impact of usage (Shapoori et al. 2015b). However, few regions have a complete and frequently metered record of groundwater usage and this inhibits estimation of the aquifer hydraulic properties from the groundwater hydrograph. Furthermore, groundwater hydrographs are likely to also be influenced by recharge and hence the reliable estimation of drawdown and recovery, and aquifer hydraulic properties, is conditional upon the reliable estimation of the recharge over time. This paper presents a nonlinear transfer function noise (TFN) model that addresses these challenges to provide jointly derived estimates of transmissivity, storativity, and gross recharge from an observed groundwater hydrograph. This is achieved by deriving estimates of historic usage that honor the infrequently metered usage at each production bore and estimates of gross recharge that are constrained by plausible bounds for evapotranspiration. The approach has been incorporated into the *HydroSight* groundwater time-series

[1]Department of Civil Engineering, Monash University, Wellington Rd, Clayton, VIC 3010, Australia
[2]Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia
[3]Corresponding author: Monash University, Clayton, Australia; tim.peterson@monash.edu

package (http://peterson-tim-j.github.io/HydroSight/) and was applied to a groundwater management area in Victoria, Australia, and results were evaluated against independent estimates.

Estimating a realistic groundwater extraction history from incomplete and infrequent measurement records has been an ongoing challenge, and one critical for quantifying the impact of extractions on groundwater dynamics and the environment. Liu et al. (2015) outlined four methods to estimate groundwater usage: (1) pump electricity consumption analysis, (2) water budget method, (3) groundwater hydrograph analysis, and (4) numerical model simulations. Recently, a fifth method was proposed that used time-invariant site attributes and generalized linear models to predict the distribution of mean usage at the property and well scale (Keir et al. 2019). The approach was applied to estimate the usage at 10,168 properties and 19,700 pumping bores, but it did not enable estimation of aquifer hydraulic properties or drawdown. When estimates of drawdown are required, numerical groundwater models are often used. However, too often the models are constructed with *a priori* fixed estimates of usage derived from independent data (e.g., Martínez-Santos et al. 2010). Recently, this input data uncertainty has been addressed using inverse groundwater modeling approaches that update the usage per model grid cell and time step from the difference between the observed and modeled groundwater head (Martínez-Santos and Martínez-Alfaro 2010; Shao et al. 2014). A more advanced approach identified usage behavior from signal processing of hourly groundwater hydrographs and then incorporated this into a groundwater model to estimate usage (Liu et al. 2015). To our knowledge, all such numerical methods rely on the prior estimation of both aquifer hydraulic properties and recharge, and do not allow for the joint estimation of usage, aquifer properties, and recharge. Consequently, and considering the covariance between usage and aquifer hydraulic properties as captured within the Theis drawdown equation, such usage estimation methods are likely to produce biased estimates of usage, aquifer hydraulic properties, and/or recharge.

In this paper, we estimate usage by calibrating a simulated groundwater hydrograph to optimally fit an observed hydrograph. The approach overcomes the difficulty with hydrograph-based methods requiring estimates of recharge (Liu et al. 2015) and overcomes the weakness of numerical models not providing well-specific usage (Liu et al. 2015) and does not require prior model calibration of aquifer properties and recharge. Hence, our approach merges usage estimate types three and four. The approach is an extension of Shapoori et al. (2015a, 2015b, 2015c), which is based upon impulse response functions (von Asmuth et al. 2002, 2005) extended to account for nonlinear partitioning of precipitation (Peterson and Western 2014). This prior work undertook time-series modeling of individual groundwater hydrographs using daily meteorological time-series data and historic monthly groundwater usage data; with the former used to simulate the meteorological drivers of groundwater head fluctuations and the latter the drawdown and recovery from groundwater extraction. By simulating the drawdown and recovery using instantaneous versions of analytical drawdown functions (e.g., Theis), then when the time-series model is fit to the observed groundwater hydrograph estimates of aquifer hydraulic properties (transmissivity and storativity) can be derived. Effectively this approach is analogous to a long-term pumping test with time-varying usage. However, considering the approach effectively simulates drawdown by weighting the historic usage at each production bore, it requires at least metered monthly and ideally daily usage over the entire history of extraction at each production bore. Assuming most groundwater management regions are like those in Victoria (Australia) where metering of usage commenced recently in 2004 and provides at best biannual estimates (Gill et al. 2017), there are likely to be many prior years of unmetered extraction and this limits application of the aforementioned approach.

To resolve this, we extended the approach to derive time series of usage at each production bore using a modified genetic algorithm. The approach is applicable to regions with (1) partial records of historic groundwater usage at each production bore and (2) infrequent metered usage at each production bore. The usage time series are jointly calibrated with the other time-series model parameters, which include a constraint on partitioning of precipitation into evapotranspiration.

We applied the new approach to the Warrion Water Supply Protection Area (WSPA) in Victoria, Australia, and then evaluated the results against independent data. This region was selected because of the availability of groundwater level and high-frequency metered usage data, and because it provides an opportunity to evaluate the effectiveness of the approach at sites having complex recharge boundaries and a high density of licensed production bores.

In the following, the study area is first detailed followed by the approach for estimation of usage, aquifer hydraulic properties, and gross recharge. The method for assessing the reliability of the results is then outlined. The results section then summarizes the outcomes from the reliability assessments followed by the results from the application to the study area and the evaluation against independent data. Last, the discussion section critically examines the results and the conclusions section summarizes the key findings.

Study Area

The Warrion WSPA is in central Victoria, Australia, north of the Colac township, and approximately 150 km west of Melbourne. Lake Corangamite and Cundare Pool form the western and northern boundaries of the management area while the Colac-Cressy Road and the Princess Highway define the eastern and southern boundaries (Figure 1).

Groundwater extraction is managed according to the Warrion Groundwater Management Plan (GMP), which was approved by the Minister for Water in 2010.
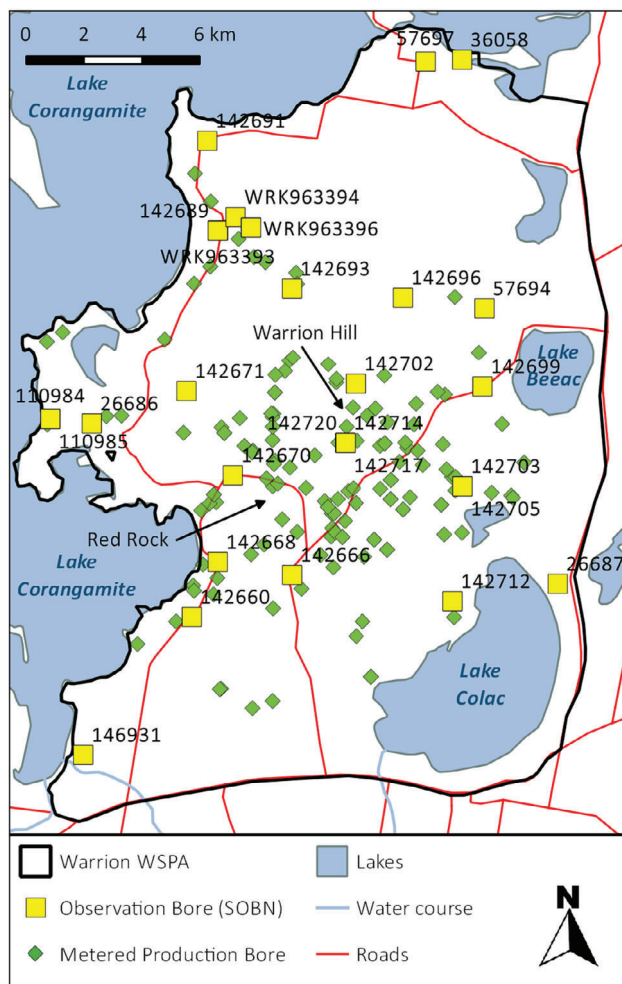
Figure 1. Study area observation and production bores.

As of June 2017, the management area has 132 licenses of which 86 are metered. The maximum total extraction volume is set at 14,086 ML/year and the total allocated volume is 14,078 ML/year. In the 2016-2017 reporting year, the reported metered volume used was 3,702 ML, which represents 26% of the maximum. In addition to the metered licenses, there are ~265 unlicensed stock and domestic bores with an estimated use of 398 ML/year (Southern Rural Water 2017), which represents 2.8% of the maximum. The dairy industry is the dominant groundwater reliant enterprise with groundwater also used for irrigation, commercial, domestic, and stock purposes. Metered groundwater extraction shows a strong seasonal pattern with most of the groundwater use occurring over the spring/summer period (November-March).

Climate

Meteorological data was sourced from the Bureau of Meteorology (BOM) Australian Water Availability Project (AWAP) gridded daily meteorological product using the *AWAPer* R-package (https://github.com/peterson-tim-j/AWAPer; Peterson et al. In press).

The mean catchment weighted precipitation is 640 mm/year and the mean annual potential evaporation

(Morton's PET) is 1240 mm/year. Average maximum temperature ranges from 25.9 degrees Celsius (°C) in January to 11.6 °C in July. July is the coolest month with an average minimum temperature of 5.3 °C. Rainfall occurs throughout the year but is highest in late winter/early spring (July to October). Peak potential evaporation occurs across the summer months (December to March) and coincides with both the period of lowest rainfall and the highest groundwater extraction. Average monthly rainfall only exceeds potential evaporation during the winter months from June to August.

Geology

The study area is situated in the Otway Sedimentary Basin, which comprises of a thick succession of Mesozoic and Cainozoic sedimentary and volcanic rocks that extends across western Victoria. The following geological units are mapped within the study area (Colac 1:250 000 Geological Sheet [GSV 1973]):

- Gellibrand Marl (Tertiary)—Predominantly grayish marl, with calcareous clay and silt, and clayey limestone. Regionally it has an observed thickness of up to 400 m.
- Moorabool Viaduct Formation (Tertiary)—Sand, variously ferruginous and calcareous, limestone, gravel, and laterite. Thickness ranges up to 30 m.
- Newer Volcanics (Tertiary to Quaternary)—"*Earlier Phase*" Tertiary volcanics are interbedded within and overly the Moorabool Viaduct Formation. They comprise of partially weathered sheet basalts which are typically 30-m thick but range up to 60 m in thickness close to eruption points and where lava flows infilled existing drainage lines. "*Later Phase*" Quaternary volcanics comprising vesicular lava flows and scoria, individual lava flows are typically between 2 and 10 m in thickness.
- Alluvial and Lacustrine deposits (Quaternary) comprising discontinuous aeolian lunette sands on the eastern shores of the lakes, lacustrine clay, silts, and sand across the central study area and alluvial river sands and clay along minor drainage lines.

Hydrogeology

The Newer Volcanics hosts the principal water supply aquifer with minor groundwater extracted from Quaternary alluvial deposits—where they have sufficient saturated thickness—and also from Tertiary sediments of the Moorabool Viaduct Formation. The latter is discontinuous across the study area but where it is present it is generally considered to be in hydraulic connection with the overlying volcanics aquifer (Coram 1996; SKM 2000).

The Newer Volcanics form an unconfined, fractured rock aquifer system ranging up to 60 m in thickness. Most of the groundwater is sourced from shallow depths with over 75% of production bores constructed at depths of less than 30 m. The Newer Volcanics aquifer is characterized by two distinct hydrostratigraphic units, pyroclastic flows, and scoria deposits in the elevated

areas around the eruption points of Warrion Hill and the Red Rock Complex (6 km WSW of Lake Beeac), and the stoney rise basalt plains which extend across the remainder of the study area. The scoria deposits around Red Rock/Mt Warrion are highly permeable and, although they cover less than 20% of the WSPA, around 80% of the licensed groundwater extraction volume is located within this zone. Hydraulically, the scoria deposits and basalt plains are considered to form a single, unconfined aquifer.

Recharge occurs via direct infiltration of precipitation with the greatest volume of recharge understood to occur on the more permeable scoria cones around Warrion Hill/Red Rock. Infiltration rates in this region have been estimated at up to 30% of precipitation (Thompson 1971), while recharge rates on the basalt plains are in the order of 5-10% of precipitation (Nolan 2002). The water quality distribution and groundwater flow pattern in the Newer Volcanics aquifer supports higher recharge rates around the Mt Warrion/Red Rock area. Groundwater quality in the Mt Warrion area is typically less than 500 mg/L (total dissolved solids—TDS) but deteriorates progressively away from the recharge zone and exceeds 3000 mg/L on the margins of the study area. Groundwater flows radially away from the recharge zone toward Lake Corangamite in the west and north, and the chain of smaller lakes in the east. The groundwater system has a complex relationship with the surrounding lakes and wetlands. Early studies (Thompson 1971) concluded that the lakes were groundwater discharge points, the rate of groundwater discharge estimated by Thompson (1971) was later used to determine the original allocation volume for the Warrion groundwater management area (SKM 1998). More recent work (Tweed et al. 2008) suggests a more complex relationship between the lakes and the groundwater system where the dynamics and function of the lake (recharge, discharge, and throughflow) changes in response to meteorological forcing.

Methods

This section first details the algorithm for downscaling infrequent metered groundwater usage to daily usage and the estimation of unmetered usage. The approach for recharge estimation is then detailed, followed by the methods for assessing the reliability of the approach. This involved assessment of (1) the reproducibility of the transmissivity, storativity, and recharge estimates when a model is calibrated many times and (2) the error in the downscaled groundwater usage over a range of time scales. Additionally, the Supporting Information presents assessment of (1) the variability in the calibrated downscaled daily groundwater usage, and subsequently estimated transmissivity, storativity, and recharge using different calibration settings and (2) the variability of the solutions to the radius of influence (i.e., the radius of the pumping drawdown cone). Last, the method for application to all the observation bores within the study area is detailed.

Estimation and downscaling of metered groundwater usage

The algorithm for the estimation and downscaling of metered groundwater usage was developed to enable the joint estimation of daily groundwater usage volumes at any number of production bores with minimal assumptions or prior knowledge about the behavior of groundwater users or their response to meteorological conditions. That is, the daily groundwater usage time series was estimated from the observation data. To achieve this, a genetic evolutionary parameter optimization scheme called *SP-UCI* (**S**huffled complex evolution with **P**rincipal components analysis—**U**niversity of **C**alifornia at **I**rvine, Chu et al. 2011) was recast from simply *evolving* a population of randomly selected model parameter sets toward a global optima (i.e., minimum error) to also *evolving* time series of daily production volumes at multiple bores toward a minimum model error with the observed groundwater hydrograph. To our knowledge, this is the first approach to jointly estimate parameters and time series. The approach was developed with the following objectives:

- Minimal prior knowledge is required about when groundwater extraction occurs. That is, no assumptions were required about the crop type or when irrigation occurs within a week, season, or year.
- No more than one calibrated parameter was required per production bore. This was required to ensure the global calibration of the model remains numerically feasible.
- Applicable to both downscaling infrequent metering and the estimation of usage during unmetered periods; that is, periods after construction of a production bore but prior to the commencement of metering.
- The output time series of usage is a single best estimate of usage. Ensemble time-series estimates of usage were not sought.

These objectives were met by first assuming that each pump has a fixed, but unknown, flow rate and that the daily duration of pumping is fixed between meter readings but can vary from one reading to the next. Numerically, this was achieved by deriving a binary daily time series of each production bore being either "on" or "off." During metered periods, the daily usage rate was estimated by dividing the metered volume by the number of days the pump was on. During unmetered periods, the daily usage was also unknown but was estimated from the median estimated usage rate during metered periods. To estimate whether the pump was on or off, a model parameter iterated through the time-series switching the pumping state. If the change in pumping state improved the fit to the observed hydrograph, then it was accepted. The approach required each pump to have at least one period of metered volumetric usage. As mentioned prior, an important assumption was that, for a given metered period that was to be downscaled, the daily extraction volume was assumed to be constant. Similarly, at a given production bore, the daily extraction

volume during unmetered periods was constant within each period and across all unmetered periods. These assumptions are obviously a simplification of human behavior and the degree of simplification will increase with the duration between meter readings because the temporal variability in the estimated daily usage will decline. However, qualitative knowledge of groundwater usage can be incorporated into the approach by assigning zero usage to periods that are known to have no pumping. In this application, usage was assumed to be zero between the wetter months of May and October. In applications where usage varies seasonally, and groundwater-level observations are infrequent, meter readings should be sufficiently frequent to allow derivation of a seasonally varying estimate of daily usage.

Importantly, the number of combinations of daily pumping state (i.e., "*on*" or "*off*") over, say, a decade far exceeds the largest unsigned long integer (18E18). The implications are that the sequential assessment of each combination of pumping state time series at each production bore, combined with the parameter estimation, is computationally near impossible. The approach developed is therefore a computationally tractable approximation. Consequently, if an identical model was recalibrated, then an alternate optimal solution may be identified, which may have a slightly different time series of pumping state and model parameters. To reduce the number of combinations to assess, and to increase the solution reproducibility, the approach starts by downscaling and estimating the pumping states at a coarse time step; herein annual. That is, each pump was assessed as being "on" or "off" for the entire period (e.g., the irrigation season). Upon convergence to an optima, this sequence of pumping states was then refined by recalibrating using a monthly time step for the downscaling and estimation. Similarly, the monthly sequence of pumping states was then refined by recalibrating to a weekly time step and then to a daily time step. For details of the algorithm see Figure S1.

The downscaling and estimation algorithm increased the calibration time of *HydroSight* models by over an order of magnitude. Significant effort was put into assessing and refining the computationally demanding components of the approach. However, on a 6-core 4GHz workstation, 6 h was required for the global calibration of a single model with 12 production bores having annual meter readings. Where such computational times are impractical, reducing the number of production bores requiring downscaling significantly reduces the computational time. Alternatively, changing the final downscaled pumping states from daily to, say, monthly also significantly reduces the computational time. This feature is provided within *HydroSight*, but the reliability of the estimates to the final downscaled time step is not investigated within this paper.

Estimation of recharge

Prior versions of *HydroSight* provided an estimate of gross recharge. Specifically, when a transfer TFN model includes a soil moisture model with vertical free drainage (Peterson and Western 2014) then, following calibration, the free drainage can be considered as an estimate of the gross recharge (i.e., phreatic evapotranspiration and discharge to surface waters are not subtracted from the estimate, Doble and Crosbie 2017). To account for the diffusion and time lag between the free drainage and its percolation through the unsaturated zone to the water table, a convolution function is then applied. Consequently, if a time series of gross recharge is required, the free-drainage time series should be averaged over a period greater than the time lag.

The long-term average free-drainage estimate is constrained by the long-term average precipitation. Fitting of the convolved free drainage to the observed hydrograph further constrains the estimate. However, in undertaking trials for this paper, we found these constraints were inadequate such that the annual free drainage was often an implausibly high fraction of the annual precipitation. Considering that the groundwater head fluctuation is estimated from the convolved free drainage minus the convolved pumping drawdown, and that the observed head fluctuation is fixed at a given time point, an overestimation of the free drainage would require a corresponding and implausible increase in the drawdown to produce the same change in head; which would be achieved by reducing the estimates of aquifer transmissivity and/or storativity. Consequently we sought hydrological *laws*, or the closest findings hydrology has to *laws*, to constrain the long-term free drainage to a plausible range.

Recently Obergfell et al. (2019) used seasonal forcing harmonics to constrain time-series estimates of head and mean recharge. We, however, sought an approach that required few assumptions about the interaction of precipitation, evapotranspiration and recharge and that was based on known long-term catchment behavior. The most appropriate *law* that we found was based on the Budyko curve (Zhang et al. 2004), which defines the expected long-term average actual evaporation for a given catchment aridity. Specifically, the curve describes the ratio of mean annual actual evapotranspiration (E) to precipitation (P) as a function of the aridity (potential evapotranspiration, E_o, over P) using only one parameter, w. In applying the curve, Zhang et al. (2004) collected mean precipitation, potential evapotranspiration, and estimated actual evapotranspiration for 470 catchments throughout Australia and, by fitting the curve, estimated w as 2.63.

The relevance of the Budyko curve to *HydroSight* recharge estimation is that the Zhang et al. (2004) equation provides an estimate of the expected actual evapotranspiration. The *HydroSight* soil moisture module also provides an estimate of the actual soil evapotranspiration, which we will call E_{HS}. Hence, when E_{HS} is significantly different from E then the *HydroSight* soil moisture model parameters can be deemed as implausible. Conversely, when E_{HS} is similar to E, then the parameters can be deemed as plausible, and consequently the recharge is hypothesized as being more plausible.

Estimating when E_{HS} is sufficiently similar to E was achieved using estimates of the uncertainty in the

Budyko curve for any value of aridity (Greve et al. 2015). They achieved this by identifying a right-skewed gamma distribution for the plausible w parameter values using hydrological data from 411 MOPEX study catchments in the United States. By randomly sampling 1000s of w values and then calculating the Budyko curve for each, bounds for the Budyko curve could be derived and, more importantly for this application, a distribution for E/P could be derived for any aridity.

Using the distribution for E/P for a given aridity, plausible bounds for the actual evaporation could then be derived; which herein were set to the 10th and 90th percentiles. In applying these bounds, we assumed that the 411 US catchments are representative of the long-term dynamics in south-eastern Australia. If future applications required narrower bounds, then Budyko extensions that account for local climatic, soil and vegetation factors (e.g., Donohue et al. 2012) could be used if such extensions provided an uncertainty distribution for E.

The 10th and 90th percentiles for E were incorporated into *HydroSight* by the construction of two new soil moisture models, a one-layer model and a two-layer model, as follows:

1. The gamma distribution for parameter w was defined with, as per Greve et al. (2015), the shape and scale parameters set to 4.54 and 0.37, respectively.
2. Ten thousand random samples of w were drawn from a gamma distribution.
3. For each sampled w and the catchment aridity (from the observed mean precipitation and Morton's areal potential evapotranspiration), the expected ratio of actual evapotranspiration to precipitation was derived for each of the random samples and then its 10th and 90th percentiles were calculated.
4. During model calibration, a candidate parameter set was derived from the simplex SP-UCI step (i.e., a multidimensional triangle is built from accepted parameter sets and used to predict a new parameter set going from the worst parameter set to between the best two).
5. E_{HS} over the calibration period was then estimated for the candidate parameter set. If, when divided by precipitation, the ratio was outside the bounds from Step 3, then the candidate parameter set was deemed to be implausible. Its objective function value was set to infinity and then a new candidate parameter set was tested at the centroid of the simplex.

In applying this approach, very few one-layer soil model parameters were found to meet the aforementioned constraints; and allowing the modeled evapotranspiration to have a nonlinear relationship with the fractional soil moisture did not resolve the issue. Instead, the soil moisture module was extended to include two serial stores (Figure 2) where the drainage from the upper store drained into the deep store, ET demand was first supplied by the upper store and the remaining demand was supplied by the deep store. Also, the evapotranspiration from each
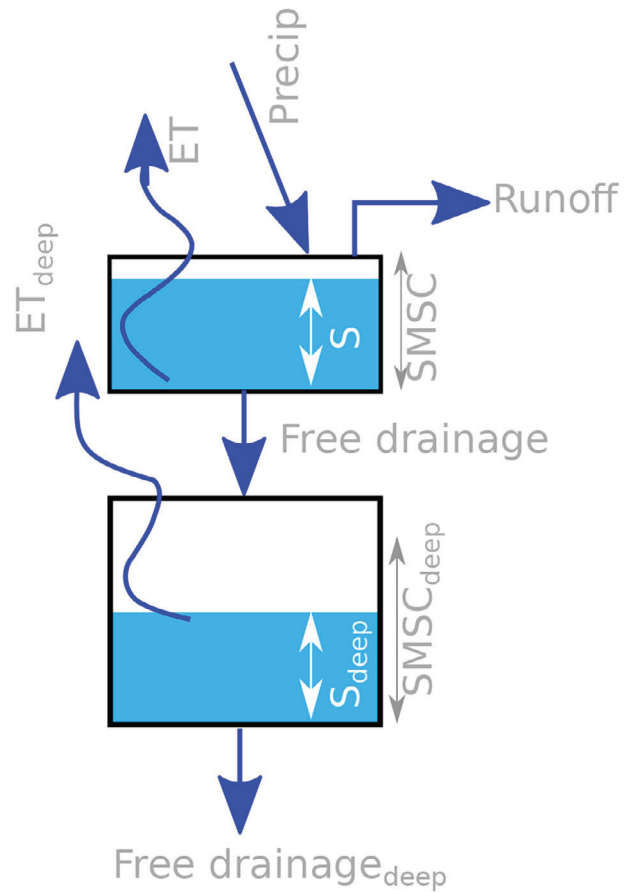
Figure 2. Two-layer soil moisture module.

layer had a nonlinear relationship with the fractional soil moisture. These modifications provided the flexibility to meet the constraints on the modeled *actual* ET.

Following Peterson and Western (2014), the upper layer soil moisture was modeled using Equation 1. The deep layer soil moisture was modeled using Equation 2:

$$\frac{dS}{dt} = P_{inf}\left(1 - \frac{S}{SMSC}\right)^{\alpha} - k_{sat}\left(\frac{S}{SMSC}\right)^{\beta} - PET\left(\frac{S}{SMSC}\right)^{\gamma} \tag{1}$$

$$\frac{dS_{deep}}{dt} = k_{sat}\left(\frac{S}{SMSC}\right)^{\beta} - k_{sat,deep}\left(\frac{S_{deep}}{SMSC_{deep}}\right)^{\beta_{deep}} - \left(PET - \left(\frac{S}{SMSC}\right)^{\gamma}\right)\left(\frac{S_{deep}}{SMSC_{deep}}\right)^{\gamma_{deep}} \tag{2}$$

where S and S_{deep} are the soil moisture in the upper and lower stores respectively at time, t; $SMSC$ and $SMSC_{deep}$ are the soil moisture capacity in the upper and lower stores respectively; P_{inf} is the daily infiltrating precipitation at t (i.e., excludes infiltration excess runoff); PET is the daily evapotranspiration at t; k_{sat} and $k_{sat,deep}$ are the vertical saturated hydraulic conductivity in the upper and deep stores respectively; and α, β, β_{deep}, γ, and γ_{deep} are

power terms controlling the nonlinearity in the infiltration, vertical drainage and ET respectively. For this paper, the calibrated soil parameters were $SMSC$, $SMSC_{deep}$, k_{sat}, β, and γ. The parameters $k_{sat,\,deep}$, β_{deep}, and γ_{deep} were set to equal k_{sat}, β, and γ. Additionally, α was fixed at 0. These simplifications were undertaken to increase the reproducibility of the calibration; however, more flexible options are available within *HydroSight*. These simplifications assume that (1) all precipitation is infiltrated, unless the upper store is full; (2) the upper and deep stores have identical hydraulic properties and differ only in their soil moisture storage capacity.

Estimation of transmissivity, storativity, and recharge reproducibility

The joint calibration of *HydroSight* model parameters and the downscaling and estimation of daily usage is a highly unusual optimization problem. Foremost, at each iteration of the model parameters, the time series of daily groundwater usage is updated. This has the implication that if the objective function was recalculated using the parameters from the prior iteration, the new objective function value would likely differ from its original value. Hence, the updating of the groundwater usage time series makes the calibration scheme highly dependent upon the sequence of parameter sets that are assessed.

To assess the significance of this complication on the reproducibility of solutions, models were built for four bores surrounding the region of greatest usage (142666, 142670, 142699, and 142702) using the structure as detailed in the following subsection and with all production bores within a 2-km radius. Each model was then calibrated 50 times with a different random seed for each calibration. For each of the 200 models, an estimate of transmissivity, storativity, and mean recharge was derived. Boxplots were then derived showing the variability in each estimate at each bore. Importantly, this analysis does not provide an estimate of the uncertainty and only informs the variability in estimates that can be expected if the calibration is repeated.

Evaluation of the error in the downscaled and estimated usage

To assess the reliability of the estimates of groundwater usage, a model was built for observation bore 142714 using the structure as detailed in the following subsection and with all production bores within a 2-km radius ($n = 12$). Bore 142714 was selected because it has a long-term water-level record and is centered within a region of dense groundwater production bores.

From January 1, 2017 to June 22, 2017, nine production bores near to observation bore 142714 had daily metered usage. To quantify the error in the downscaling, the daily usage was summed to weekly, monthly, quarterly, and biannual volumes and then downscaled to daily by the recalibration of the model for observation bore 142714. This was repeated 10 times and then the daily usage error from each aggregated time step was estimated at each pump. Importantly, the use

of 10 model trials is modest. Ideally, 100+ trials would have been undertaken. However, to do so would have increased the number of model calibration from 40 to 400 and increased the calibration time from 10 days (6 h per model) to 100 days.

Modeling of all study area observations bores

Models of the following form were built for each of the 30 observation bores within the study area. Prior to the modeling, each groundwater hydrograph was analyzed for erroneous and outlier observations using Peterson et al. (2018). All models were then calibrated over the full observation record length to December 31, 2017 using all production bores within 2-km of each observation bore and the calibration used 16 SP-UCI complexes (see Figures S2 and S3 for an assessment the latter user decisions). Each model was calibrated only once. Below are details of additional settings:

- The two-layer soil model with constrained soil evapotranspiration was used to produce a time series of free drainage from the deep layer.
- Pumping downscaling and estimation was undertaken using the calibration scheme outlined above and detailed in Figure S1.
- Convolution of the free drainage was done using the three-parameter Pearson's *HydroSight* function.
- Convolution of the downscaled pumping at each production bore was done using the *Ferris-Knowles* drawdown function (Ferris and Knowles 1963), which is an instantaneous form of the Theis equation. Following Shapoori et al. (2015b), Ferris-Knowles was reformulated as an impulse response function for the drawdown per unit pumping volume:

$$\theta_F(t) = \frac{\alpha}{t} \exp\left(\frac{-r^2\beta}{t}\right) \qquad (3)$$

where r is the distance between the observation bore and the production bore and α and β are calibration parameters. Following Shapoori et al. (2015c), transmissivity (T) and storativity (S) were then estimated:

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

$$S = 4T\beta. \qquad (5)$$

- The drawdown from all production bores for a given observation bore were modeled using a single convolution function—and hence assumed to all have an equal transmissivity and storativity. Neither recharge or no-flow boundaries were included in the drawdown function. Lake Corangamite is likely to act as a recharge boundary. However, because it is spatially nonlinear and because of the large number of production wells, the simulation of the boundary using the standard approach of image wells was considered overly complex.

Importantly, at some observation bores, the drawdown from pumping may be very modest. Modeling such bores with a drawdown component would probably produce highly uncertain estimates of transmissivity, storativity, and the downscaled usage time series, and potentially also the meteorological forcing related parameters and fluxes. To assess where observation bores are best modeled with, and without, production bores all 30 observation bores were also modeled using the above form but without the pumping downscaling and *Ferris-Knowles* weighting functions. Upon calibration of all models, those observation bores best modeled with, or without, pumping were identified by the model having the lowest *Akaike information criterion—corrected* (AICc); which penalizes the objective function model fit to the observed hydrograph by the number of model parameters and number of observation points, and hence allows the selection of the most parsimonious model. The AICc was calculated from (Burnham and Anderson 2002):

$$AICc = 2k + 2Ln(\widehat{L}) + \frac{2k(k+1)}{N-k-1} \qquad (6)$$

where k is the number of calibrated model parameter, N is the number of groundwater-level observations, and $\widehat{L}$ is the maximized likelihood value; which was approximated using the modified least-squares objective function from Peterson and Western (2014) (adapted from von Asmuth and Bierkens 2005) and which accounts for serial correlation in the model residuals and irregular observation time steps.

Results

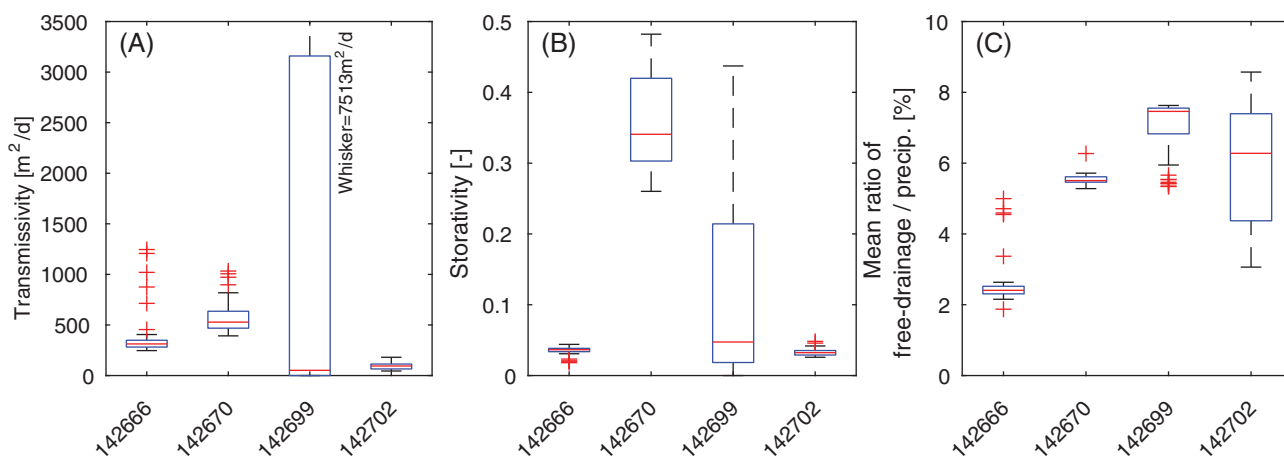### Estimation of transmissivity, storativity, and recharge reproducibility

Figure 3 shows the variability in the estimated transmissivity, storativity, and mean recharge at four observation bores when the model for each bore was calibrated 50 times. For three bores, the variability in the transmissivity estimates were modest (142666, 142670, 142702) with an interquartile range varying from 49 m^2/d (142702) to 167 m^2/d (142670); or when the interquartile range is divided by the median, varying by a factor of 0.22 (142666) to 0.51 (142702). For the fourth bore (142699), the interquartile range was extremely large (3,159 m^2/d) and varied by a factor of 61.

For two bores, the variability in the storativity estimates was low with an interquartile range varying from 0.0047 (142666) to 0.0063 (142702); or when the interquartile range is divided by the median varying by a factor of 0.13 (142666) to 0.2 (142702). For the other two bores, the interquartile range increased to 0.12 (142670) and 0.2 (142699), or by a factor of 0.34 and 4.1, respectively.

For three bores, the variability in mean recharge (as a percentage of mean annual precipitation) was low, with an interquartile range varying from 0.15% (142666) to 0.73% (142699); or by a factor of 0.03 (142670) to 0.1 (142699). For the fourth bore, the interquartile range was higher at 3% (142666), or by a factor of 0.48.

Overall, the transmissivity and recharge estimates appear to be the most reproducible variables and so undertaking a single calibration is likely to produce a well constrained estimate. However, the reproducibility of storativity estimates appears bore dependent. Looking at Figure 1, the bores with highly variable storativity estimates (142670, 142699) substantially differ in the number of pumping bores within the radius of influence (21 and 4 respectively) but are both near lakes. Conversely, the bores with low variability in the storativity estimates (142666, 142702) are not near lakes. Considering that the only highly variable estimate of transmissivity (142699) is also near a lake, this suggests that recharge boundary effects increase the variability in aquifer property estimates. Consequently, the estimates from a single model calibration appear most reliable when the observation and pumping bores are not influenced by recharge boundaries. In summary, the calibration scheme produced generally reproducible estimates of transmissivity, storativity,

Figure 3. Box-plots plots of (A) transmissivity, (B) storativity, and (C) mean recharge at four observation bores surrounding the region of greatest usage. The estimates were derived by calibrating each model 50 times.
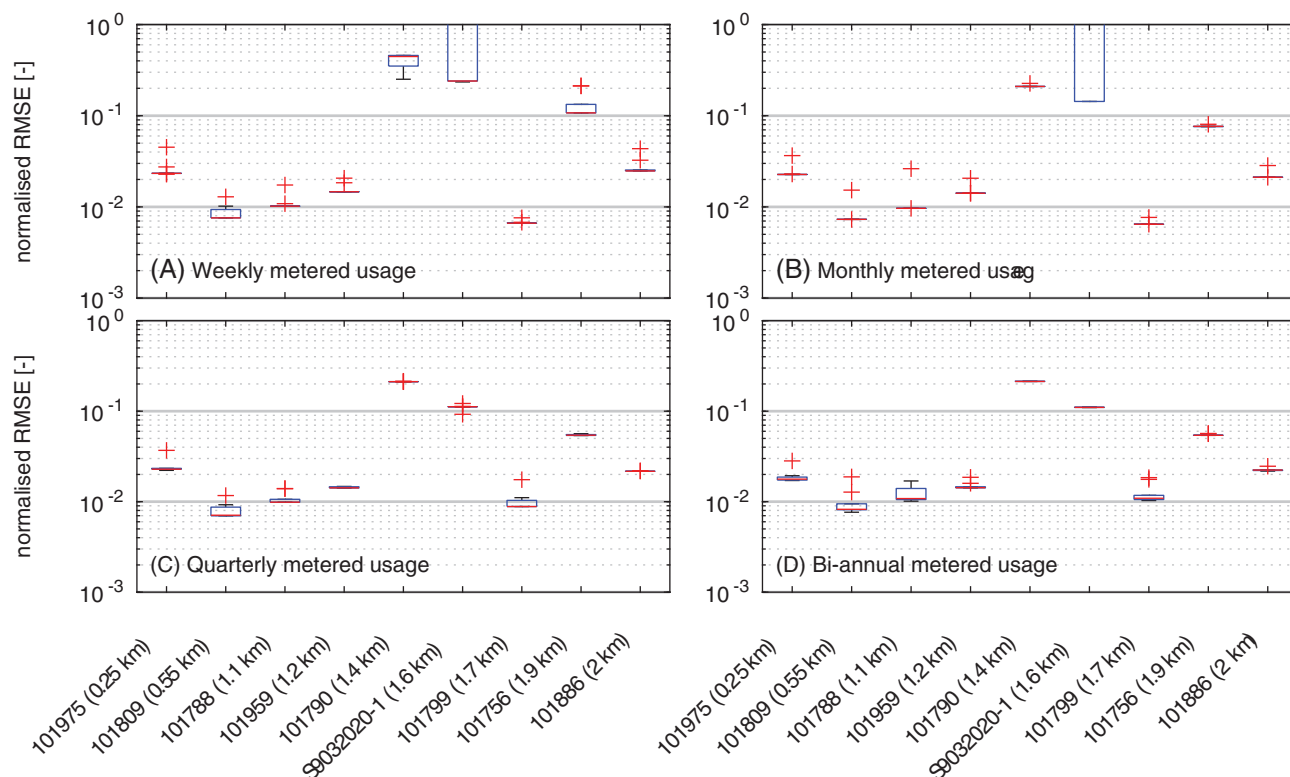
Figure 4. Box-plots plots of the normalized root mean square error in the downscaled metered usage at nine pumps from 10 replicates when the daily metered usage from January 1, 2017 to June 22, 2017 was aggregated to (A) weekly usage, (B) monthly usage, (C) quarterly usage, and (D) biannual usage and then downscaled. The analysis was undertaken for only observation bore 142714. The sum of the daily metered usage over the aforementioned dates was adopted for the normalization. Pumps 101933 and 101841 were removed from the analysis because the daily metered usage was zero. Note, pumps 101790 and S9032020 had a very low metered total usage of 55 and 30 m^3, respectively.

and mean recharge that varied by a factor of 0.22-0.32, 0.13-0.2, and 0.03-0.48, respectively when boundary effects were minimal.

Evaluation of the error in the downscaled and estimated usage

Figure 4 summarizes the error in the downscaled usage estimates. It shows the root-mean-square-error in usage normalized by the total metered usage from January 1, 2017 to June 22, 2017 for the four time scales and at the nine pumps. It shows that for the downscaling of monthly, quarterly, and biannual metered usage the error was <10% at seven of the nine pumps and approaching 1% at three pumps (see Figure 4A, C, and D). For monthly metered usage, the error was <10% at six of the nine pumps. Furthermore, at all time steps, the variability from the 10 trials was very low at all but weekly and monthly metering at production bores 101790 and S9032020, which had a very low metered total usage of 55 and 30 m^3, respectively, that caused small usage errors to be inflated to produce a large normalized RMSE. To further illustrate the downscaling, SI Figures S57-S60 show scatter plots of the observed daily usage against estimated daily usage for each data point used in Figure 4. Similarly, SI Figures S61-S64 and SI Figures S65-68 show scatter plots of the weekly and

monthly usage volumes, respectively. Overall, they show that daily usage estimates may be acceptable only when weekly metering is available. However, the weekly usage estimates are often reasonable when monthly and quarterly metering is available. Similarly, monthly usage estimates are often reasonable when quarterly and biannual metering is available.

Study area analysis

HydroSight groundwater hydrograph time-series models were built and calibrated for each of the 30 observation bores shown in Figure 1. For each bore, three models were produced: (1) meteorological-only forcing using the one-layer soil model from Peterson and Western (2014), (2) meteorological-only forcing using the two-layer soil model with the constraint on soil evapotranspiration, and (3) the prior two-layer model with downscaled metered usage. Comparing the first two models informed if the addition of a second soil moisture layer and the evapotranspiration constraint improved the hydrograph fit. Comparing the latter two informed if the simulation of pumping improved the hydrograph fit. For each comparison, the most parsimonious model was identified by the minimum AICc.

The comparison of the two models without pumping is summarized in Table 1. It shows the AICc from the

Table 1

Comparison of HydroSight AICc and Evidence Ratio for the One-Layer Soil Model and the Two-Layer Soil Model with the Constraint on Evapotranspiration

Bore ID	AICc (-)		Evidence Ratio (-)	
	1-Layer	2-Layer	2-Layer over 1-layer	1-Layer over 2-Layer
142660	**107.4**	107.6		1.1
142671	−226.7	**−242.7**	3016.2	
142689	−154.2	**−163.8**	122.3	
142693	−54.3	**−72.3**	8231.3	
142705	**−566.6**	−563.4		4.9
WRK963392	**−154.6**	−151.8		4.1
WRK963393	**−159.1**	−154.9		8.0

Notes: The lowest AICc model is in bold (i.e., most parsimonious). The evidence ratio shows the improvement of the best model to the other. An evidence ratio <10 indicates very little improvement of the best model to the other.

bores found not to be influenced by pumping. The relative difference between the AICc values is also shown using the *evidence ratio* (Burnham and Anderson 2002):

$$ER = \exp\left(\frac{AICc - AICc_{min}}{2}\right) \qquad (7)$$

where $AICc_{min}$ is the minimum AICc from the two models. While the evidence ratio has no threshold for significance, like a *t*-test, a value less than approximately 10 indicates very little support for the best model relative to the alternate model (Burnham and Anderson 2002). Looking at Table 1, the evidence ratio indicates that when the two-layer model was best it was substantially better than the one-layer model. Conversely, when the one-layer model was best, it offered only a very slight improvement over the two-layer model. This indicates that the two-layer model is the most, or almost the most, parsimonious model for all seven bores found not to be influenced by pumping. That is, the two-layer model generally produces the best fit to the observed hydrograph while also producing plausible estimates of evapotranspiration. Henceforth, only the two-layer model is further considered.

Figure 5 shows results from the comparisons with and without pumping. It shows the coefficient of efficiency from the most parsimonious model at each observation bore, where one denotes a perfect fit and zero denotes a fit no better than predicting the observed mean. It shows that the CoE was ≥0.80 at 22 bores and ≥0.90 at 12 bores. Further details of each model calibration are provided in SI Figures S4-S33 and the estimated contribution from meteorological forcing and pumping drawdown from each production bore are shown in SI Figures S34-S56.

Where the meteorological-only model was parsimonious, groundwater pumping can be considered as having no discernible impact. However, for completeness, the transmissivity (T) and storativity (S) were estimated at these sites using the alternate pumping and meteorological model and were included in the following mapping,

but are designated with a cross overlaid on the bore location.

Transmissivity

The estimated T for the observation bores best modeled using pumping and meteorological data was 0.1 to 2424 m²/d (median of 354 m²/d). Figure 6 shows the estimates. Table 2 compares the estimated transmissivity values with conventional transmissivity values from pumping tests. Twenty-six point estimates of transmissivity were identified in the literature. Pumping tests were available from Thompson (1971) and Nolan (2000). The Thompson (1971) tests were largely short duration (6-h) tests without observation bores except for one 48h test using a single monitoring bore. The Nolan (2000) test was of 48h duration with three observation bores—this is considered the most reliable test/transmissivity estimate. All pumping tests are spatially constrained to a relatively small area in the east of the WSPA. Figure 6 and Table 2 also include point estimates of transmissivity from Coram (1996). These were derived from drawdown analysis and limited pumping test data using Hurr (1966). There is significantly greater uncertainty with these point estimates as they make assumptions about test duration and specific yield.

The harmonic mean *HydroSight* transmissivity (84 m²/d) in Table 2 compares favorably with the mean values from the conventional analysis (58-287 m²/d). The upper range of the *HydroSight* estimates (2424 m²/d) is around 150% higher than the maximum T from pumping tests (935 m²/d) and 40% higher than the maximum from Coram (1996)—1754 m²/d.

These upper range *HydroSight* T estimates are largely based in the center of the WSPA at a volcanic eruption point characterized by highly permeable scoria deposits. Near this point are the nested monitoring bores 142714, 142717, and 142720, which are constructed in the shallow volcanics, deeper volcanics, and underlying Tertiary sediments, respectively. *HydroSight* T estimates increased with bore depth from 1413 m²/d in the shallowest bore (142714, screens 20-23 m) to 2013 m²/d in the
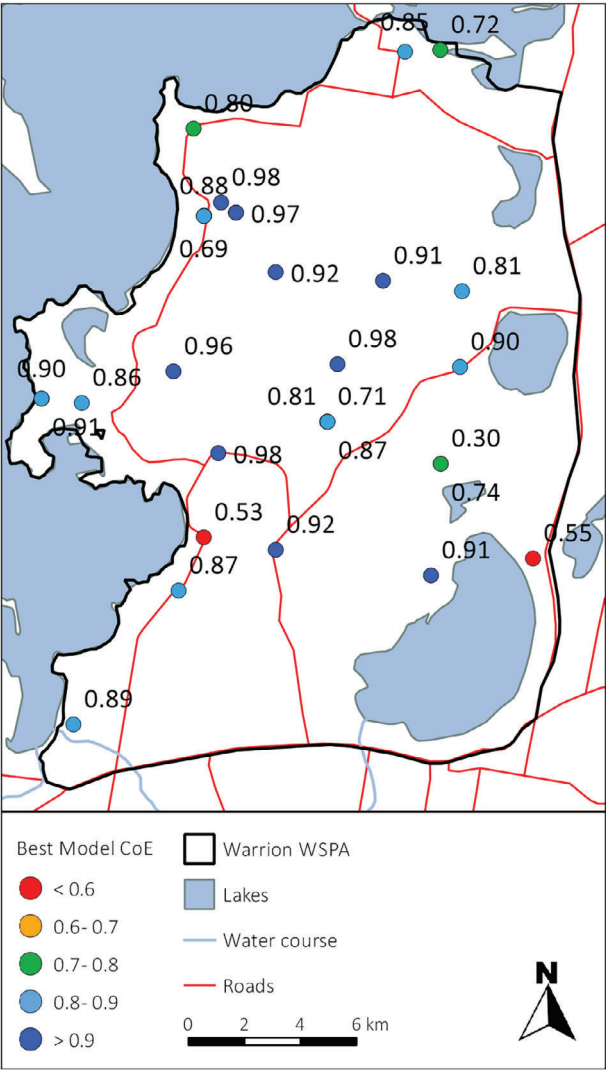
T.J. Peterson and S. Fulton Groundwater 57, no. 6: 860–876   869

Figure 5. Coefficient of efficiency (CoE) fit to the observed hydrograph as derived from the most parsimonious model.

Figure 6. Comparison of *HydroSight* estimations of transmissivity against independent estimates.

intermediate bore (142717, screens 47-50 m) to 2424 m^2/d in the deeper bore constructed in Tertiary sediments (142720, screens 66-69 m). The most permeable sequence typically occurs in the shallow volcanics, so the trend of increasing T with depth appears counter intuitive and
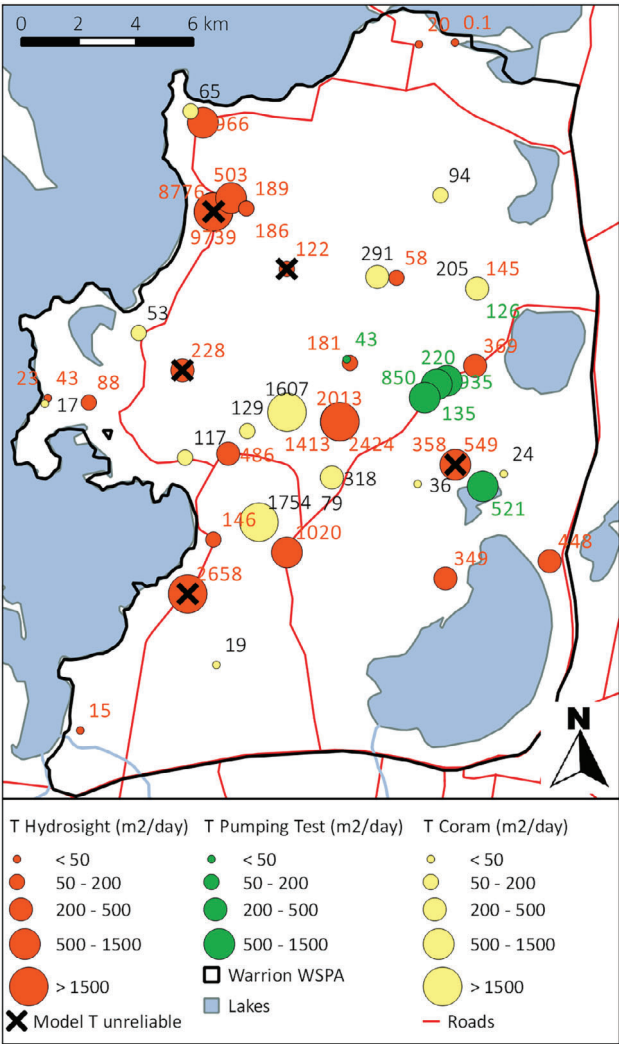
the magnitude of T for the Tertiary bore (142720) is unrealistically high. Hence, the upper range of T for the volcanics is 2013 m^2/d.

There are no pumping test results within the eruption point, but a nearby estimate from Coram (1996)

Table 2
Comparison of HydroSight Transmissivity with Values from Conventional Pumping Tests and Analytical Solutions

|	Transmissivity (m²/d)			
Source	Min.	Max.	Harmonic Mean	Comment
HydroSight modeling	0.1	2424	84	Excludes models where the inclusion of pumping did not improve model performance ($n = 23$)
Thompson (1971) Pump tests	43	935	111	6-h single bore tests. Bores are clustered in the same area ($n = 7$)
Nolan-ITU (2000) Pump test	135	850	287	48-h pumping test with three obs. bores. Results vary significantly with distance & direction. ($n = 3$)
Coram (1996)	17	1754	58	Analysis Gov. data using Hurr (1966) method. ($n = 16$)
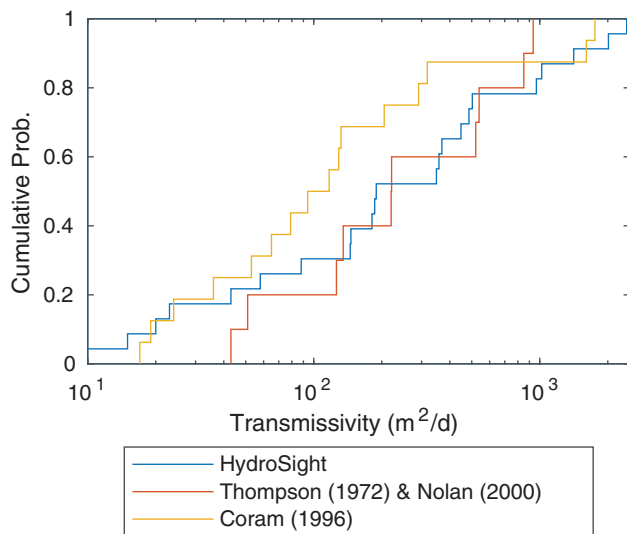
Figure 7. Cumulative histogram of *HydroSight* transmissivity estimates vs. conventional estimates from pumping tests and other analytical methods.

places the transmissivity at 1607 m²/d. This suggests that the *HydroSight T* estimates for this location are reasonable. Also, Shapoori et al. (2015c) demonstrated that *HydroSight* estimates of *T* are typically higher than *T* estimates from one-day conventional pumping tests. This was found to be due to the *HydroSight* modeling acting like a long-term pumping test that characterizes aquifer behavior on the scale of years. In contrast, most of the study area pumping tests are of short duration (6h) and may not reflect aquifer behavior at steady state conditions.

Figure 7 shows the empirical cumulative histogram for estimates from *HydroSight*, pumping tests (Thompson 1971 and Nolan 2000) and analytical approaches (Coram 1996). It shows that approximately 60% of the *HydroSight* and pumping test estimates have a very similar distribution; that is, when the transmissivity is less than approximately 1000 m²/d. Relative to Coram (1996), Figure 7 shows *HydroSight* to be comparable only above approximately 1000 m²/d. To statistically test if the median estimates are equal (null hypothesis), the Wilcoxon rank sum test was undertaken using a 5% significance. Comparing the *HydroSight* estimates against the pumping tests gave a *p*-value of 0.87, indicating that there is insufficient evidence to reject the null hypothesis. Similarly, comparing the *HydroSight* estimates against Coram (1996) indicated that the null hypothesis could not be rejected ($p = 0.17$). Overall, these results suggest that the *HydroSight* point estimates are samples from a similar distribution as the pumping test estimates.

Storativity

The estimated *S* for the observation bores that were best modeled using pumping and meteorological data ranged from 2.5×10^{-4} to 0.32 (median of 4.3×10^{-3}). Figure 8 shows the spatial distribution of storativity values across the study area. Bores have been nominally
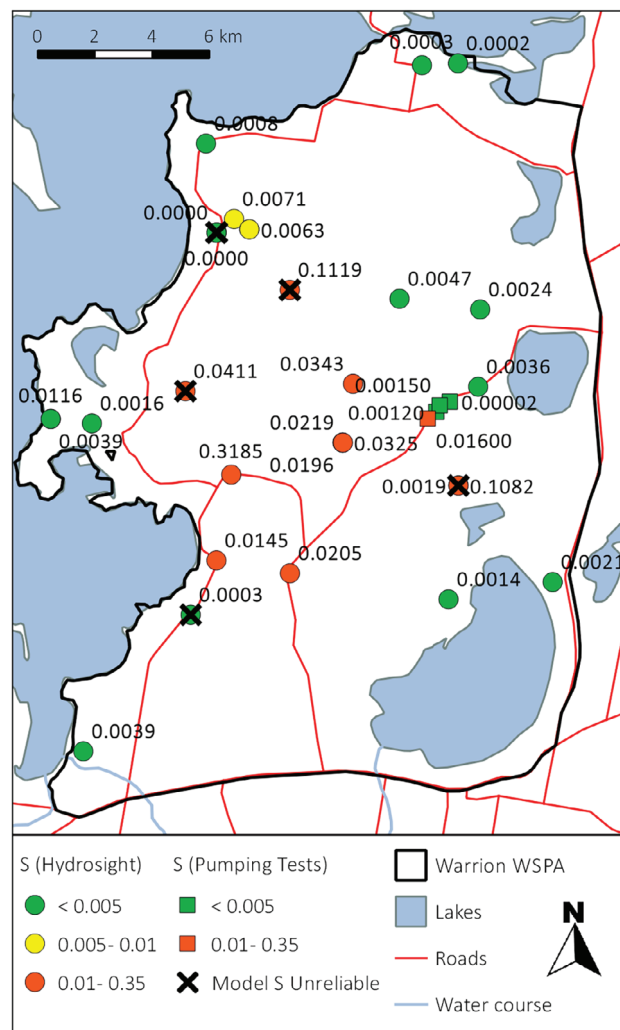

Figure 8. Comparison of *HydroSight* estimations of storativity against independent estimates.

assigned as unconfined, semiconfined, and confined based on storativity values adopted from Kruseman and de Ridder (1994). The unconfined model results are largely constrained to the area around Warrion Hill and Red Rock, where the volcanic aquifer is characterized by scoria deposits. These results are consistent with the existing conceptual model of the Warrion aquifer system, which identifies this area as more transmissive and with greater storage capacity than the surrounding basalt plains aquifer (Nolan 1999).

To independently evaluate these estimates, there are only a limited number of pumping test-derived storativity estimates available and all are from the same area to the west of Lake Beac. Thompson (1971) reports on a storage value of 1×10^{-5} from a 6-h pumping test but concludes that this result is not representative of the basalt aquifer and assumed a value of 0.05. A 48-h pumping test by Nolan (2000) utilizing three observation bores produced a storativity ranging from 0.0012 to 0.016. A pumping test on bore 57517, constructed in the Newer Volcanics and located 10-km north of the study area reported a storativity of 1.5×10^{-4}.

Table 3
Comparison of HydroSight-Derived Storativity with Values from Conventional Pumping Test Analysis and Reported Literature Values

Source	Storativity (-)			Comment
	Min.	Max.	Arith. Mean	
HydroSight modeling	2.5E-4	0.32	4.3E-3	Excludes models where the inclusion of pumping did not improve model performance. ($n = 23$)
Study area pump tests	2E-5	0.016	1.5E-3	Estimates from two pump tests both located in the same area. ($n = 4$)
Other Newer Volcanics pump tests	1.5E-4	4.5E-3	1.2E-3	($n = 3$)
Literature values	0.01	0.3	—	Based on generalized estimates of specific yield in basalt aquifers, source of values not available.
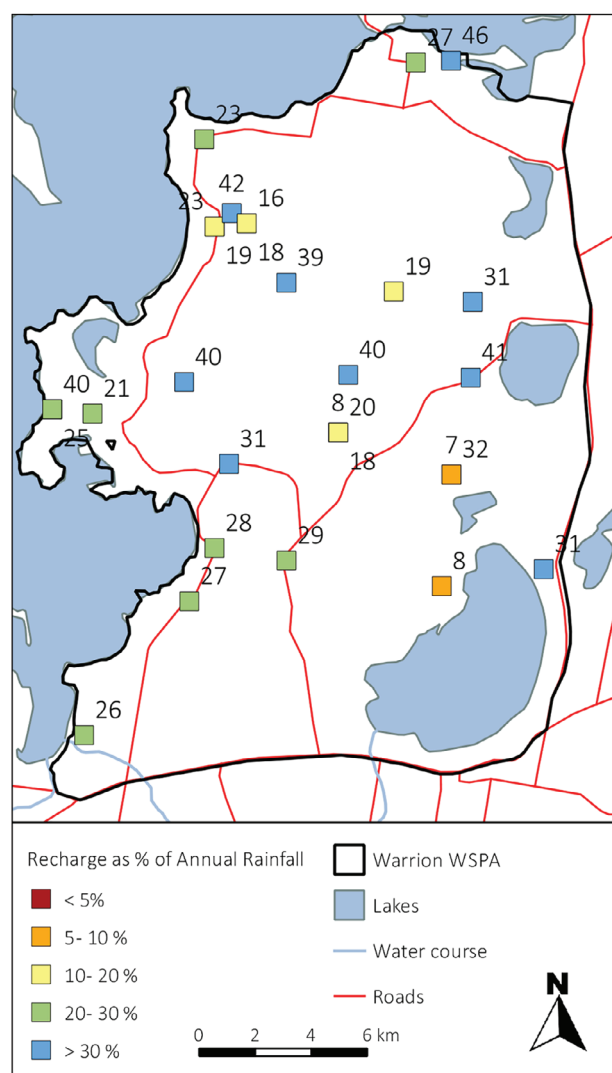
Figure 9. Mean recharge as percentage of annual precipitation.

Nolan (2007) in a review of the groundwater management for the study area reported a specific yield range of 0.01 to 0.03 for basalt aquifers (Sinha and Sharma 1988), porosity estimates of permeable basalt of 0.02 to 0.05 (Bouwer 1978) and a specific yield range of Victorian

basalts of 0.02-0.3 (Leonard 1992). Outside the study area, pumping test estimates of S for the Quaternary basalts and scoria are 0.0012 and 0.0045, respectively.

A summary of the different estimates of S is shown in Table 3. The *HydroSight* modeling produced a storativity of 2.5×10^{-4} to 0.32, though it should be noted that 0.32 is an outlier value and if this single result is excluded then the upper limit is an order of magnitude lower with results ranging from 2.5×10^{-4} to 0.034. The mean storativity from the 23 modeled bores in Warrion is 0.0043. Although the storativity range from the literature is very broad, this mean result is consistent with the Nolan (2000) pumping test results (0.0012-0.016) and results from Newer Volcanics/scoria aquifers in other Victorian pumping tests (0.0012-0.0045).

Recharge

Estimated gross recharge ranged from 7% to 46% of annual precipitation (580 mm/year) with a mean of 27%. This range excludes bore 142703 which was poorly calibrated. The spatial distribution of the estimated recharge across the study area is shown in Figure 9. Importantly, the gross recharge is physically modeled as the free drainage from the deep storage (Figure 2) derived from the numerical integration of the instantaneous free drainage at the start and end of each day; that is, to convert the instantaneous flux to a daily water depth. By assuming minimal deep drainage is lost to interflow or deep unsaturated transpiration, the free drainage can be considered as an upper estimate of the gross recharge; that is net recharge plus phreatic evapotranspiration.

Table 4 shows the recharge estimates from the *HydroSight* modeling and a range of independent estimates of net recharge. The *HydroSight* recharge estimates range from 41 to 267 mm/year with an average of 155 mm/year. The upper end of the *HydroSight* recharge range (267 mm/year) exceeds the upper range of the independent recharge estimates (175 mm/year) by around 50%. However, like *HydroSight* S estimates, the independent estimates all assume much higher recharge rates in the Warrion Hill/Red Rock area where the prevalent aquifer is fractured scoria. Unlike the *HydroSight* S estimates, which broadly reflected this geological distribution, the

highest *HydroSight* recharge estimates largely occur to the north of the Warrion Hill/Red Rock area on the basalt plains.

Discussion

This paper presented an extension to *HydroSight* that allows the time-series modeling of groundwater hydrographs when metered groundwater usage does not cover the history of usage and metering is infrequent. Such a lack of extraction data is characteristic of many groundwater management areas, and resolving this issue allows the point estimation of aquifer transmissivity, storativity, and recharge. We presented an evaluation of the consistency of the results and error associated with the approach. The approach was then trialed using observation bore data from the Warrion WSPA (Victoria, Australia) to generate point estimates of transmissivity, storativity, and recharge that were evaluated against independent estimates.

The *HydroSight* extension to accommodate incomplete usage data involved downscaling infrequent monthly and annual metered usage volumes to a daily time step using a new genetic evolutionary global optimization scheme (Figure S1). Model parameters and time series of daily pumping volumes were jointly optimized to minimize model error with the observed groundwater hydrograph while honoring metered usage volumes; however, the joint optimization is computationally demanding and can require ~6h per model. Additionally, the *HydroSight* method for estimating gross recharge was refined by constraining the long-term free drainage to a plausible range using the Budyko Curve, a function which defines the expected long-term actual evapotranspiration in a catchment based on its aridity.

In light of the challenges in jointly estimating model parameters and usage time series, four trials were undertaken, each involving repeated calibration of models: (1) estimation of the variability in transmissivity, storativity, and recharge values at four bores when each model was calibrated 50 times; (2) estimation of the reliability of the groundwater usage estimates; (3) identification of calibration settings that producible the most reproducible solutions (see SI); and (4) estimation of the radius of influence for the selection of pumping bores for each observation bore model (see SI).

The trial into the variability of estimates produced generally reproducible values of transmissivity, storativity, and mean recharge that varied by a factor of 0.22-0.32, 0.13-0.2, and 0.03-0.48, respectively, when boundary effects were low. Some such variability is an outcome of the probabilistic nature of global optimization (Arsenault et al. 2014) but the joint estimation of the parameters and time series of usage does increase the variability and if all of the models for the study area were recalibrated some variability would likely occur. Evaluation of the usage estimates found that the volumetric error was generally within 10% of weekly and biannual metered data. Reasonable estimation of daily usage was found to require weekly metering, while reasonable monthly usage generally required quarterly to biannual metering. Extending *HydroSight* to jointly model multiple observation bores is expected to improve the reliability of the usage estimation, while also eliminating the issue that usage estimates are likely to be inconsistent when observation bores are modeled individual and eliminating the requirement for the user to estimate the radius of influence. On the latter, for this study, a radius of 2 km was found to produce the most parsimonious models (Figure S3). However, the radius is likely to be site specific and for other studies, we advise undertaking trials using a range of radii and where uncertainty exists, or trials are not possible, choosing a radius on the larger side.

The study area was a groundwater management region in the western districts of Victoria, Australia. Groundwater extraction is predominantly from the newer volcanics, which forms an unconfined, fractured rock aquifer up to 60 m in thickness. *HydroSight* models were built and calibrated for 30 observation bores using daily precipitation and Morton's potential evapotranspiration estimates and groundwater extraction data for 136 licensed production bores.

Three model types were constructed and calibrated for each observation bore: (1) meteorological-only forcing using the one-layer soil model from Peterson and Western (2014), (2) meteorological-only forcing using the two-layer soil model with the constraint on soil evapotranspiration, and (3) the prior two-layer model with downscaled metered usage. Comparing the former two we found that the two-layer model was consistently the most, or nearly the most, parsimonious model (Table 1). Comparing the latter two models, we identified those observation bores likely to be influenced by pumping and then derived estimates of aquifer hydraulic properties (n=23). The estimated transmissivity ranged from 0.1 to 2424 m^2/d (harmonic mean of 84 m^2/d) and compared favorably with the average values from independent estimates (58-287 m^2/d) sourced from pumping test analysis, other analytical methods and literature. The upper range of the *HydroSight* estimates (2424 m^2/d) was ~150% higher than the maximum value from pumping test analysis (935 m^2/d), though the highest *HydroSight* transmissivity estimates were from a nested bore site and are not considered representative. Storativity across the 23 *HydroSight* models ranged from 2.5×10^{-4} to 0.32 (mean of 4.3×10^{-3}), noting that the 0.32 was an outlier value. Relative to transmissivity, there were a limited number of storativity estimates available from independent sources. However, the mean storativity (0.0043) is consistent with the more reliable pumping test results (0.0012-0.016) and the estimated range was commensurate with adopted specific yield values for the Newer Volcanics from local studies (0.01-0.3). Estimated gross recharge rates were derived for all 30 observation bores and ranged from 7% to 46 % of rainfall. For comparison with conventional recharge estimates, these were converted to an annual recharge range of 41-267 mm/year (mean of 155 mm/year). Although not implausibly high, the

T.J. Peterson and S. Fulton Groundwater 57, no. 6: 860–876

Table 4
Warrion *HydroSight* Estimates of Recharge and Conventional Recharge Estimates

Source	Recharge (mm/year)			Comment
	Min.	Max.	Arith. Mean	
HydroSight modeling	41	267	155	Includes all observation bores ($n = 30$)
Rainfall Recharge	58	174	—	Using Thompson (1971), management report assumed 30% of rainfall around Warrion Hill and 10% elsewhere
Regional Groundwater Models	12	60	—	Regional groundwater models calibrated using 55-60 mm/year at Warrion Hill and 12-37 mm/year elsewhere (Hyder 2010)
Environmental tracers	0.2	10	—	Chloride mass balance and carbon-14 data suggests 10 mm/year around Warrion Hill and 0.2-1 mm/year elsewhere (Hyder 2010)

HydroSight estimates of gross recharge are generally at the upper end or exceed the range from conventional estimates of net recharge (0.1-174 mm/year).

The most anomalous T and S results, and the poorest *HydroSight* model performance (as indicated by the CoE calibration) occurred in bores located within 1 km of Lake Corangamite. These results were not unexpected as previous research (Shapoori et al. 2015c) has shown that unaccounted boundary effects from surface water bodies can produce uncharacteristically low S and high T estimates. The results reinforce the importance of considering boundary affects using image-well analysis, functionality currently available within *HydroSight*. However, there are practical limitations to the application of image well analysis when the boundary is nonlinear and when there are many production bores; specifically, each production bore would require one image well equidistant from the boundary but on the other side and when the boundary is nonlinear positioning is nontrivial and may require multiple image wells per production bore. Where it is not practical to consider image well analysis for bores located in proximity to surface water bodies, lower reliability should be associated with the T and S estimates. Future research aims to implement the more flexible image-well method by Kuo et al. (1994) within *HydroSight*.

Analysis of estimated T and S values from a nested site with three bores (142714, 142717, and 142720) revealed T increased with bore depth. This trend was found to be inconsistent with the prevailing conceptualization of the hydrogeology where groundwater extraction largely occurs from a very shallow, basalt aquifer. The deepest bore was constructed in the Tertiary Sands, which although hydraulically connected shows a more muted response to pumping. As *HydroSight* assumes the observation bore is constructed within the pumped aquifer, the model accounts for the subdued water-level response by increasing the estimated T. However, a more plausible explanation that considers the site hydrogeology is that the deeper bore is not fully connected to the pumping aquifer in the upper Newer Volcanics. Findings at this site reinforce the need to interpret the *HydroSight* T and S estimates within the context of existing hydrogeological knowledge. Estimates of hydraulic properties are likely to be more reasonable when the observations bores are unambiguously constructed in the same aquifer as the production bores and even more so when the screened intervals between the production bores and observation bores are comparable.

Ideally, the appropriateness of the selected observation and production bores would be apparent in uncertainty estimates of T, S, and recharge. Such estimates would also aid the evaluation against the pumping test results. However, deriving such uncertainty estimates would require uncertainty estimates of the usage. Being time series, this would require ensembles of plausible time series for each production bore. Unfortunately, our approach does not allow derivation of such ensembles. *HydroSight* does however include a Markov Chain Monte Carlo (MCMC) calibration scheme that allows derivation of parameter uncertainties. If the MCMC scheme was modified to derive and store plausible ensembles of usage, then uncertainty estimates of T, S, and recharge would be possible.

A final broader comment is required on research into the estimation of aquifer hydraulic properties. *HydroSight* and many pumping test analysis methods use the Theis analytical drawdown equation, or a variant, to estimate T and S. These methods provide an effective T for the region from the observation bore to the pump bore(s) used in its estimation. Consequently, the heterogeneity within this scale is ignored. Zech and Attinger (2016) recently showed that for heterogeneous aquifers, steady state estimates of T (using the Theim equation) are biased. Zech and Attinger (2016) and Zech et al. (2016) presented pumping drawdown equations for the unbiased estimation of T and drawdown within heterogeneous aquifers and which included parameters for the spatial variance and spatial correlation in T. Synthetic applications show that unbiased local and aquifer average T values can be derived. However, no method has been published for the joint estimation of local and aquifer average S.

Additionally, the methods assume a constant pumping rate. When the research progresses, it is foreseeable that such an approach could be incorporated into *HydroSight* to enable local and aquifer average estimates of T and S. However, in the meantime it is important to acknowledge that the point estimates of T from *HydroSight* are likely to underestimate the aquifer heterogeneity, and for very heterogeneous aquifers they may be biased.

Conclusions

This paper presented an approach for the joint estimation of aquifer hydraulic properties and gross recharge from incomplete metered groundwater usage. Following testing of the approach and the application to the Warrion study area, the following conclusions can be made: (1) the calibration scheme produced generally reproducible estimates of transmissivity, storativity, and mean recharge that varied by a factor of 0.22-0.32, 0.13-0.2, and 0.03-0.48, respectively, when boundary effects are low; (2) downscaling of monthly, quarterly, and biannual metered usage was reasonable (volumetric error was <10%) but reasonable estimation of daily usage required weekly metering while reasonable monthly usage generally required quarterly to biannual metering; (3) the estimated aquifer transmissivity and storativity were reasonable and were consistent with independent estimates, while mean gross recharge may be slightly overestimated; and (4) reasonable estimation of transmissivity, storativity, and recharge requires that observation bores are screened within the pumped aquifer, boundary effects are minimal and that those models simulating pumping drawdown are more parsimonious than models simulating only meteorological dynamics.

Acknowledgments

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article. Supporting Information is generally *not* peer reviewed.

Appendix S1. Supporting Information.

References

Arsenault, R., A. Poulin, P. Côté, and F. Brissette. 2014. A comparison of stochastic optimization algorithms in hydrological model calibration. *Journal of Hydrologic Engineering* 19: 1374–1384. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000938

Bakker, M., K. Maas, and J.R. Von Asmuth. 2008. Calibration of transient groundwater models using time series analysis and moment matching. *Water Resources Research* 44: W04420. https://doi.org/10.1029/2007WR006239

Bouwer, H. 1978. *Groundwater Hydrology*, 480. New York: Mc-Graw Hill.

Burnham, K.P., and D.R. Anderson. 2002. *Model Selection and Inference. A Practical Information–Theoretic Approach. R.*, 2nd ed., xix+353. Heidelberg, Germany: Springer-Verlag ISBN 0-387-98504-2.

Chu, W., X. Gao, and S. Sorooshian. 2011. A new evolutionary search strategy for global optimization of high-dimensional problems. *Information Sciences* 181: 4909–4927. https://doi.org/10.1016/j.ins.2011.06.024

Coram, J.E. 1996. Groundwater-surface water interactions around shallow lakes of the western districts Victoria. Master's Thesis, University of Melbourne, April, 1996.

Doble, R.C., and R.S. Crosbie. 2017. Review: Current and emerging methods for catchment-scale modelling of recharge and evapotranspiration from shallow groundwater. *Hydrogeology Journal* 25, no. 1: 3–23. https://doi.org/10.1007/s10040-016-1470-3

Donohue, R.J., M.L. Roderick, and T.R. McVicar. 2012. Roots, storms and soil pores: Incorporating key ecohydrological processes into Budyko's hydrological model. *Journal of Hydrology* 436-437: 35–50. https://doi.org/10.1016/j.jhydrol.2012.02.033

Geological Survey of Victoria. 1973. *Colac Geological Mapsheet (1:250 000)*. Victoria, Australia: Mines Department.

Gill B., Webb J., Stott K., Cheng X., Wilkinson R. and Cossens, B. (2017), Economic, social and resource management factors influencing groundwater trade: Evidence from Victoria, Australia *Journal of Hydrology*, 2017, 550, 253 – 267, https://doi.org/10.1016/j.jhydrol.2017.04.055

Greve, P., L. Gudmundsson, B. Orlowsky, and S.I. Seneviratne. 2015. Introducing a probabilistic Budyko framework. *Geophysical Research Letters* 42: 2261–2269 https://doi.org/10.1002/2015GL063449

Ferris, J.G., and D.B. Knowles. 1963. The slug-injection test for estimating the coefficient of transmissibility of an aquifer. In *Methods of determining permeability, transmissibility, and drawdown*, ed. R. Bentall US Geological Survey Water Supply Paper, 299–304. U.S. Geological Survey: Washington, D.C., USA

Hurr, R.T. 1966. A new approach for estimating transmissibility from specific capacity. *Water Resources Research* 2, no. 4: 657–664. https://doi.org/10.1029/WR002i004p00657

Hyder. 2010. Review of technical work to inform the Warrion WSPA Groundwater Management Plan. Discussion paper prepared for Southern Rural Water. February 2010. Hyder Consulting: Melbourne.

Keir, G., N. Bulovic, and N. McIntyre. 2019. Stochastic modeling of groundwater extractions over a data-sparse region of Australia. *Groundwater* 57: 97–109. https://doi.org/10.1111/gwat.12658

Kruseman, G.P., and N.A. de Ridder. 1994. *Analysis and Evaluation of Pumping Test Data*, 2nd ed., Bulletin 11. Wageningen, The Netherlands: International Institute for Land Reclamation and Improvement.

Kuo, M., W.L. Wang, D.S. Lin, C.C. Lin, and C.J. Chiang. 1994. An image-well method for predicting drawdown distribution in aquifers with irregularly shaped boundaries. *Groundwater* 32: 794–804. https://doi.org/10.1111/j.1745-6584.1994.tb00921.x

Leonard, J.G. 1992. *Port Phillip Region Groundwater Resources—Future Use and Management*. Victoria, Australia: Department of Water Resources.

Liu, H.J., N.S. Hsu, and W.W.G. Yeh. 2015. Independent component analysis for characterization and quantification of

regional groundwater pumping. *Journal of Hydrology* 527: 505–516. https://doi.org/10.1016/j.jhydrol.2015.05.013

Obergfell, C., M. Bakker, and K. Maas. 2019. Estimation of average diffuse aquifer recharge using time series modeling of groundwater heads. *Water Resources Research* 55: 2194–2210. https://doi.org/10.1029/2018WR024235

Martínez-Santos, P., and P. Martínez-Alfaro. 2010. Estimating groundwater withdrawals in areas of intensive agricultural pumping in central Spain. *Agricultural Water Management* 98: 172–181. https://doi.org/10.1016/j.agwat.2010.08.011

Martínez-Santos, P., D. Pedretti, P.E. Martínez-Alfaro, M. Conde, and M. Casado. 2010. Modelling the effects of groundwater-based urban supply in low-permeability aquifers: Application to the Madrid Aquifer, Spain. *Water Resources Management* 24: 4613–4638. https://doi.org/10.1007/s11269-010-9682-0

Nolan-ITU. 2007. Review of permissible annual volume—Warrion GMA. Prepared for Southern Rural Water, Victoria, Australia.

Nolan-ITU. 2002. Groundwater technical report—Warrion GMA. Prepared for Southern Rural Water. January 2001, Victoria, Australia.

Nolan-ITU. 2000. Warrion pumping test—Kawana Dairies Pty Ltd. Letter report prepared for Southern Rural Water, Victoria, Australia.

Nolan-ITU. 1999. Warrion GMA assessment of groundwater levels and related issues. Prepared for Southern Rural Water, Victoria, Australia.

Peterson, T.J., C. Wasko, M. Saft, and M. Peel. In review. AWAPer: An R package for area weighted catchment daily meteorological data anywhere within Australia. *Hydrological Processes*

Peterson, T.J., A.W. Western, and X. Cheng. 2018. The good, the bad and the outliers: Automated detection of errors and outliers from groundwater hydrographs. *Hydrogeology Journal* 26: 371–380. https://doi.org/10.1007/s10040-017-1660-7

Peterson, T.J., and A.W. Western. 2014. Nonlinear time-series modeling of unconfined groundwater head. *Water Resources Research* 50: 8330–8355. https://doi.org/10.1002/2013WR014800

Shao, J., Y. Cui, Q. Hao, Z. Han, and T. Cheng. 2014. Study on the estimation of groundwater withdrawals based on groundwater flow modeling and its application in the North China Plain. *Journal of Earth Science* 25: 1033–1042. https://doi.org/10.1007/s12583-014-0493-8

Shapoori, V., T.J. Peterson, A.W. Western, and J.F. Costelloe. 2015a. Decomposing groundwater head variations into meteorological and pumping components: A synthetic study. *Hydrogeology Journal* 23, no. 7: 1431–1448. https://doi.org/10.1007/s10040-015-1269-7

Shapoori, V., T.J. Peterson, A.W. Western, and J.F. Costelloe. 2015b. Top-down groundwater hydrograph time series modeling for climate-pumping decomposition. *Hydrogeology Journal* 23: 819–836. https://doi.org/10.1007/s10040-014-1223-0

Shapoori, V., T..J. Peterson, A.W. Western, and J..F. Costelloe. 2015c. Estimating aquifer properties using groundwater hydrograph modelling. *Hydrological Processes* 29: 5424–5437. https://doi.org/10.1002/hyp.10583

Sinha, B.P.C., and S.K. Sharma. 1988. Natural groundwater recharge estimation methodologies. In *Estimation of Natural Groundwater Recharge*. NATO ASI Series C, Vol. 222, ed. I. Simmers. Dordrecht: Springer.

SKM. 2000. Warrion Water Supply Protection Area—Drilling and monitoring bore construction program. Prepared for Southern Rural Water, Victoria, Australia.

SKM. 1998. Permissible annual volume project. The Warrion GMA. January 1998. Monograph Series ISSN 1328-4495. Prepared for the Department of Natural Resources and Environment, Victoria, Australia.

Southern Rural Water. 2017. *Warrion Groundwater Management Plan Annual Report 2016-17*. Melbourne, Australia: Southern Rural Water.

Thompson, B. 1971. The geology and hydrogeology of the Corangamite region. Thesis for masters of Science. University of Melbourne, Australia.

Tweed, S., M. Leblanc, and I. Cartwright. (2008), Remote sensing and hydrochemistry of lakes-groundwater interaction. Abstract for the World Water Congress 2008.

von Asmuth, J.R., and M.F.P. Bierkens. 2005. Modeling irregularly spaced residual series as a continuous stochastic process. *Water Resources Research* 41: W12404. https://doi.org/10.1029/2004WR003726

von Asmuth, J.R., M.F.P. Bierkens, and K. Maas. 2002. Transfer function-noise modeling in continuous time using predefined impulse response functions. *Water Resources Research* 38: 1287. https://doi.org/10.1029/2001WR001136

Zech, A., and S. Attinger. 2016. Technical note: Analytical drawdown solution for steady-state pumping tests in two-dimensional isotropic heterogeneous aquifers. *Hydrology and Earth System Sciences* 20: 1655–1667. https://doi.org/10.5194/hess-20-1655-2016

Zech, A., S. Muller, J. Mai, F. Heße, and S. Attinger. 2016. Extending Theis' solution: Using transient pumping tests to estimate parameters of aquifer heterogeneity. *Water Resources Research* 52: 6156–6170. https://doi.org/10.1002/2015WR018509

Zhang, L., K. Hickel, W.R. Dawes, F.H.S. Chiew, A.W. Western, and P.R. Briggs. 2004. A rational function approach for estimating mean annual evapotranspiration. *Water Resources Research* 40: W02502. https://doi.org/10.1029/2003WR002710