Hydrogeol J (2018) 26:371–380
DOI 10.1007/s10040-017-1660-7

PAPER

The good, the bad and the outliers: automated detection of errors and outliers from groundwater hydrographs

Tim J. Peterson [1] · **Andrew W. Western** [1] · **Xiang Cheng** [2]

Received: 22 February 2017 / Accepted: 15 August 2017 / Published online: 26 September 2017

Abstract Suspicious groundwater-level observations are common and can arise for many reasons ranging from an unforeseen biophysical process to bore failure and data management errors. Unforeseen observations may provide valuable insights that challenge existing expectations and can be deemed outliers, while monitoring and data handling failures can be deemed errors, and, if ignored, may compromise trend analysis and groundwater model calibration. Ideally, outliers and errors should be identified but to date this has been a subjective process that is not reproducible and is inefficient. This paper presents an approach to objectively and efficiently identify multiple types of errors and outliers. The approach requires only the observed groundwater hydrograph, requires no particular consideration of the hydrogeology, the drivers (e.g. pumping) or the monitoring frequency, and is freely available in the *HydroSight* toolbox. Herein, the algorithms and time-series model are detailed and applied to four observation bores with varying dynamics. The detection of outliers was most reliable when the observation data were acquired quarterly or more frequently. Outlier detection where the groundwater-level variance is nonstationary or the absolute trend increases rapidly was more challenging, with the former likely to result in an under-estimation of the number of outliers and the latter an overestimation in the number of outliers.

Keywords Groundwater monitoring · Statistical modeling

✉ Tim J. Peterson
timjp@unimelb.edu.au

[1] Department of Infrastructure Engineering, The University of Melbourne, Victoria, Australia

[2] Department of Economic Development, Jobs, Transport and Resources, Victoria, Australia

Introduction

Unreliable groundwater level observations can arise from observation bore failures (e.g. bore collapse or bore flooded) and data management (e.g. recording in the field, entry into databases) and can be deemed errors. Such errors are often not recorded by quality codes, in which case the observed hydrograph contains spurious observations that may compromise trend analysis and model calibration, especially when using a least-squares objective function which will give undue weight to groundwater levels appearing as *spikes*. Outliers can also arise when an observation does not conform to our expectations of the aquifer and, while further examination may identify them as errors, they may also offer insights to the aquifer that should not be overlooked. Ideally, such errors and outliers should be identified and considered for omission from analysis. However, the lack of appropriate automated methods results in errors and outliers either being ignored or a reliance on manual subjective inspection of the data, a process which is not reproducible, inefficient when multiple parties analyze the same data and impossible for analysis of large (e.g. national) datasets. This paper presents an approach to objectively and efficiently identify multiple types of spurious observations, with the implausible deemed errors and the plausible but unusual deemed outliers. The approach requires only the observed groundwater hydrograph and requires no consideration of the local hydrogeology, meteorology or other drivers (e.g. pumping).

The identification of erroneous groundwater level observations has only received modest attention to date. Tremblay et al. (2015) identified outliers using a simple threshold maximum deviation from the mean water level, while Li et al. (2016) identified outliers as those points more than three standard deviations away from a moving-window smoothed hydrograph. More recently Berendrecht and van Geer (2016) presented a multi-bore groundwater time-series approach that was shown capable of graphically identifying single outliers when neighboring bores where frequently monitored. Outside hydrogeology, time-series outlier detection methods have been investigated for many decades. For time-series outlier detection, statistical time-series modeling has been the predominant approach. Four types of outliers have been characterized, with approaches developed for each—following Chan (1995), these are additive outliers (AO), innovational outliers (IO), level shifts (LS) and temporary changes (TC).

An AO affects only the observation(s) in question, appearing as a spike in groundwater level for one or more isolated observations. IOs are a more challenging scenario. With IOs, serial correlation causes the outlier to affect all subsequent observations, and in this application could result from an obstruction to the bore that influences all future observations. Similarly, LS affects all subsequent observations but by a step-change in the system rather than from the serial correlation, and in this application might result from a shift in the measurement datum. Lastly, TC affects subsequent observations but, unlike IOs, the impact decays with time. Stepping back from this typology, outliers can be deemed as a subset of anomalies—that is, unusual observations that may be erroneous or may provide clues of new processes that warrant investigation, not removal (Chandola et al. 2009). The automated removal of hydrograph errors and outliers risks overlooking potentially important clues (e.g. a bypass recharge event appearing as a spike), and it is very difficult to distinguish between errors and true but unusual system responses. In this paper, the outliers identified cannot be objectively separated into these two groups and, in the end, the outliers may be valuable clues omitted from further consideration.

Despite these caveats, AOs are likely to be the most common type of outliers in groundwater hydrographs and are the focus herein. However, temporary LSs can also appear, in the authors' experience, as periods of constant head offset arising from data management failures or the water level declining below the screen. The paper therefore also presents a series of hydrogeologically relevant heuristics to identify LS errors, and other spurious observations, after which AOs are identified using a time-series modeling approach. It is however important to remain mindful that more challenging types of anomalies, e.g. IO and TC, are unlikely to be detected by the presented approach and their prevalence may increase with increased adoption of automated sensors.

The basis of the AO detection is a comparison between individual observed data points and a time-series model fitted to the data—for example, Battaglia and Orfei (2005) used a nonlinear time-series model to identify the timing of outliers and their amplitude from regularly sampled nonlinear data. They found that the detection of outliers required iterative application of the model to overcome the challenge of outliers remaining within the dataset affecting the parameters of the time-series model used to identify outliers. In this paper, a similar iterative approach was found to be important in outlier detection. A further challenge, and one not known to be addressed in the outlier detection literature, is that groundwater hydrographs very rarely have a regular monitoring time step, are likely to be first-order and second-order nonstationary (i.e. mean and variance change with time) and, with aquifers being an open system, the structure of the time-series model and drivers are inherently unknown. Furthermore, the drivers causing the observed groundwater level time-series may not be known, nor observed, and they may vary over time. With these challenges, it is not surprising that an automated outlier detection approach has not yet been developed for groundwater hydrographs. Advances in groundwater time-series modeling present a means to account for irregular observation time steps and serially correlated residuals (von Asmuth and Bierkens 2005) within a numerically robust approach (Peterson and Western 2014). In the following section, an algorithm is presented for detecting common errors and then the aforementioned research is drawn upon to present an algorithm for detecting outliers. Both have been developed without limitations on the type of aquifer and monitoring regime. Four applications are presented, which are then discussed.

Methods

This section presents an overview of the approach for automatic identification and removal of erroneous and outlier observations. The main algorithm is in section "Error and outlier detection algorithm". It first uses heuristics to identify and remove physically implausible observation (e.g. observations below the screen or before the bore construction date), which herein are considered to be errors. Next, the main algorithm calls the sub-algorithm to identify plausible but unusual observations, which are considered to be outliers in this paper. The approach is available as a graphical module within the HydroSight toolbox (Peterson et al. 2017), as is the source code (see files *doDataQualityAnalysis.m*, *outlierDetection.m* and *ExpSmooth.m* for details of the error detection, outlier detection and exponential smoothing respectively).

Error and outlier detection algorithm

The algorithm for the automated identification and removal of erroneous and outlier observations is detailed as algorithm 1, in Fig. 1. The required inputs are the groundwater hydrograph to be analyzed (observation date-time, water level elevation and ideally with known errors already removed, e.g., quality codes denoting, say, a collapsed bore) and eight user-set parameters controlling the error and outlier analysis; four of these are based on the physical attributes of the bore, and guidance is given for the remaining four in section "Algorithm application".

Fig. 1 Outline of the groundwater hydrograph error and outlier detection (algorithm 1). Variables are italic, vectors are bold italic, the scope for each "For", "If" and "Else" term is defined by the indentation, and comments are denoted by /*... */

Line number	**Data**: Groundwater hydrograph head, $\boldsymbol{h}$, and time, $\boldsymbol{t}$, vector. **Parameters:** bore construction date, $t_{\min}$ [T]; bore termination date, $t_{\max}$ [T]; screening lower elevation, $h_{\min}$ [L]; top of casing elevation, $h_{\max}$ [L]; maximum plausible absolute rate of change, $r_{\max}$ [L T^{-1}]; minimum constant head duration, t_{const} [T]; minimum number of obs. in constant head period, n_{const} [-]; number of noise standard deviations denoting an outlier, η [-]
1	Sort hydrograph by date/time in ascending order.
2	*/*Identify implausible observation dates*/*
3	**For** (i = first to most recent observation)
4	**If** ($t_i < t_{\min}$ Or $t_i >$ min(Today's date, $t_{\max}$))
5	Remove observation i.
6	*/* Identify duplicate observations for a time point. */*
7	**For** (i = first to second most recent observation)
8	**If** ($\lvert t_{i+1} - t_i \rvert <$ sqrt(floating point *machine precision*))
9	Remove observation i.
10	*/* Identify implausibly high or low head values. */*
11	**For** (i=first to most resent observation)
12	**If** ($h_i < h_{\min}$ Or $h_i > h_{\max}$)
13	Remove observation i.
14	*/* Identify implausibly rapid change in the head.*/*
15	**For** (i = first to second most recent observation)
16	r_i = abs(($h_{i+1} - h_i$)/(t_{i+1}-t_i))
17	**If** ($r_i > r_{\max}$)
18	$isError_i$ = True
19	*/* Filter hydrograph to non-erroneous observations.*/*
20	$\boldsymbol{h} = \boldsymbol{h}$(Not *isError*)
21	clear *isError*
22	*/* Identify periods of constant head. First the periods of constant head are identified, then their duration is assessed.*/*
23	**For** (i=first to second most recent observation)
24	$IsFlat_i$ = False
25	**If** ($h_{i+1} - h_i$ == 0)
26	$IsFlat_i$ = True
27	**For** (i=second to most recent observation)
28	**If** ($IsFlat_i$ And Not $IsFlat_{i-1}$)
29	$t_{\text{startFlat}} = t_i$;
30	$t_{\text{endFlat}} = 0$;
31	$j = 1$
32	**Else If** ($t_{\text{startFlat}} > 0$ And Not $IsFlat_{i+1}$)
33	$t_{\text{endFlat}} = t_i$;
34	**If** ($t_{\text{endFlat}} - t_{\text{startFlat}} > t_{\text{const}}$ And $j \geq n_{\text{const}}$)
35	$isError_i$ = True
36	$t_{\text{startFlat}} = 0$
37	$j = 0$
38	**Else If** ($t_{\text{startFlat}} > 0$)
39	$j = j + 1$
40	*/* Filter hydrograph to non-erroneous observations.*/*
41	$\boldsymbol{h} = \boldsymbol{h}$(Not *isError*)
42	clear *isError*
43	*/* Undertake outlier analysis on the remaining data*/*
44	**Call Function:** *IsOutlier* = outlierDetection($\boldsymbol{h}$,$\boldsymbol{t}$, η)
45	*/* Filter out outlier observations.*/*
46	$\boldsymbol{h} = \boldsymbol{h}$(Not *IsOutlier*)
47	clear *IsOutlier*

The algorithm sequentially applies a series of heuristics to identify errors, with the order of each chosen to maximize its effectiveness. The first set of heuristics analyzes all observations for plausible observation dates (i.e. after the construction date and prior to the minimum of the bore termination data and the current date), duplicate observations for a single time-point, and plausible heads (i.e. above the elevation of the bottom of the screen and below the top-of-casing elevation). Next, the sequence of remaining observations is analyzed for errors—see lines 14–42 of algorithm 1 (Fig. 1). This involves searching for periods of implausible absolute change in head (see algorithm 1, lines 14–18), followed by periods of constant head exceeding a user defined duration (see algorithm 1, lines 22–42). Lastly, the remaining observations are assessed for outliers (see algorithm 1, lines 43–47 and section "Outlier detection algorithm").

Clearly not all of the preceding heuristics are applicable to all observation bores—for example, the check for the water level being below the top-of-casing would be inappropriate for artesian bores. Likewise, checking for duplicate observations may be inappropriate when automatic pressure transducers are checked against manual observations. The selected heuristics should therefore be chosen to reflect the physical conditions of the study area. HydroSight allows selection of each individual heuristic.

Outlier detection algorithm

The algorithm for outlier detection is given as algorithm 2, in Fig. 2. Central to the algorithm is the construction of a double exponential smoothing time-series model and its fit to the observed hydrograph using the SP-UCI (Chu et al. 2011) global calibration scheme (section "Outlier detection time-series model"). After an outlier is identified and removed, the time-series model is recalibrated and, while computationally demanding, it was found necessary to identify sequences of outliers and is consistent with other outlier detection approaches (Battaglia and Orfei 2005).

Conceptually, the key steps of the outlier detection algorithm, starting from observation $i = 1$, are as follows:

1. Calibrate the time series model using the remaining data from the error detection process.
2. Removing observation i from the hydrograph and run the time-series model using the remaining data.
3. Calculate the model residuals and, to minimize the impacts of remaining outliers other than observation i, remove the minimum and maximum residual values from the vector or residuals (note: all minimum and maximum residual values are removed, and at least two values are always removed).
4. Estimate the hydrograph noise (σ_t) at observation i using Eq. (15).
5. Predict observation i water level using the calibrated model parameters and Eq. (3).
6. Calculate the error in the prediction. If the error is greater than $\eta \times \sigma_t$ (η is a user set number of noise standard deviations, herein set to 4), then identify the observation as an outlier, remove it and go to step 7. Otherwise, increment i and repeat steps 2–6 for all observations.
7. Set $i = 1$ and repeat steps 1–6 until no outlier is detected.

In assessing if an observation is an outlier (the aforementioned step 6), the residuals were assumed independent and identically distributed (iid). With η set to four and assuming normality of the residuals, the probability of an observation being four noise standard deviations away from the time-series model prediction would be < 6.3e-05. Assuming the time-series model is adequate and the residuals are iid and normal, there is a low probability of incorrectly removing nonoutliers. Given the simplicity of the model (described in the next section), it may produce residuals that are serially correlated, in which case the aforementioned probabilities should be considered as a lower limit.

Outlier detection time-series model

Smoothing groundwater hydrographs presents a number of challenges. Specifically, groundwater hydrographs are often irregularly sampled and are influenced by temporally varying

Fig. 2 Outline of the outlier detection algorithm (algorithm 2). Variables are italic, vectors are bold italic, the scope for each "For", "If" and "Else" term is defined by the indentation, and comments are denoted by /*... */

```
Data: Groundwater hydrograph head, h, and time, t, vector.
Parameters: number of noise standard deviations denoting an outlier, η[-]
While (outlier.fraction > outlier.expected)
    Build double exponential smoothing time-series model.
    Calibrate time-series model to h and t.
    /* Loop through each observation, re-running the time-series smoothing model without
    the current observation and then assessing if the residual from the smoothed estimate
    is >  η * σ.*/
    Initialize IsNewOutlier to False.
    For (i = first to last non-outlier observation)
        Remove h_i from h and set to h.trim
        Run double exponential smoothing time-series model using h.trim
        Calculate residual, i.e. h.trim minus the smoothed hydrograph.
        Remove the minimum and maximum residual.
        Calculate the standard deviation of the noise from the remaining residuals, σ.
        Estimate the smoothed head at t_i.
        Estimate the residual at t_i, Δ.
        /* Exit For-loop if the current point is an outlier */
        If (abs(Δ) > η * σ)
            IsNewOutlier_i =True
            Break
    /* Calculate the fraction of newly identified outliers and the expected fraction */
    outlier.fraction = sum(IsNewOutlier) / No. non-outlier observations
    outlier.expected = normal PDF at η * σ  from a distribution with a mean of 0 and
    standard deviation of σ.
/* While loop ended. Filter out outliers.*/
```

external drivers such as climate, pumping or land-use change, that produce water level fluctuations on timescales ranging from daily to decadal or longer. Combined, these challenges result in an irregularly sampled first and second order nonstationary time-series not suitable for analysis with standard exponential smoothing techniques. An alternative is damped double exponential smoothing for irregular data (Cipra 2006). The approach smooths both a damped trend and the component not explained by the trend. Combined with Wright's method for irregularly spaced data, a smoothing algorithm can be derived that is applicable to groundwater hydrographs. From Cipra (2006), the smoothed estimate at time t_n, S_{t_n}, is estimated from the weighted sum of the current observation, y_{t_n} (i.e., the observed groundwater level), the smoothed estimate from the prior time step, $S_{t_{n-1}}$, and the damped trend, $T_{t_{n-1}}$:

$$S_{t_n} = \alpha_{t_n} \times y_{t_n} + (1-\alpha_{t_n})\left(S_{t_{n-1}} + \sum_{i=1}^{t_n-t_{n-1}} \varphi^i T_{t_{n-1}}\right) \tag{1}$$

where α_{t_n} is a time-varying smoothing coefficient and the sum term dampens the prior trend $T_{t_{n-1}}$ using a damping coefficient ($0 < \varphi < 1$) to the power of the duration from t_{n-1} to t_n, causing the trend to decay to zero over time. Thus the trend at the current time point is:

$$T_{t_n} = \gamma_{t_n} \frac{\varphi^{(t_n-t_{n-1})}}{\sum_{i=1}^{t_n-t_{n-1}} \varphi^i} (S_{t_n} - S_{t_{n-1}}) + \left(1-\gamma_{t_n}\right)\varphi^{(t_n-t_{n-1})} T_{t_{n-1}} \tag{2}$$

where γ_{t_n} is a time-varying smoothing coefficient. The forecast, required at line 12 of algorithm 2 (Fig. 2), is calculated as:

$$\hat{y_{t_{n+k}}} = S_{t_n} + \sum_{i=1}^{k} \varphi^i T_{t_{n-1}}, k \geq 0 \tag{3}$$

where k is the number of time steps forecast. Additionally, the two time-varying smoothing coefficients, α_{t_n} and γ_{t_n}, are estimated as follows:

$$\begin{aligned} \alpha_{t_n} &= \frac{\alpha_{t_{n-1}}}{(1-\alpha)^{t_n-t_{n-1}} + \alpha_{t_{n-1}}} \\ \gamma_{t_n} &= \frac{\gamma_{t_{n-1}}}{(1-\gamma)^{t_n-t_{n-1}} + \gamma_{t_{n-1}}} \end{aligned} \tag{4}$$

where the initial values, α_{t_0} and γ_{t_0}, are approximated from the mean observation frequency (q) and input parameters α and γ:

$$\alpha_{t_0} = 1-(1-\alpha)^q \tag{5}$$

$$\gamma_{t_0} = 1-(1-\gamma)^q \tag{6}$$

$$q = \frac{1}{n} \sum_{i=1}^{n-1} (t_{n+1} - t_n) \tag{7}$$

Departing from Cipra (2006), the initial values for S_{t_0} and T_{t_0} were estimated using a smoothing spline (see Matlab *csaps* function) fit to the entire observation record, from which the initial spline head and slope are extracted. Additionally, inclusion of the damping coefficient ($0 < \varphi < 1$) was found to add little benefit at the cost of increased calibration demand. Therefore, it was removed by replacing the sum terms in Eqs. (1), (2) and (3) with the time step, $t_n - t_{n-1}$, and the term $\varphi^{(t_n-t_{n-1})}$ in Eq. (2) by 1. The final equations used in the approach were:

$$S_{t_n} = \alpha_{t_n} \times y_{t_n} + (1-\alpha_{t_n})(S_{t_{n-1}} + (t_n - t_{n-1}) T_{t_{n-1}}) \tag{8}$$

$$T_{t_n} = \frac{\gamma_{t_n}}{t_n - t_{n-1}} (S_{t_n} - S_{t_{n-1}}) + \left(1-\gamma_{t_n}\right) T_{t_{n-1}} \tag{9}$$

$$\hat{y_{t_{n+k}}} = S_{t_n} + (t_n - t_{n-1}) T_{t_{n-1}}, k \geq 0 \tag{10}$$

The application of the exponential smoothing model to a hydrograph requires two parameters, α and γ. Considering the natural variability in groundwater hydrographs and monitoring regimes, it is appropriate that the parameters be estimated for each hydrograph and that this be objective and automated. This was achieved using a variant of the von Asmuth and Bierkens (2005) objective function, which accounts for the serial correlation between the residuals by using the innovations (i.e. the change in the model residuals over irregular time steps) rather than the model residuals themselves, and then weights the innovations by the time step, which can be irregular. In the following, the relevant equations from the modifications to the von Asmuth and Bierkens (2005) by Peterson and Western (2014) are reproduced.

The residuals at a time point were calculated as:

$$\varepsilon_t = h_t - h_t^* - d \tag{11}$$

where ε_t [L] is the residual at time t; h_t [L] is the observed groundwater elevation at time t; h_t^* [L] is the smoothed groundwater level at time t; and d [L] is a drainage elevation constant. In modeling the groundwater level, the mean of the residuals was assumed to equal zero and this allowed estimation of the drainage elevation constant:

$$d = \overline{h} - \overline{h_t^*} \tag{12}$$

where $\overline{h}$ [L] is the arithmetic mean observed groundwater level over the calibration period; and $\overline{h_t^*}$ [L] is the arithmetic mean smoothed estimate, h_t^* [L], over the calibration period. Use of the mean levels does however assume that the observed hydrograph has a fairly regular monitoring frequency. If the hydrograph has long periods of, say, seasonal observations followed by a period of daily observations, then the arithmetic mean may be overly weighted toward the period of daily observations; care should therefore be taken in application of the approach to hydrographs having highly varying observation frequencies.

Next, the innovations (υ_t [L]) over an irregular time step were defined by assuming an exponential decay of the residuals:

$$\upsilon_t = \varepsilon_t - \varepsilon_{t-\Delta t} e^{-\beta \Delta t} \tag{13}$$

where $\varepsilon_{t-\Delta t}$ [L] is the residual at the previous groundwater observation time; Δt [T] is the time step to the previous groundwater observation; and β [1/T] is a parameter defining the decay rate of the residuals. The standard deviation of the model noise (σ_n [L]) is then estimated from the residuals:

$$\sigma_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} \left(\frac{\upsilon_i^2}{1 - \exp(-2\beta \Delta t_i)} \right)} \tag{14}$$

In this application, the noise needs to be time-dependent to account for the varying time separation between the observation being assessed as an outlier and the preceding observation. To estimate the time-dependent noise, Eq. A7 in von Asmuth and Bierkens (2005) was extended as follows using the arithmetic mean of the innovations:

$$\sigma_t = \sqrt{\frac{\frac{1}{n-1} \sum_{i=1}^{n-1} \upsilon_i^2}{1 - \exp(-2\beta \Delta t)}} \tag{15}$$

where in this instance Δt [T] is the time separation between the time-point (t) at which an estimate is required and the time at the proceeding observation. Using the aforementioned Eq. (10), von Asmuth and Bierkens (2005) derived an objective function used to optimize all model parameters, including the noise parameter β. To increase numerical stability, Peterson and Western (2014) modified it to the following:

$$\chi^2 = \sum_{i=1}^{n-1} \frac{\exp\left\{ \frac{1}{n-1} \sum_{j=1}^{n-1} \left[\ln\left(1 - \exp\left(-2\hat{\beta}\Delta t_i\right)\right) \right] \right\}}{1 - \exp\left(-2\hat{\beta}\Delta t_j\right)} \upsilon^2\left(t_j, \hat{\beta}\right) \tag{16}$$

where $\hat{\beta}$ is the anti-log transform of β and $\hat{\beta} = 10^{\beta}$. To improve the performance of the global calibration scheme, SP-UCI, all parameters should be encoded to have similar sensitivity to each other. To achieve this, all three calibrated parameters were $\log_{10}$ transformed.

Algorithm application

Four unconfined observation bores where selected to demonstrate the error and outlier detection algorithms (Table 1). The bores are located in Victoria, Australia, and were selected to illustrate the performance of the algorithms under a range of scenarios. Specifically, the bores were chosen with periods of constant head (bore 27174), approximately linear trends (bore 27174), varying trends (bore 99), possible sequential outliers (bores 27174 and WRK957283) and periods of high and low variance (bore 1101). In each application of the algorithms, the minimum constant head duration (t_{const}) was 90 days, the minimum number of observations in the constant head period (n_{const}) was three, the number of noise standard deviations denoting an outlier (η) was four and the maximum plausible absolute rate of change ($r_{\max}$) was set to 10 m/day; future applications could adopt a site specific value derived from the within-season water level variability. All other user-set parameters were taken from Table 1.

The outlier detection algorithm is clearly the most significant aspect of the approach. To further explore its behavior, the application to the four bores was repeated using (1) water level data resampled to lower frequencies and (2) a range of values for η. For the resampling, the water level data were resampled to 49 time steps having a minimum frequency of 28 days up to 364 days (7-day increments). Additionally, the identified outliers from the complete data were inserted into each resampled data set to allow comparison of the results. To analyze the results, the following measures were extracted for each bore and time step: coefficient of efficiency, providing a measure of the model fit to the overall time-series; each of the three time-series model parameters; the standard deviation of the noise (Eq. 14) and the exponential model variogram range of the innovation, providing a measure of the serial correlation; the number of outliers detected; and the number of the outliers detected from the complete data that were identified using each resampled data set.

To provide some indication of the computational demands of the error and outlier detection, the run-time for the aforementioned analysis was estimated at $\tau = 2.79m + 6.51$($r^2 = 0.793$) where τ is the run-time in seconds and m is the number of outliers detected (note, the analysis was undertaken using a 6-Core 3.6GHz i7-6850 PC). For the complete data, analysis of each of the four bores required a run-time mean and standard deviation of 33 s and 28 s respectively. Lastly, to investigate the sensitivity to η the analysis was repeated using the complete data and values of η from 2 to 6 at 0.25 increments. The results were then summarized by extracting the number of outliers detected and the number of those outliers detected from when $\eta = 4$.

Results

Figure 3 summarizes application of the error and outlier detection algorithm to the four selected bores. Figure 3a shows a hydrograph having an approximately linear trend and irregular sampling, with monthly samples pre-2007 and annual samples thereafter. Two periods of constant head were identified by algorithm 1 and eight outliers by algorithm 2. Most notably, a sequence of outliers with different head values was detected

Table 1 Summary of the four unconfined groundwater observation bores selected to demonstrate the error and outlier detection algorithms

Bore ID	Lat.	Long.	Construction year	Land surface elevation (m AHD)	Bore depth (m)	No. of observations	Mean time step (days)	SD of time step (days)
27,174	−34.13	141.07	1991	26.3	10	160	51	75
WRK957283	−36.50	146.45	1998	169.1	14.94	155	37	25
1101	−35.93	145.67	1975	114.4	9.5	191	34	15
99	−36.38	145.07	1971	99.3183	13.4	404	37	24

AHD Australian Height Datum, *SD* standard deviation

in 1993 and another sequence, but with similar head values, was detected in 2006. Figure 3b shows a hydrograph with approximately regular monthly monitoring but with more complex subannual and interannual trends. Four outliers were detected in 2010 that are nonsequential with one nonoutlier between each, despite appearing to be a sequence of outliers.

Figure 3c shows a hydrograph again having complex trends but also a temporal trend in the variance, with the within-year head variance increasing following the late 1980s; the cause of this is unknown but it may be due to the commencement of irrigation and/or groundwater pumping within the region. The fitted noise is effectively an average of the two periods, with the noise in the earlier period overestimated and that in the latter period underestimated. No outliers were identified in either period. This example illustrates a limitation of the outlier detection approach in that it assumes a stationary variance. Outliers within periods of lower variance would need to be quite extreme to be detected.

Lastly, Fig. 3d shows a hydrograph having an initial interannual period of rapid water level rise followed by a period of relatively constant water level. Nine outliers were identified, but inspection of the hydrograph suggests some of those identified in the 1980s and 1990s are questionable. To investigate, the insert to Fig. 3d shows an iteration of the algorithm 2 *while-loop* analyzing the fourth outlier. It shows the trimmed observed hydrograph with the point under consideration omitted and the fitted smoothed hydrograph. To assess the observation (denoted by a white circle), the time-series model is first used to forecast the point to assess (denoted by a black square). If the observed point is, in this example, greater than four times the standard

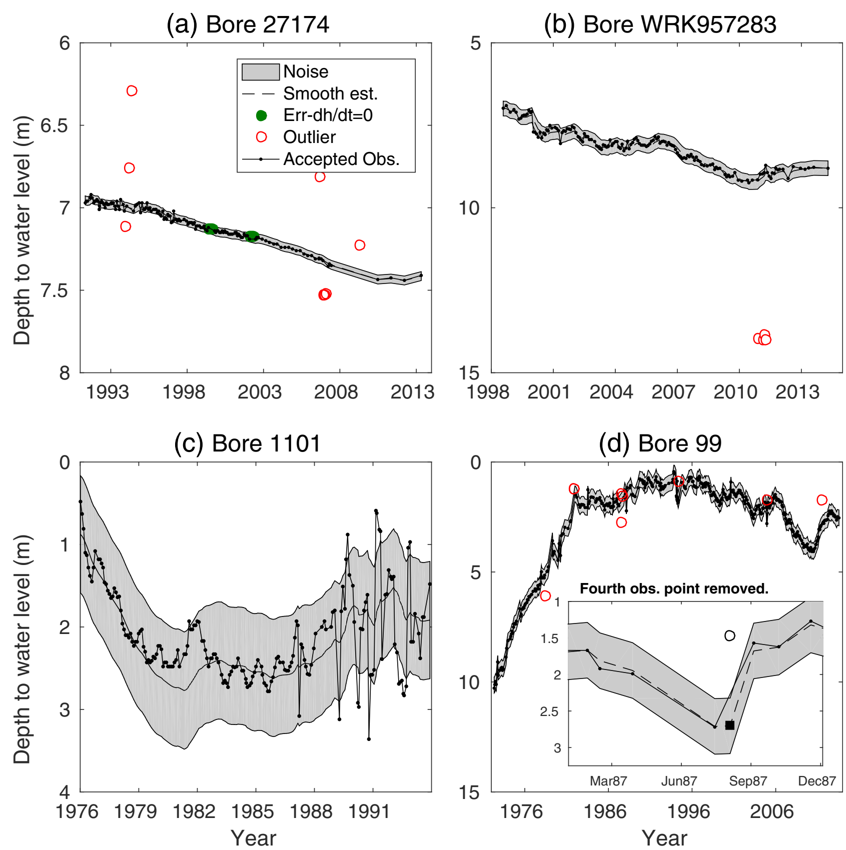

Fig. 3 Application of the error and outlier detection algorithms to four observation bores in Victoria, Australia. Each plot presents the smoothed head, the 5th–95th percentile envelope of noise (±1.68 standard deviations), the identified errors and outliers and the final hydrograph with errors and outliers removed. Note, only (**a**) contained constant head errors and no outliers were detected in (**c**). The insert in (**d**) details the identification of the fourth outlier and challenges that can arise when the head rapidly rises or falls

deviation of the noise from the forecast point then it is deemed an outlier. From Eq. (3), the forecast is derived from the sum of the weighted previous observation and the trend at the previous observation. The point to be assessed is not used in the trend estimation and, as shown by the insert, when there is a rapid water level change (relative to the recent interannual water level variance) the forecast may significantly under or overestimate the observed water level. In Fig. 3d, this resulted in the questionable identification of six observations as outliers (two in 1981, three in 1987, one in 2004).

To explore the performance of the outlier detection algorithm as the observation frequency changes, Fig. 4 shows results from resampled water level data with maximum time steps of 28 up to 364 days (7-day increments), with that at a zero time step showing the results from the complete data (see Table 1 for details of the complete data). Figure 4a shows that the coefficient of efficiency for the fit between the time-series model and the resampled water level data, with 1 denoting a perfect fit and < 0 denoting a fit worse than from using the mean water level. It shows that for 3 of the 4 bores the fit was very good when the sample time step is < 100 days. For bore 1101, the fit was around 0.5, and this is most likely due to the water level variance changing over time. Considering the purpose of the time-series model is to produce a smoothened head for estimation of the noise, and not an excellent fit to the hydrograph, the more relevant aspect from Fig. 4a is that the fit is stable at all bores up to a time step of 100–200 days. Figure 4b–d show that the corresponding model parameters were more variable but do show a similar change at about 100 days, and most notably so in Fig. 4b–c.

Figure 4e shows the estimated residual noise standard deviation after all identified errors and outliers are removed. Similar to the proceeding plots, it shows a notable increase between 100 and 200 days for bores WRK957283 and 99 and slightly so for 1101. At WRK957283, the noise increased by over an order of magnitude, and also at 1101 but only beyond 250 days observation time step. Figure 4f shows a similar increase in the correlation of the innovations beyond 70 days to a variogram range of above 200 days. Interestingly, below 70 days the variogram range was between 14 and 60 days for three bores.

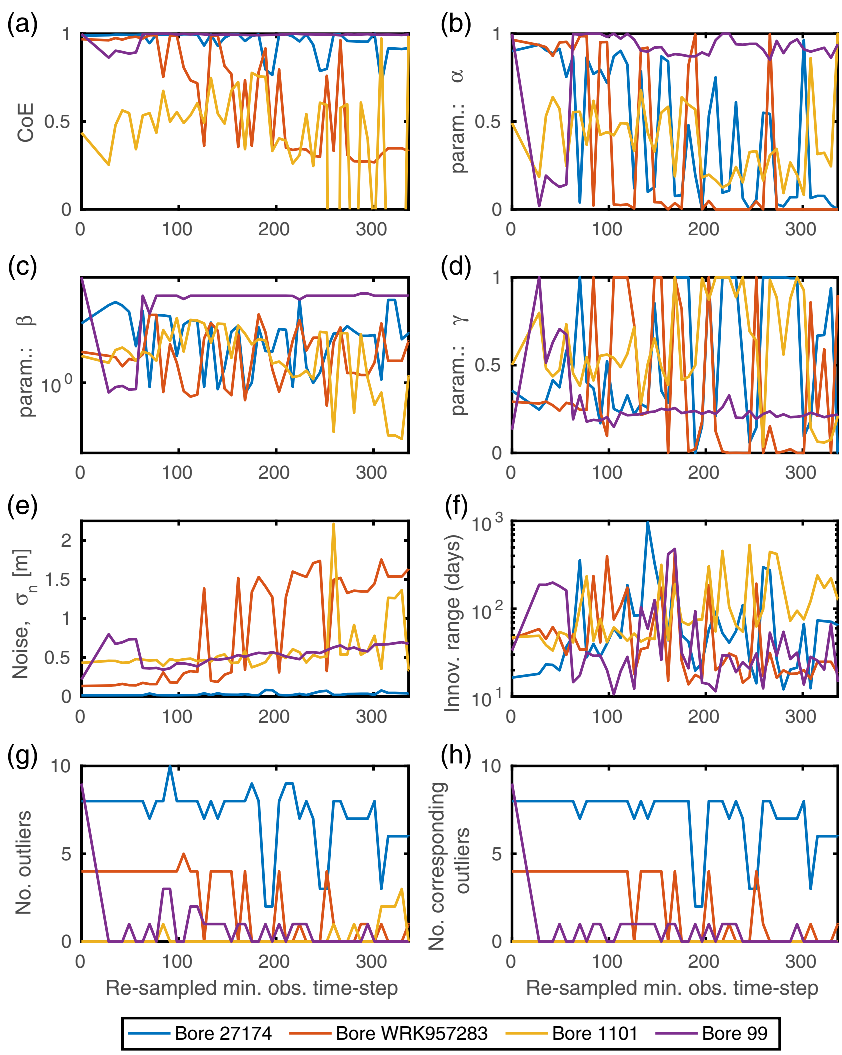

Fig. 4 Application of the error and outlier detection algorithms to the four bores from Fig. 3 using observations resampled to a minimum frequency of 28 days up to 364 days (7-day increments). Note, the results at an x-axis value of zero are those from the nonresampled data. Also, the identified outliers from Fig. 3 were inserted into each resampled data set to allow comparison of the results. **a** Shows the model fit after the removal of errors and outliers, expressed as the coefficient of efficiency; **b–d** show the corresponding time-series model parameters; **e** shows the standard deviation of the model noise; **f** shows the serial correlation of the model innovations as given by the model variogram range; **g** shows the number of outliers detected; **h** shows the number of outliers detected using the resampled data that correspond to those detected using the complete data

This indicates that the innovations remained serially correlated (note, given the exponential variogram, the variogram is at 95% of the sill for a time step 3 times the range: that is, some correlation remains at up to 180 days). The innovations are therefore not iid and hence the indicative probabilities for detection of outliers (section "Outlier detection algorithm") should be treated with caution. Figure 4e–f does however suggest that for seasonal or greater observation frequencies, the time-series model produces plausible estimates of the noise and innovations with relatively low serial correlation.

To investigate the resulting outlier detection, Fig. 4g shows the number of outliers detected. It shows that at each bore the number of outliers first fluctuates by 1–2 points with the sample time step, and beyond approximately 100 days, the fluctuations become larger, especially at 27174 and WRK957283. Figure 4h shows the number of outliers from the resampled data that correspond with outliers from the complete data. Comparing Fig. 4g–h shows that, when outliers are detected, all but 1–2 of them are the outliers detected from the non resampled data. This suggests that, while the probability of detecting an outlier decreases as the observation frequency declines, the outliers that are detected are likely to be consistent with those that would have been detected had there been higher frequency data.

The other factor likely to heavily influence the performance of the outlier detection is the user set parameter for the number of noise standard deviations for an outlier (η). Figure 5a shows the change in the number of outliers, relative to the number of outliers shown in Fig. 3, when η is varied from 2.5 to 6. Unsurprisingly, it shows that the number of outliers is inversely related to η. More interestingly, it shows that for each bore there exists a range for η over which the number of outliers is constant. Clearly the outliers that are detected are likely to be those that are more extreme, but encouragingly the number of outliers does not decline to zero when $\eta = 6$ for 3 of the four bores. Furthermore, Fig. 5b shows that below the threshold value for η the outliers that are detected include those that were detected in Fig. 3 at all bores except bore 99. Obviously the value of η at which the number of outliers converges is likely to be dependent upon the hydrograph dynamics, but this threshold point for η appears to provide a means for objectively setting η and in doing so objectively identifying the more extreme outliers.

Discussion and conclusion

This paper presents applied algorithms for automated detection of erroneous and outlier groundwater hydrograph observations. The algorithms detect a range of simple errors that can arise from issues ranging from bore condition to the process of recording data in the field to inputting data to agency databases. It was achieved using simple heuristics to remove implausible water level observations, with the remaining observations being examined for outliers by iterative application of a calibrated double-exponential smoothing time-series model.

Application of the algorithms requires only the observed groundwater hydrograph and no knowledge or assumptions of the aquifer characteristics (e.g. hydraulic properties) or drivers. The algorithms were demonstrated on four groundwater hydrographs with varying dynamics and, with the caveats from the "Introduction", were found to be capable of detecting errors and outliers where their number and timing is unknown a priori, including where the outliers are sequential with varying or similar values. Objectively assessing whether the identified outliers are accurate or erroneous data points is, however, intrinsically problematic because the cause of each outlier is unknown and may arise from recording errors or unusual but real natural processes.

The forecasting aspect of the outlier detection has the implication that rapid water level changes (relative to the entire hydrograph) are more likely to be identified as an outlier. Preferential or flood recharge, and possibly rapid pumping recovery, would cause such a rapid water level rise, while significant groundwater pumping from a low transmissivity aquifer may cause a rapid decline. If few water level observations are made during such events then, from the analysis in Fig. 4, the

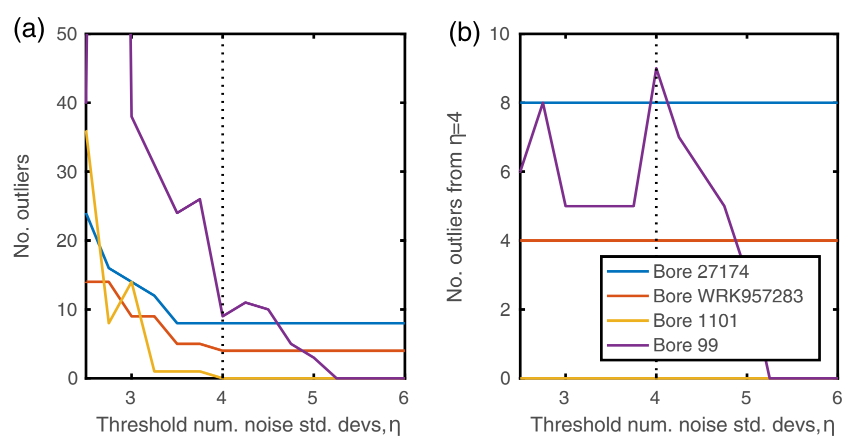

Fig. 5 Application of the error and outlier detection algorithms to the four bores from Fig. 3 using values for the number of noise standard deviations, η, ranging from 2.5 to 6. **a** Shows the number of outliers detected; **b** shows the number of outliers detected within the nonresampled data that were also detected using the resampled data

observations of the event are likely to be falsely detected as outliers. More generally, the analysis of sample frequency indicates that the detection of outliers becomes less reliable when the observations are less frequent than seasonal. The number of noise standard deviations (η) was shown to significantly influence the number of outliers detected and for each bore a range for η was identified, within which the number of outliers was constant and often greater than zero. For an observation frequency less than quarterly, trials varying η may improve the outlier estimation, but the interaction of observation frequency and η remains an open question. Furthermore, both are likely to be influenced by the observation record length, which van der Spek and Bakker (2017) showed to be more influential on groundwater time-series model prediction uncertainty than the observation frequency, a finding which clearly has implications for the residual noise estimation and hence the outlier detection.

A more challenging scenario was the detection of outliers when the groundwater level variance changes, with false negative outliers being more likely during the periods of low variance and false positive outliers being more likely during the periods of high variance. These challenges, while speaking to the diversity of groundwater hydrographs and the ongoing mathematical challenges of outlier detection (Chandola et al. 2009), illustrate that some caution is required when applying the outlier detection approach and when considering if identified outliers should be omitted from subsequent analysis. For such challenges, each period of relatively uniform variance and trend could be individually analyzed for outliers. Alternatively, more numerically demanding approaches may be required that use forcing data. Transfer function noise models are the obvious choice for such an extension and to date they have been developed to account for climate driven processes (Yihdego and Webb 2011; Peterson and Western 2014), groundwater pumping (von Asmuth et al. 2008; Shapoori et al. 2015a, b, c) and surface-water interactions (von Asmuth et al. 2008).

No automated method is however a replacement for local hydrogeological and hydrographic knowledge, and no matter the progress made in error and outlier detection, manual inspection of groundwater hydrographs is likely to remain a basic hydrogeological skill. While a numerical approach may identify an observation as an error or outlier, a decision will always be required whether to omit the point from subsequent analysis or to regard it as a valuable insight into unexpected aquifer behavior. For this reason, the implementation of the algorithms into the HydroSight toolbox (Peterson et al. 2017) allows the graphical manual editing of each type of error and outlier from each hydrograph. Coupled with the inbuilt summary statistics of the analysis, HydroSight offers an efficient, objective and reproducible means to analyze 1000s of observation bores and to prioritize those requiring manual inspection.

Acknowledgements This research was funded by the Australian Research Council Linkage Project LP130100958 and funding partners: Bureau of Meteorology (Australia); Department of Environment, Land, Water and Planning (Vic., Australia); Department of Economic Development, Jobs, Transport and Resources (Vic., Australia); and Power and Water Corporation (N.T., Australia). The authors are grateful to Dr. Elisabetta Carrara (Bureau of Meteorology) for her constructive input during the development of the algorithms.

References

Battaglia F, Orfei L (2005) Outlier detection and estimation in nonlinear time series. J Time Ser Anal 26:107

Berendrecht W, van Geer F (2016) A dynamic factor modeling framework for analyzing multiple groundwater head series simultaneously. J Hydrol 536:50–60

Chan W-S (1995) Understanding the effect of time series outliers on sample autocorrelations. TEST 4:179–186. https://doi.org/10.1007/BF02563108

Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Comput Surv 09(2009):15

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

Cipra T (2006) Exponential smoothing for irregular data. Appl Math :597

Li L, Wen Z, Wang Z (2016) Outlier detection and correction during the process of groundwater lever monitoring base on Pauta criterion with self-learning and smooth processing. In: Zhang L, Song X, Wu Y (ed) Theory, methodology, tools and applications for modeling and simulation of complex systems. Springer, Singapore, pp 497–503

Peterson TJ, Western AW (2014) Nonlinear time-series modeling of unconfined groundwater head. Water Resour Res 50:8330–8355. https://doi.org/10.1002/2013WR014800

Peterson TJ, Western AW, Shapoori V (2017) HydroSight: a toolbox for data-driven hydrogeological insights. http://peterson-tim-j.github.io/HydroSight/. Accessed August 2017

Shapoori V, Peterson T, Western A, Costelloe J (2015a) Estimating aquifer properties using groundwater hydrograph modelling. Hydrol Process 29:5424–5437. https://doi.org/10.1002/hyp.10583

Shapoori V, Peterson TJ, Western AW, Costelloe JF (2015b) Top-down groundwater hydrograph time-series modeling for climate-pumping decomposition. Hydrogeol J. https://doi.org/10.1007/s10040-014-1223-0

Shapoori V, Peterson TJ, Western AW, Costelloe JF (2015c) Decomposing groundwater head variations into meteorological and pumping components: a synthetic study. Hydrogeol J 23: 1431–1448. https://doi.org/10.1007/s10040-015-1269-7

Tremblay Y, Lemieux J-M, Fortier R, Molson J, Therrien R, Therrien P, Comeau G, Talbot Poulin M-C (2015) Semi-automated filtering of data outliers to improve spatial analysis of piezometric data. Hydrogeol J 23:851–868

van der Spek JE, Bakker M (2017) The influence of the length of the calibration period and observation frequency on predictive uncertainty in time series modeling of groundwater dynamics. Water Resour Res 53:2294–2311. https://doi.org/10.1002/2016WR019704

von Asmuth JR, Bierkens MFP (2005) Modeling irregularly spaced residual series as a continuous stochastic process. Water Resour Res 41. https://doi.org/10.1029/2004WR003726

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. https://doi.org/10.1111/j.1745-6584.2007.00382.x

Yihdego Y, Webb J (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. https://doi.org/10.1007/s10040-011-0726-1