AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2013WR014800

Nonlinear time-series modeling of unconfined groundwater head

T. J. Peterson[1] and A. W. Western[1]

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

Key Points:
- Groundwater hydrograph models cannot simulate episodic recharge and droughts
- Nonlinear transformation of climate forcing overcame these weaknesses
- Robust calibration required reformulation of existing time-series equations

Supporting Information:
- Readme
- Summary of model structures
- Groundwater statistical framework

Correspondence to:
T. J. Peterson,
timjp@unimelb.edu.au

Citation:
Peterson, T. J., and A. W. Western (2014), Nonlinear time-series modeling of unconfined groundwater head, *Water Resour. Res.*, *50*, doi:10.1002/2013WR014800.

Received 26 SEP 2013
Accepted 15 SEP 2014
Accepted article online 17 SEP 2014

Abstract This paper presents a nonlinear transfer function noise model for time-series modeling of unconfined groundwater hydrographs. The motivation for its development was that existing groundwater time-series models were unable to simulate large recharge events and multiyear droughts. This was because existing methods do not partition rainfall to runoff and do not account for nonlinear soil water drainage. To account for these nonlinear processes, a vertically integrated soil moisture module was added to an existing transfer function noise model. The soil moisture module has a highly flexible structure that allowed 84 different forms to be built. Application of the time-series model requires numerical calibration of parameters for the transfer functions, noise model and, for the nonlinear models, the soil moisture module. This was undertaken using the *Covariance Matrix Adaptation Evolutionary Strategy* (CMA-ES) global calibration scheme. However, reproducible calibration to the global optima was challenging and a number of modifications were required to the transfer function noise model. In trialing the 84 nonlinear models and 2 linear models, each was applied to eleven observation bores within a paired catchment study area in Great Western, Victoria, Australia. In comparison with existing groundwater hydrograph time-series models, the proposed nonlinear time-series model performed significantly better at all observation bores during calibration and evaluation periods. Both the linear and nonlinear models were also used to quantify the impact of revegetation within the paired catchment; however, results were inconclusive, which is likely due to time-series data for the state of the revegetation being unavailable. By analyzing the application of 84 nonlinear models to each bore, an optimal structure for the soil moisture module was identified. It is unlikely, however, that this model structure would be appropriate for all climates and geologies. To encourage further investigations, open-source code for the highly flexible groundwater time-series modeling framework is available and we invite others to develop new models.

1. Introduction

In most catchments, the water table level is the only observed hydrological state variable. Other observations, such a stream flow and precipitation, are fluxes and other state variables such as soil moisture are rarely observed, except through remote sensing of the top few centimeters. The water table level is also an aggregation of the drivers and processes within the catchment, and potentially provides quantitative insights into the catchment dynamics and impacts of management actions. Groundwater hydrographs have been widely utilized to qualitatively inform catchment dynamics, derive linear temporal trends and map the potentiometric surface. More recently, they have been widely used to calibrate groundwater models such as ModFlow [*Harbaugh et al*., 2000]. However, calibrating such models is often exceedingly difficult, resource intensive, frequently requires considerable simplification of the ecohydrology and hydrogeology and often does not achieve encouraging predictive performance. Despite these challenges, such bottom-up deterministic models remain a powerful tool for exploring hydrological dynamics and management scenarios; such as changes in groundwater pumping.

For undertaking tasks such as interpolating or extrapolating a groundwater hydrograph, predicting groundwater levels under different climate scenarios or decomposing the hydrograph to individual drivers, a top-down time-series approach is potentially a powerful yet simple option that has received insufficient attention. Development of such top-down approaches would enable significantly more quantitative use of groundwater data and to contribute to this, this paper presents a new nonlinear groundwater hydrograph time-series modeling framework. A range of model structures are developed and applied but, just like

rainfall-runoff modeling, no one model structure should be expected to work best everywhere and, hence, we encourage others to use the our open-source groundwater time-series modeling framework to test, develop and share their own groundwater time-series models.

The motivation for developing the groundwater time-series modeling framework was that existing linear groundwater time-series models [*von Asmuth and Bierkens*, 2005; *von Asmuth et al*., 2002; *Ferdowsian et al*., 2001, 2002] were unable to simulate the large recharge events and multiyear droughts common throughout Australia. One of the first groundwater hydrograph time-series models was the HARTT model [*Ferdowsian et al*., 2001, 2002; *Ferdowsian and Pannell*, 2009]. It has been widely used within Australia and models an unconfined groundwater level as the sum of the cumulative monthly rainfall residual (CRR), a linear time trend and, where appropriate, linear components for land cover change and pumping. Scalars on these components are estimated by multiple linear regression and, with the exception of *Ferdowsian and Pannell* [2009], serial correlation has not been included. A common use of the HARTT model is to quantify a linear trend in the head not explained by the CRR [*CSIRO*, 2009; *Sinclair Knight Merz*, 2005, 2006a; *Yihdego and Webb*, 2011]. However, the residuals often exhibit significant serial correlation. Furthermore, and more significantly, the coefficients of the linear time trend and CRR components can have a very large covariance, rendering the trend estimates very unreliable. This is most pronounced when applied to a hydrograph with an approximately linear temporal trend and where the CRR over the concurrent period is approximately linear [*Peterson and Western*, 2011].

Application of the HARTT model presents additional challenges. Its use of monthly CRR requires the omission (or averaging) of submonthly groundwater level observations from the model, and hence high frequency water level dynamics are omitted. The remaining observations must also be shifted to the end of each month. Additionally, the use of the CRR requires the assumption that the climate is first-order stationary and that historic rainfall events say, 100 years ago, influence the current groundwater level.

Alternatively, *von Asmuth and Bierkens* [2005] and *von Asmuth et al*. [2002] proposed a more complex approach applicable to water level observations with an irregular time step that does not assume a stationary climate. It is a transfer function noise model (TFN) composed of a deterministic component for the transformation of climate observations to a groundwater level displacement (from a constant) and an exponential noise model comparable to an auto-regressive lag 1 (AR1) model. Like other groundwater TFN models, such as *Fabbri et al*. [2011], the deterministic component is effectively the sum of the weighted historic precipitation and potential evapotranspiration record. For example, this component could be approximated for a given time point by applying, say, an exponentially decaying weight to all daily historic precipitation and summing to estimate the water level displacement from precipitation. The weights are, however, unknown. To estimate the weights, *von Asmuth and Bierkens* [2005] adopted a Pearson's Type III distribution function for the weights. It has three parameters and a flexible structure that allows the estimation of weighting function shapes ranging from a skewed Gaussian-like shape, simulating a time lag in the response of the groundwater level, to an exponential-like decay with no lag. The noise component simulates an exponential decay in the water level response not explained by the deterministic components. Adoption of an exponential decay allows the calibration of the model to irregularly spaced water level observations. This is possible because the exponential decay function allows the derivation of a relationship between the residuals and the innovations (i.e., the change in the residuals over an irregular time step when noise is accounted for, see *von Asmuth et al*. [2002]) and this allows the derivation of a weighted least squares objective function and avoidance of a computationally expensive Kalman Filter. With this, *von Asmuth and Bierkens* [2005] and *von Asmuth et al*. [2002] then apply a single start Levenberg-Marquardt gradient-based calibration to optimize the four to five model parameters to the observed water level. Despite use of such nonlinear regression techniques the TFN model is, however, a linear model. This is because the head at any time point is estimated as the weighted sum of past precipitation plus the noise component. Consequently, the response of the head to the climate is linear so that a doubling of a rainfall event produces twice the head response.

The linear TFN model of *von Asmuth and Bierkens* [2005] and *von Asmuth et al*. [2002] has been applied in a number of papers [*Lehsten et al*., 2011; *Yihdego and Webb*, 2011; *Manzione et al*., 2010] and, for most applications, the performance during the calibration and evaluation periods is very good. However, *Yihdego and Webb* [2011] found that the linear TFN model performed poorly at 30 of the 82 observations bores investigated throughout south west Victoria, Australia, and for the remaining 52 bores the calibration scheme had

to be repeatedly applied to achieve acceptable results. In other studies, the TFN model appeared to perform best when the hydrographs show only many small recharge events and no long-term drainage [*Lehsten et al*., 2011; *Manzione et al*., 2010]. To account for drivers in addition to climate, physically based weighting functions (henceforth, referred to as *impulse response functions*) have been added for pumping [*von Asmuth et al*., 2008; *Shapoori et al*., 2011] and stream flow [*von Asmuth et al*., 2008]. Like the inclusion of precipitation and PET, these additional drivers were assumed to act independently of one another, despite evidence to the contrary [*Zhou*, 2009]. For example, drawdown from pumping was assumed to not change recharge, evaporative loss or the interaction with streams. Recently, TFN models models have also been used to gain insights into the aquifer hydrological properties. For example, aquifer transmissivity and specific yield has been estimated for observation bores influenced by pumping [*Obergfell et al*., 2013], and *Bakker et al*. [2008] proposed a moment-based inverse modeling approach for estimating the transmissivity and specific yield using the Pearson's Type III impulse response function.

These applications, while encouraging, assume all rainfall produces a groundwater level response and, hence, ignore the nonlinear partitioning of rainfall to evapotranspiration and runoff and the nonlinearity of soil drainage. This is obviously not realistic but is unlikely to have been problematic because many of the above applications were to sites with small frequent recharge events. If the linear TFN model were applied to a site having large infrequent rainfall events that produce runoff and a large head response, then the model linearity would likely result in the poor simulation of large recharge events and acceptable simulation of smaller events, or vice versa, but not acceptable simulation of both large and small recharge events. If an additional driver, such a groundwater pumping, were input to the model then the numerical calibration may attempt to adjust this driver to make up for these climatic deficiencies, resulting in compromised estimation of the impact from the pumping and the derived parameters, such as the aquifer hydraulic properties. A time-series model that did account for such nonlinearity should partition rainfall so that large events produce runoff and small events produce little, if any, runoff. Assuming runoff is driven by saturation excess, then the runoff from both events should be a function of the antecedence catchment wetness and, most importantly, the soil vertical drainage should be a function of the soil wetness. A time-series model should also account for the nonlinear response to PET whereby vegetation may preferentially uptake groundwater during dry periods. Clearly such features of a nonlinear model are driven by soil moisture dynamics, but to date the only known study accounting for such nonlinearity is *Yi and Lee* [2004]. They used a 1-D finite element vadose zone model to transform the precipitation to an infiltration rate, which was then input to a Kalman filter TFN model. While the nonlinear filtering of precipitation produced a clear improvement, the soil and TFN models were not jointly calibrated, nor were alternative soil model structures explored or PET accounted for within the TFN model. In this paper, these weaknesses are addressed by using a simple yet highly flexible soil moisture model that can transform precipitation and PET for input to the impulse response functions. Hence, following *Box et al*. [2008], the model developed herein is not strictly a nonlinear time-series model, but a nonlinear transfer function time-series model.

In the following, the methods section presents the study area; the modified linear and nonlinear transfer function noise models; the selection, implementation and testing of a global calibration scheme; and the implementation of models to the study area. The results section then details the reproducibility of the global calibration scheme; the performance of the linear and nonlinear models for the study area; estimation of the impacts from the revegetation using the linear and nonlinear TFN models; and an example of the temporal decomposition of hydrographs. Lastly, the discussion and conclusions section summarizes the findings and discusses the optimal structure for the nonlinear model and possible further developments using the open-source modeling framework.

2. Methodology

The effectiveness of a nonlinear time-series model is a function of both the model structure and the global calibration scheme. A complex model structure may be physically realistic but if, in optimizing its parameters, there are numerous optima, the objective function response surface is not smooth or the parameters are highly correlated, then the calibration may be ineffectual and/or be exceedingly slow, rendering the time-series modeling impractical. Such challenges were reported by *Yihdego and Webb* [2011], and all arose in the development of our linear and nonlinear TFN models. This occurred despite the linear TFN model

simulating the head as a linear sum of weighted climate forcing, because the weights are from an analytical equation that must be fitted to the observed heads using nonlinear regression. For these reasons, we put considerable effort into developing robust TFN model structures and an efficient and reproducible global optimization scheme. In the subsections below, the application study area is outlined, following which the linear TFN is outlined and a number of refinements are detailed. The nonlinear TFN model is then presented, followed by the calibration scheme and the implementation of the linear and nonlinear TFN models to the study area.

2.1. Study Area

The TFN models were applied to a paired catchment study area, located seven kilometers east the town of Great Western in western Victoria, Australia. The catchments are on the north-west face of Tucker Hill and are within the upper catchment of the Wimmera River. This region contains significant dryland salinity and contributes a significant mass of salt to the Wimmera River [*Sinclair Knight Merz*, 2006b]. Figure 1 details the topography, catchment boundaries, land cover, geological units, stream courses and observation bore locations. The annual average precipitation and ETo were estimated to be 487 mm yr^{-1} and 1066 mm yr^{-1} respectively [*Peterson*, 2009] and the site is classified as Köppen-Geiger climate type *Cfb*, that is, temperate with a warm summer and without a dry season [*Peel et al*., 2007]. Furthermore, as shown by the box-plots of the seasonal climate forcing in Figure 2, the winters are cool and wet, the summers dry and the variability in monthly ETo is low relative to precipitation.

The south-west catchment is the control catchment and has annual pasture vegetation, the dominant land cover for grazing within the region. It is estimated to be 33.9 ha in area [*McIntyre*, 2003] and the catchment is relatively uniform in width, with modest convergence. The north-western boundary is uncertain because of the low slope.

The north-east catchment is the treated catchment and the land cover consists of tall wheatgrass and perennial pastures in the lower catchment and various tree species in the upper catchment. It is estimated to be 54.2 ha in area [*McIntyre*, 2003] and the catchment width is relatively narrow in the upper catchment, widens within the midslopes and converges toward the outlet. The northern boundary is uncertain because of the low slope. The perennial pastures consist of phalaris, cocksfoot and subclovers and approximately 29.4 ha were sown in June 1996. In July 1996, approximately 14.9 ha were planted with 10,720 native seedlings (*Acacia implexa*, *Allocasuarina vertticillata* and various *Eucalyptus* species) [*McIntyre*, 2003]. A survey estimated the average stand height at 2.8 m and the survival rate at 90.3% in 2003 [*McIntyre*, 2003].

The catchments both have two distinct landforms and geologies. Drilling and field mapping found the upper slopes to be Lower Ordovician meta-sediments (approximately 470 million years old) and the lower slopes to be intrusive Devonian biolite granites (approximately 400 million years old) [*HydroTechnology*, 1995]. The Devonian granites occasionally outcrop, but more commonly have a sand dominated soil profile. They are heavily weathered to a depth of 15–35 m [*Sinclair Knight Merz*, 2006b]. The Ordovician meta-sediments occur mainly as meta-siltstone and shales, form the aquifer basement for the region, and are slightly to moderately weathered to a depth of approximately 50 m. Twelve observation bores were drilled in 1995 and six of the twelve observation bores were monitored electronically twice daily and the remaining six bores were monitored monthly. Records span from July to August 1995 to January 2008. For further details of the observation data see Table 1 and *Peterson* [2009, section 7.2].

2.2. Time-Series Models

2.2.1. Linear Transfer Function-Noise Model

The *von Asmuth and Bierkens* [2005] and *von Asmuth et al*. [2002] transfer-noise function (TFN) model, if applied to regularly spaced observation, is effectively an auto-regressive moving-average externally forced (ARMAX) time-series model with a lag 1 auto-correlation and a weighting of all historic external forcing data. However, unlike traditional ARMAX models, it does not require uniformly spaced observations without gaps, which makes it ideal for groundwater applications. This was achieved by adopting an exponential noise model for simulation of the part of the water level response not explained by the deterministic components. As stated above, this allowed the derivation of a relationship between the residuals and the innovations (i.e., the change in the residuals over an irregular time step when noise is accounted for, see *von Asmuth et al*. [2002]). This allowed the derivation of a weighted least squares objective function, avoiding a computationally expensive Kalman filter.

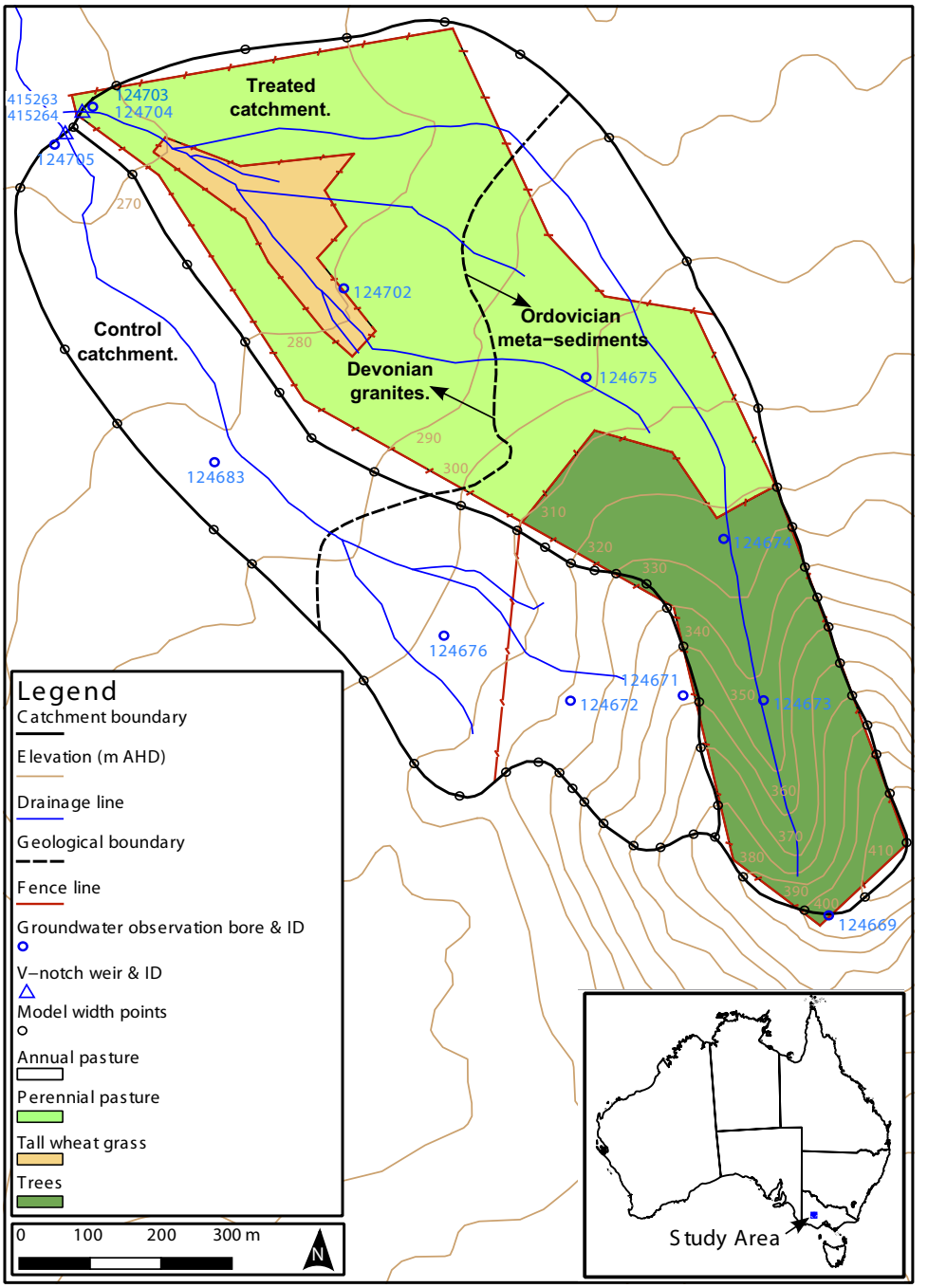

Figure 1. Control and treated catchment boundaries, elevation, bores and weir locations and land cover. Note: bores 124703 and 124704 are nested. All mapped data were adapted from *Sinclair Knight Merz* [1996, 1997].

Following *von Asmuth and Bierkens* [2005] and *von Asmuth et al.* [2002], the observed groundwater level, h_t [L], at time, t [T], is modeled as:

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

where h_t^* [L] is the deterministic component predicting the groundwater level at t due to all the external forcing, such as climate, and relative to a constant water level. This constant, d [L], is a variable set after model calibration to produce a zero mean error. ϵ_t [L] is the residual groundwater level at time, t. Expanding h_t^* to a weighted integral of past precipitation gives the simplest linear TFN model:

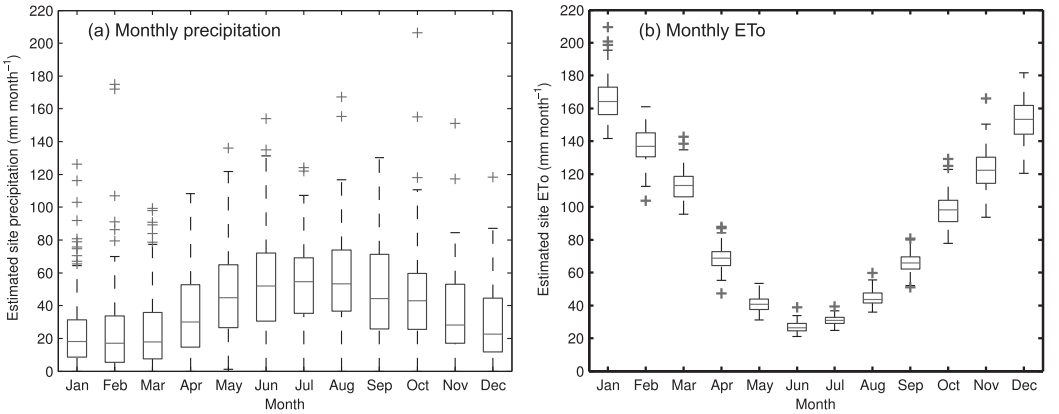

Figure 2. Box-plots of monthly SILO (a) precipitation and (b) ETo. The whiskers extends a maximum of 1.5 times the interquantile range. Values beyond the whisker are marked as outliers and are denoted as +.

$$h_t = \int_{\tau=-\infty}^{t} P_\tau \theta(t-\tau)\partial\tau + \epsilon_t + d \tag{2}$$

where P_τ [L T^{-1}] is the daily precipitation at time τ [T] and $\theta(t-\tau)$ is the impulse response function. The impulse response function was defined by *von Asmuth and Bierkens* [2005] as a Pearson type III function:

$$\theta(t) = A\frac{b^n t^{n-1}\exp(-bt)}{\Gamma(n)} \tag{3}$$

where A, b and n are nonphysical parameters and Γ denotes a gamma function. Using the Pearson type III function and a one parameter exponential noise model for ϵ_t (see equation (22)), the linear TFN model (equation (2)) has four parameters. By accounting for PET, *von Asmuth et al*. [2008] extended equation (2) to a five parameter model:

$$h_t = \int_{\tau=-\infty}^{t} P_\tau \theta(t-\tau)\partial\tau - \int_{\tau=-\infty}^{t} f_E E_\tau \theta(t-\tau)\partial\tau + \epsilon_t + d \tag{4}$$

where E_τ [L T^{-1}] is the daily PET and f_E is a parameter scaling the impulse response function, θ, from the precipitation to that for the PET. Throughout this paper, the five parameter linear TFN model for the non-revegetated catchment is that given by equation (4) but with a rederived impulse response function.

Table 1. Summary of Observation Bore Construction, Lithology, and Salinity

Bore ID	Elevation (m AHD[a])	Depth (m)	Screening (m) From	To	Lithology[b]	Salinity (μS cm^{-1})	Data-logger?
Control Catchment							
124705	269.18	7.9	3	7	Dgr	1030	N
124683	286.03	28	21.5	26.5	Dgr	1015	Y
124676	302.10	23	19.2	22.2	Oms	1114	N
124672	323.01	45.8	37	40	Oms	1610	N
124671	357.22	61.1	55	60	Oms	1105	Y
Treated Catchment							
124704	269.59	7.64	2.6	6.6	Dgr	7440	Y
124703	269.35	35.4	29	35	Dgr	9080	N
124702	280.72	10	7	9	Dgr	5110	Y
124675	299.22	38	35	37	Oms	2138	N
124674	323.30	45	39	44	Oms	738	Y
124673	341.01	50	40	42	Oms	1097	N
			46	49			
124699	404.00	122	61	66	Oms	1069	Y
			115	121			

[a]AHD stands for Australian Height Datum.
[b]*Dgr* denotes Devonian granites and *Oms* denotes Ordovician meta-sediments.

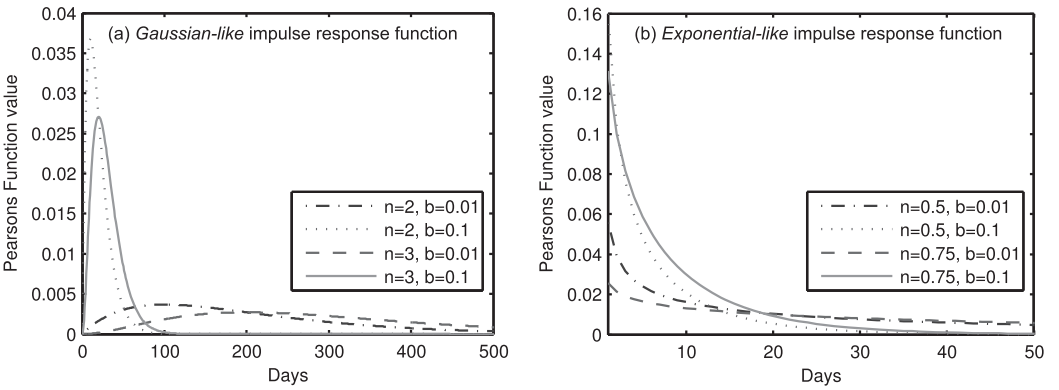

Figure 3. Example Pearsons Type III impulse response functions for $A = 1$ and eight combinations of parameters b and n that produce (a) a *Gaussian-like* and (b) an *exponential-like* response function. It shows that b and n change the function peak value, which is also influenced by parameter A, and this indicates parameter covariance.

von Asmuth et al. [2008] also proposed an approach to account for vegetation clearing. The approach assumes clearing only changes evapotranspiration and does not change interception, infiltration or recharge. The approach consists of a third integral added to those within equation (4) and is a weighting of the ET integral by an additional parameter, f_L, and the historic fraction of the clearing, L_τ:

$$h_t = \int_{\tau=-\infty}^{t} P_\tau \theta(t-\tau)\partial\tau - \int_{\tau=-\infty}^{t} f_E E_\tau \theta(t-\tau)\partial\tau - \int_{\tau=-\infty}^{t} f_L L_\tau E_\tau \theta(t-\tau)\partial\tau + \epsilon_t + d \tag{5}$$

Throughout this paper, the six parameter linear TFN model for the revegetated catchment is that within equation (5).

The impulse response function, θ, is central to the TFN models. *von Asmuth and Bierkens* [2005] and *von Asmuth et al.* [2002] adopted a Pearson type III function because of its flexible form. To illustrate, Figure 3 shows θ calculated for $A = 1$ and eight combinations of parameters b and n. In addition to showing the exponential-like and skewed Gaussian-like shapes, it shows that b and n change not only the shape of the function but also its peak value. Considering that parameter A of equation (3) rescales the response function, and hence the peak value, parameters A, b and n of equation (3) have considerable covariance. This weakness was one of five arising when calibrating TFN models using the Pearson type III function. Solutions to all five weakness are detailed in Appendix A and, throughout this paper, all application of the TFN models used the revised impulse response function within Appendix A. An overview of the additional four weaknesses is given below.

The second modification addressed parameter transformations. In trialling the linear (and nonlinear) TFN models, the parameter n (see equation (A11)) ranged over five orders of magnitude from less then 0.001 one to over three hundred, postcalibration. A similar range also occurred for the other two parameters, A and b. This caused the calibration problem to often be poorly scaled, which compromised the calibration reproducibility. This problem was overcome by undertaking a $\log_{10}$ transformation of the parameters.

The third modification addressed the problem of exponential-like functions with very low values of b having a significant value at long times, as illustrated in Figure 10b. This implies that the climate several decades or even a century ago influences the current groundwater level. This is unrealistic and led to very low values of the drainage elevation constant (see d within equation (4)). To address this problem, the exponential-like function was modified by subtracting its value at a fixed number of decades priors to the first climate observation.

The fourth modification addressed the integration of the function tail from the time lag at the first climate observation to negative infinity (i.e., infinity days prior to the first climate observation). If the function was exponential-like, integration was undertaken to the fixed number of decades prior to the first climate observation. This integration was undertaken because, without the tail integration, spurious parameter estimates resulted from the calibration and the climate decomposition (see section 3.4) often showed the influence of the climate at the longest time lag to increase throughout the simulation period.

The fifth modification was undertaken to minimize rounding errors in the numerical estimation of the integrals. Specifically, in equation (3), the term t^{n-1} can become very large when $n \gg 1$ and a very long daily climate record is input. When numerically approximating the integrals they can, and did, exceed machine precision and rounded to ∞. To address this problem the order of operations was modified when $n \gg 1$.

2.2.2. Nonlinear Transfer Function-Noise Model

The TFN model of section 2.2.1, despite its apparent complexity, is a linear model. This is because the deterministic component of the groundwater level, at a time point, is estimated as the sum of historic precipitation weighted by the impulse response function. Each day's climate is independent from each other day and no transformation is applied to the climate data. Hence, being a linear model a rainfall event of 50 mm would produce double the groundwater rise of a 25 mm rainfall event. Physically, however, the water table does not respond linearly to the climate. Large rainfall events will produce recharge but there may be considerable *loss* to interception and runoff, which themselves are functions of antecedent conditions such as the soil moisture. Small events may produce very little recharge because of interception and the rainfall that does infiltrate may be transpired by vegetation or be insufficient to increase soil moisture enough to produce recharge. Furthermore, during very dry periods, vegetation may transpire groundwater. In summary, many unconfined aquifers respond very nonlinearly to climatic events and is influenced by various antecedence conditions resulting from soil and interception storages.

Clearly, a nonlinear groundwater time-series model that is of practical use cannot account for all such nonlinearities and its parameters must be computationally efficient to calibrate. In building a nonlinear groundwater time-series model, the linear transfer function from section 2.2.1 was adopted and a means was sought for transforming the climate forcing for input into the transfer function. A biophysically based means of transforming climate data is a 1-D soil moisture model. For a groundwater hydrograph with infrequent, say seasonal, observations there is insufficient justification for adopting a vertically discretized soil moisture model. That is, the soil model is only to be used to transform the precipitation signal into a recharge or infiltration signal and the soil parameters are to be estimated only from fitting to an infrequently observed groundwater signal; not subdaily soil moisture observations. Furthermore, while a vertically discretized model would allow simulation of heterogeneous soils and nonfree draining conditions, the computational demands would be prohibitive considering simulation is required for each day of the input climate data. Hence, a parsimonious vertically lumped soil model was sought. The implications of this are likely to be minimal if the vegetation can extract soil water from locally wet regions of the soil profile [*Guswa et al*., 2002].

In selecting a vertically integrated soil model, there is no shortage of choices. The chosen model should, however, produce a smooth calibration response surface when the recharge, infiltration or the soil moisture signals are input to the TFN model; have minimal parameters; capture the major dynamics of infiltration, ET and vertical drainage; and allow exploration of different structural forms because the most parsimonious model structure is unknown. While there are multiple models meeting these requirements (and the groundwater time-series framework does allow exploration of alternative structures), the nonlinear saturated pathway model of *Kavetski et al*. [2006] was adopted:

$$\frac{dS}{dt} = P\left(1 - \frac{S}{\hat{S_{cap}}}\right)^{\alpha} - \hat{k_{sat}}\left(\frac{S}{\hat{S_{cap}}}\right)^{\hat{\beta}} - E_t\left(\frac{S}{\hat{S_{cap}}}\right)^{\gamma} \tag{6}$$

where $S \in (0, \infty)$ [L] is the soil water storage; t [T] is time; P_t [L T^{-1}] is the precipitation rate at time t; E_t [L T^{-1}] is the potential evapotranspiration rate at time t; $\hat{S_{cap}} \in [1, \infty)$ [L] is an antilog transformed parameter for the maximum soil water storage ($\hat{}$ denotes a transformed parameter); $\hat{k_{sat}} \in (0, \infty)$ [L T^{-1}] is an antilog transformed parameter for the vertical saturated conductivity; $\alpha \in [0, \infty)$ is a dimensionless parameter controlling the infiltration rate as a function of catchment wetness; $\hat{\beta} \in [1, \infty)$ is an antilog transformed dimensionless parameter controlling the free drainage as a function of wetness; and $\gamma \in [0, \infty)$ is a dimensionless parameter controlling the soil evaporation as a function of wetness. Additionally, the aforementioned antilog transformations were undertaken to scale the parameters to be of a similar range and sensitivity to the other time-series parameters. This was undertaken to improve the calibration reproducibility. Specifically, transformations were:

$$\hat{S_{cap}} = 10^{S_{cap}} \tag{7}$$

$$\hat{k_{sat}} = 10^{k_{sat}} \quad (8)$$

$$\hat{\beta} = 10^{\beta} \quad (9)$$

where $S_{cap} \in [0, \infty)$ [L] is the nontransformed parameter for the maximum soil water storage; $k_{sat} \in [\infty, \infty)$ [L T^{-1}] is the nontransformed parameter for the vertical saturated conductivity; and $\beta \in [0, \infty)$ is the dimensionless nontransformed parameter controlling the free drainage as a function of wetness.

A practical feature of the soil model in equation (6) is that different functional forms of the model can be explored by fixing parameters to zero or one. For example, the full nonlinear soil model requires five parameters but fixing α, $\hat{\beta}$ and γ to one converts the model to a linear model and requires calibration of only two soil model parameters; S_{cap} and k_{sat}. Furthermore, components can be *removed* by setting selected parameters to zero. For example, fixing α to zero causes all rainfall to infiltrate until the soil layer is saturated; fixing $\hat{k_{sat}}$ to zero removes the free-drainage component; and fixing γ to zero forces the actual ET to equal to potential ET.

Using equation (6) in the TFN model requires solution of the ODE at a daily time step from the first climate observation to the last day of simulation. Traditionally, such hydrological ODEs have been solved by rearranging to a fixed step explicit solution. However, *Kavetski and Clark* [2010] demonstrated that such an approach can produce considerable discontinuities in the calibration response surface, produce spurious local optima and shift the parameter location of the global optima. A fixed time step semi-implicit solver with Newton's method convergence to an error approaching machine precision was shown to eliminate such numerical artifacts, and in solving equation (6) such an approach was implemented. For the numerically inclined, details are given within the open-source file named *forcingTransform_soilMoisture.c*.

With the ODE solved, there are a number of options for how to use it to transform the precipitation and PET for input to the TFN model. With regard to replacing precipitation in equation (4), the soil infiltration rate could be adopted. This transformation would effectively remove runoff from the precipitation signal. Alternatively, the free-drainage from the soil layer could be adopted. Assuming $10^{\beta} \gg 1$, this would transform the precipitation to a more event driven signal whereby only large precipitation events would increase the soil moisture sufficiently for free-drainage to occur. With regard to the evaporation term within the transfer function (equation (4)), the soil evapotranspiration could be adopted in place of the potential ET. Alternatively, the groundwater potential evapotranspiration (i.e., potential ET minus the simulated soil ET) could be adopted to simulate longer term drying and wetting dynamics. During dry periods, the soil would be dry and soil evaporation would be low. The groundwater potential ET would, however, be high and could be used to simulate drought conditions. Biophysically, this can be interpreted as the soil moisture being insufficient to supply the vegetation water demand so the vegetation uptakes groundwater. While the uptake may not be at the modeled observation bore, uptake in nearby areas would cause a change in the lateral flux at the observation bore, and hence the water level.

In developing this nonlinear time-series model, it is not known which of these flux combinations is best. Clearly fluxes relating to inflow or uptake from the water table are the most physically plausible. However, if the soil model is considered as just a nonlinear climate signal filter then potentially less physically plausible fluxes may be most applicable at selected bores. For example, at sites with a shallow water table the dominant nonlinear process may be saturated excess runoff, and not soil free drainage, making the soil infiltration rate the most appropriate flux to replace precipitation in the time series model. Therefore it is difficult to justify a particular soil model structure and fluxes a priori and so all model structure combinations were investigated using each combination of the four fluxes. Furthermore, models using groundwater potential ET were only able to adequately fit the observed head when the ET component of the TFN model had its own impulse response function, rather than one scaled from the precipitation. Hence, separate impulse response functions were adopted for the transformed precipitation, θ_P, and transformed ET, θ_E. This added two parameters to the time-series model.

Importantly, each of the adopted soil moisture model fluxes change throughout the daily time step and, in this application, the fluxes for input to the TFN model were taken as the integral over the daily time step. This was undertaken so that the evaluation of the nonlinear TFN models was consist with the linear TFN models; which required input of daily integrated climate forcing (i.e., daily precipitation and PET) and subsequent multiplication by the trapezoidal integral of the impulse response function over the corresponding

day (note, the time-series framework allows adoption of instantaneous fluxes (e.g., soil free drainage) and hence more accurate numerical integration; specifically, 3/8 Simpson's composite integration). In calculating the daily integral of the soil moisture model fluxes, fixed subhourly time steps were found to produce very low mass balance errors. Use of daily time steps did produce significantly larger mass balance errors but this was found to be a result of the trapezoidal integration of the soil fluxes, and not due to errors in the solution of equation (6). Finally, in transforming the precipitation term within the linear TFN model to either the infiltration rate or the normalized soil free drainage, the fluxes were calculated as:

$$I_t = \frac{S_t - S_{t-1}}{\Delta t} + \frac{\Delta t}{2}\frac{S_t^{\hat{\beta}} + S_{t-1}^{\hat{\beta}}}{\hat{S_{cap}}^{\hat{\beta}}} + \frac{\Delta t}{2}\frac{S_t^{\hat{\gamma}} + S_{t-1}^{\hat{\gamma}}}{\hat{S_{cap}}^{\hat{\gamma}}} \tag{10}$$

$$D_t = \frac{\Delta t}{2}\frac{S_t^{\hat{\beta}} + S_{t-1}^{\hat{\beta}}}{\hat{S_{cap}}^{\hat{\beta}}} \tag{11}$$

where I_t [L T^{-1}] is the infiltration rate and D_t [L T^{-1}] is the normalized soil free drainage rate at time t. Importantly D_t does not depend upon k_{sat} and, hence, the normalized free drainage can be used in the TFN model without the soil model simulating free drainage. This simplification to the biophysics was more parsimonious and was considered acceptable because free drainage may be a small percentage of total soil water flux and, hence, an adequate transformation of the climate data into a normalized recharge signal may be achievable without simulation of free drainage within the soil moisture model.

For transforming the PET term within the linear TFN model to either the soil evaporation or the groundwater potential evaporation, the fluxes were calculated as follows:

$$V_t = E_t \frac{\Delta t}{2}\frac{S_t^{\hat{\gamma}} + S_{t-1}^{\hat{\gamma}}}{\hat{S_{cap}}^{\hat{\gamma}}} \tag{12}$$

$$G_t = E_t - V_t \tag{13}$$

where V_t [L T^{-1}] is soil evapotranspiration rate and G_t [L T^{-1}] is the groundwater potential evapotranspiration rate at time t. Drawing on these four flux equations, the first nonlinear time-series model with the drainage-soil ET transformation was as follows. Note that the right most integral is for the revegetation and, like the linear TFN model, it simulates a change in groundwater update. For the non-re-vegetated catchment this integral was omitted from all nonlinear TFN models.

$$h_t = \int_{\tau=-\infty}^{t} D_\tau \theta_P(t-\tau)\partial\tau - \int_{\tau=-\infty}^{t} V_\tau \theta_E(t-\tau)\partial\tau - \int_{\tau=-\infty}^{t} f_L L_\tau V_\tau \theta_E(t-\tau)\partial\tau + \epsilon_t + d \tag{14}$$

The second nonlinear time-series model with the drainage-groundwater potential evaporation transformation was:

$$h_t = \int_{\tau=-\infty}^{t} D_\tau \theta_P(t-\tau)\partial\tau - \int_{\tau=-\infty}^{t} G_\tau \theta_E(t-\tau)\partial\tau - \int_{\tau=-\infty}^{t} f_L L_\tau G_\tau \theta_E(t-\tau)\partial\tau + \epsilon_t + d \tag{15}$$

The third nonlinear time-series model was the infiltration-soil ET transformation model.

$$h_t = \int_{\tau=-\infty}^{t} I_\tau \theta_P(t-\tau)\partial\tau - \int_{\tau=-\infty}^{t} V_\tau \theta_E(t-\tau)\partial\tau - \int_{\tau=-\infty}^{t} f_L L_\tau V_\tau \theta_E(t-\tau)\partial\tau + \epsilon_t + d \tag{16}$$

And, the fourth nonlinear time-series model for the infiltration-groundwater potential evaporation transformation was:

$$h_t = \int_{\tau=-\infty}^{t} I_\tau \theta_P(t-\tau)\partial\tau - \int_{\tau=-\infty}^{t} G_\tau \theta_E(t-\tau)\partial\tau - \int_{\tau=-\infty}^{t} f_L L_\tau G_\tau \theta_E(t-\tau)\partial\tau + \epsilon_t + d \tag{17}$$

2.3. Calibration Objective Function and Noise Model

To apply the linear and nonlinear TFN models, the parameters for the impulse response functions and the constant, d, need to be estimated and the residuals, ϵ_t, statistically modeled. Following *von Asmuth and Bierkens* [2005], the parameters were numerically optimized to best fit an observed hydrograph using a weighted least squares objective function. The weighted objective function was adopted so that a statistical model for the residuals could be derived that is applicable to irregularly spaced groundwater level observations. Following *von Asmuth and Bierkens* [2005], for practical reasons, the residuals were modeled as an exponential weighting of previous time-step residuals. For regularly spaced data, this reduces to an autoregressive lag 1 model (AR1) plus a white noise process having a zero-mean, homoscedastic and serially uncorrelated Gaussian distribution. Additionally, in applying this model to irregularly spaced observations, the weighted least squares objective function used the innovations (i.e., the change in the model residuals over an irregular time step) rather than the model residuals. In the following, the equations from *von Asmuth and Bierkens* [2005] are reproduced and a modification is presented to increase the numerical stability.

The residuals at a time point were calculated as:

$$\tilde{\epsilon_t} = h_t - h_t^* - d \tag{18}$$

where $\tilde{\epsilon_t}$ [L] is the residual at time t; h_t [L] is the observed groundwater elevation at time t; h_t^* [L] is the deterministic component predicting the groundwater level at t due to all external forcing; and d [L] is the 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:

$$d = \bar{h} - \overline{h_*} \tag{19}$$

where $\bar{h}$ [L] is the mean observed groundwater level over the calibration period; and $\overline{h_*}$ is the mean of the deterministic components, h_t^*, over the calibration period. Next, the innovations, υ_t [L], over an irregular time step were defined by assuming an exponential decay of the residuals:

$$\upsilon_t = \tilde{\epsilon_t} - \tilde{\epsilon_{t-\Delta t}} e^{-\mu \Delta t} \tag{20}$$

where $\tilde{\epsilon_{t-\Delta t}}$ [L] is the residual at the previous groundwater observation time; Δt [T] is the variable time step to the previous groundwater observation; and μ is a dimensionless parameter defining the decay rate of the residuals. Using this equation for the innovations, *von Asmuth and Bierkens* [2005] derived an objective function used to optimize all model parameters, include the noise parameter μ. The objective function is a simplified likelihood function. It is referred to as the sum of weighted squared innovations (SWSI) criterion, and was defined by *von Asmuth and Bierkens* [2005] as:

$$\chi^2 = \sum_{j=1}^{N} \frac{\left\{\prod_{i=1}^{N} \left(1 - e^{-2\mu \Delta t_i}\right)\right\}^{-N}}{1 - e^{-2\mu \Delta t_j}} \upsilon^2\left(t_j, \Omega\right) \tag{21}$$

where χ^2 is the SWSI criterion; N is the number of groundwater observations within the calibration period; and $\upsilon^2\left(t_j, B\right)$ is the innovations at time step j and for parameters Ω. The fraction to the left of $\upsilon^2\left(t_j, B\right)$ defines the weight applied to each innovation and is a function of the time step size and the noise parameter. Effectively, it is the mean noise variance across all time steps divided by the noise variance at each time step.

A weakness of equation (21) is that for small values of μ and small time steps, the numerator $\left(1 - e^{-2\mu \Delta t_j}\right)$ approaches zero. When a large number of such time steps are multiplied, the numerator within equation (21) approaches machine precision and rounds to zero. This causes a zero weight to be applied to all innovations. In developing this paper, this problem frequently occurred. To overcome this weakness, the term $1 - e^{-2\mu \Delta t_j}$ was transformed to a natural log. Values close to zero hence became large negative values. Furthermore, this required the product term of equation (21) to become a summation in log space, an operation less likely to approach machine precision. To convert out of log space, the mean was first derived and then the exponential calculated. This modification significantly increased calibration stability. The objective function applied to all models was as follows:

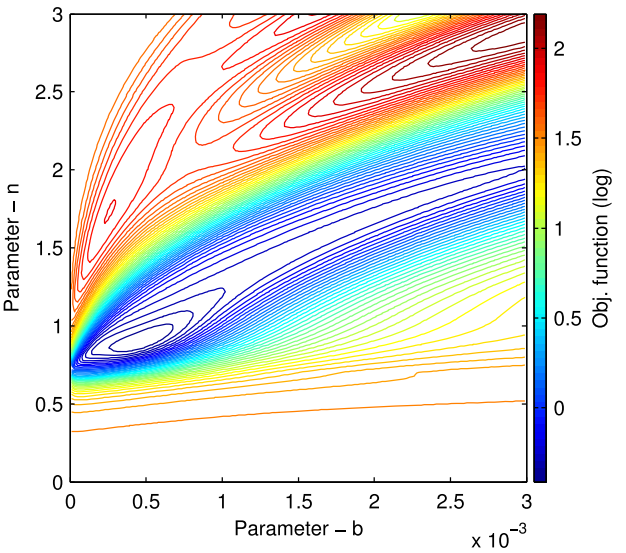

Figure 4. Two-parameter response surface illustrating the complexity of the model calibration; specifically, the multiple optima, complex valleys and regions of low sensitivity. The response surface was derived using the five parameter linear TFN model. Note, the remaining three parameters were fixed to that at a prior determined optima.

$$\chi^2=\sum_{j=1}^{N}\frac{\exp\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\ln\left(1-e^{-2\hat{\mu}\Delta t_i}\right)\right]\right\}}{1-e^{-2\hat{\mu}\Delta t_j}}\upsilon^2\left(t_j,B\right) \quad (22)$$

where $\hat{\mu}$ is the antilog transform of μ and $\hat{\mu}=10^{\mu}$. Like the soil moisture and impulse response functions, this transformation was undertaken to scale the parameter to a range and sensitivity similar to the other model parameters.

2.4. Model Calibration

The successful application of the linear or nonlinear TFN models is dependent upon a robust global calibration scheme. That is, assuming there is only one global optima, repeatedly applying a global calibration scheme should result in the same parameter estimates. However, calibration of either the linear or nonlinear TFN models is nontrivial. To illustrate the challenges, Figure 4 shows a two-parameter response surface derived from the five parameter linear TFN model. It shows two optima (at b $\leq$ 0.0005), three valleys and two regions of very low sensitivity. Following *von Asmuth et al*. [2002, 2012], the highly efficient Levenberg-Marquardt local calibration [*Marquardt*, 1963; *Levenberg*, 1944] scheme was initially applied. However, as expected from Figure 4, the optima to which such a local calibration scheme converges was very dependent upon the initial parameter values and, furthermore, partitioning a parameter region likely to contain the global optima was highly dependent upon the forcing and observation data. To address this, Levenberg-Marquardt was extended to a trust region approach [*Fan and Pan*, 2006] using eigenvalues and cluster analysis to dynamically select the number and location of initial starts [*Tu and Mayne*, 2002]. However, despite considerable extension to *Tu and Mayne* [2002] and subsequent testing, repeated application of a five parameter linear TFN and nine parameter nonlinear TFN model to four control catchment bores (see section 2.5) failed to achieve reproducible identification of the global optima.

In light of these challenges, multistart local calibration was abandoned in favor of a robust real-value global calibration scheme. In selecting a scheme from the many global calibration schemes available, the following criteria were adopted: (1) demonstrated robustness for 5–10 dimensional real-value problems with highly variable response surface shape (i.e., the global optima is highly likely to be identified when the scheme is repeatably ran); (2) applicability to problems with bounded and unbounded parameter ranges (because numerous time-series model parameters have a lower bound but no physical upper bound); and (3) sufficient simplicity for a user to run without detailed knowledge of the scheme. Clearly there are a number of schemes that would meet these criteria but a review, or testing, of each is beyond the scope of this paper. The selected scheme, and one scheme meeting criteria 2 and 3 that has been reviewed, tested and found to be very robust [*Langdon and Poli*, 2007; *Das and Suganthan*, 2011], is the *Covariance Matrix Adaptation Evolutionary Strategy* (CMA-ES) [*Hansen*, 2006; *Hansen et al*., 2003]. In an evaluation of ten stochastic global calibration schemes for hydrological modeling, CMA-ES was shown to perform as well or better than the other nine methods with respect to both reliability of the solution and convergence speed [*Arsenault et al*., 2014]. Interestingly, the widely applied Shuffled Complex Evolution (SCE) scheme [*Duan et al*., 1992] performed best only when the model complexity was low and there were few parameters calibrated. Additionally, like many schemes, SCE requires the user to input numerous settings for reliable and reproducible identification of the global optima and, hence, requires the users to have an understanding of the calibration scheme. In contrast, an appealing aspect of CMA-ES is that the user only defines the number of reruns of the scheme and all other settings are adaptively controlled by the scheme. For these reasons CMA-ES was adopted for calibration of the TFN models.

Critically important to the performance of CMA-ES is that the parameters are encoded so they have similar sensitivity. To achieve this, selected parameters were $\log_{10}$ transformed (see sections 2.2.1, 2.2.2, and 2.3 for details). This resulted in a likely range of -10 to 10 for all parameters. Lastly, CMA-ES requires the input of a standard deviation for the initial sampling distributions for each parameter. This was set very widely to half the physical parameter range if bounded, or half the plausible range if unbounded.

2.5. Assessment of the Calibration Scheme

Considering the importance of the calibration scheme to successful groundwater time series modeling, the robustness of the CMA-ES scheme was tested for both the five parameter linear TFN model and a nine parameter nonlinear TFN model. The latter used groundwater potential ET (equation (13)) and the normalized soil free drainage (equation (11)) with two calibrated soil parameters; specifically, the soil storage parameter, S_{cap}, and the power term for free drainage β. The ET power parameter, γ, was fixed to one and the infiltration parameter, α, was fixed to zero and k_{sat} was fixed to $-\infty$ (note, the latter omitted free-drainage from the soil model. See section 2.2.2 for a discussion of using free drainage in the time series despite it not being included in the soil model). In undertaking the testing, the aim was to assess the reproducibility of the calibration results and to identify the number of CMA-ES reruns required to achieve reproducible results. The testing was undertaken by repeating the calibration ten times for four bores within the control catchment and assessing the distribution of objective function and parameter values from the ten calibration runs. The trials were repeated for 0, 2, 4, 6 and 8 CMA-ES reruns. The underlying assumption of this testing was that there is a single global optima.

2.6. Model Implementation

In testing the linear and nonlinear TFN models, all structural forms of each model were applied to eleven observation bores within the study area. Specifically, two linear TFN models and 84 nonlinear TFN models were applied to each bore (see the supporting information for details the 86 models). For the linear TFN models, the models differed only in the impulse response function, θ, for ET. For the nonlinear TFN models, the 84 models structures arose from 16 to 17 different soil moisture models applied to each combination of soil fluxes input to the TFN model (equations (14–17)). An additional 16 subforms were derived for the free drainage nonlinear TFN models by removing the free-drainage term from the soil model (equation (6)) but retaining the free-drainage power parameter, β, within the transfer functions.

In applying each model, a split-sample approach [*Klemes*, 1986] was adopted whereby the calibration was undertaken on the first 10 years of water level observations and evaluation used the remaining 3 years. For computational efficiency, those bores with data loggers were resampled to monthly frequency. All simulations used SILO [*Jeffrey et al*., 2001] daily precipitation and ETo data (at 142.90° longitude and $-37.15°$ latitude) from 1890 to 2008 scaled by the monthly site precipitation. In calibrating each model, the CMA-ES was rerun four times (see section 3.1). In assessing the nonlinear TFN models against the linear models, the best performing nonlinear TFN model at each bore was defined as that having the maximum evaluation period unbiased coefficient of efficiency; that is, the standard coefficient of efficiency (CoE) [*Nash and Sutcliffe*, 1970] extended so that the evaluation period bias is subtracted from the simulation.

3. Results

3.1. Reproducibility of the Model Optimization

To assess the reproducibility of the CMA-ES calibration scheme, it was applied ten times to four bores using the five parameter linear TFN model and a nine parameter nonlinear TFN model. To summarize the assessment, Figure 5 shows the minimum, maximum and median CoE and the required number of model evaluations against the number of CMA-ES reruns.

For the linear TFN model, Figures 5a and 5b show that after four CMA-ES reruns the ten trials for three bores each converged to a single optima after between 27,425 and 119,329 model evaluations (0.56–2.44 h respectively, or 0.4429 s per model evaluation per core, using a 6-Core AMD Phenom II 3.2 Ghz desktop). However for bore 124705, the maximum difference in the CoE across the 10 trials was 0.015 for 4 reruns and it declined to 0.008 for 10 reruns (results for 10 reruns not presented, they required a total of between 1,720,853 and 2,223,605 model evaluations). Thus the calibration of bore 124705 did not reproducibly converge to a single optima. Similarly, for the nonlinear TFN model, Figures 5c and 5d show that after four

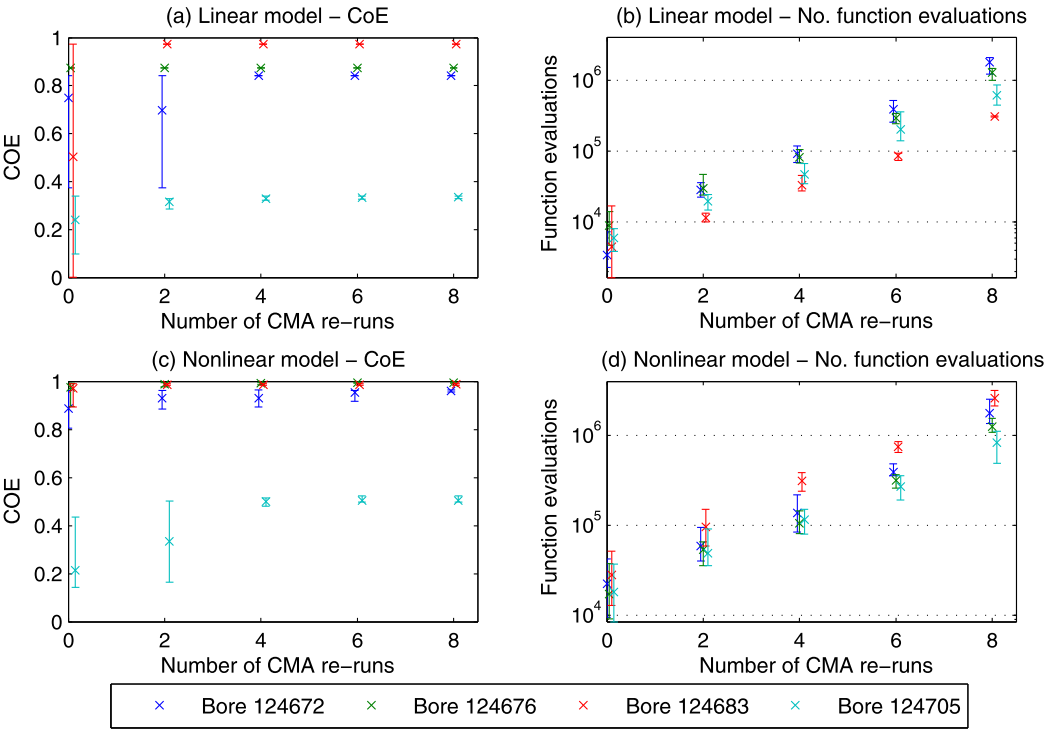

Figure 5. Summary of (a, c) CMA-ES calibration solution reproducibility and (b, d) the required number of model evaluations. A linear and nonlinear TFN model for four bores was calibrated ten times using 0, 2, 4, 6, and 8 CMA-ES reruns. Note, the whiskers denote the minimum and maximum and the cross the mean.

CMA-ES reruns the ten trials for all four bores failed to converge to a single optima with the maximum difference in CoE over each of the ten trials ranging from 0.005 to 0.07. This required 80,249 to 384,739 model evaluations. However, after 8 reruns requiring between 488,917 and 3,189,327 model evaluations (6–8 times greater computational load), the maximum difference in CoE only reduced to 0.003–0.026. The failure to repeatedly converge to a single optima highlights the probabilistic nature of this, and most, global calibration schemes; a finding supported by *Arsenault et al.* [2014] where, in calibrating the 10 parameter MOHYSE rainfall-runoff model to 11 basins, 40 CMA-ES trials were undertaken and any one trial was deemed to have converged if it was within 95% of the best optima. Overall, the increased complexity of the nonlinear TFN model does reduce the calibration reproducibility and, given that a 6–8 times increase in model evaluations only slightly improved reproducibility, an impractical number of evaluations appear required to achieve 100% calibration reproducibility. In light of Figure 5, four CMA-ES reruns appears to be sufficient to achieve a near-global optimal solution and this was adopted for the following model runs.

3.2. Performance of Optimal Models

Figures 6 and 7 show the observed and modeled groundwater hydrographs at the control and treated catchments respectively. Each plot presents simulation results from the five parameter linear TFN model (i.e., without a unique impulse response function for ET), the seven parameter linear TFN model (i.e., with a unique impulse response function for ET) and the best performing nonlinear TFN model, as defined by having the maximum evaluation period unbiased CoE. For the treated catchment, all simulations included the land use change component and hence they have an additional model parameter.

For the upper slopes of the control catchment, Figures 6c and 6d show that the nonlinear TFN model achieved exceptionally good reproduction of the large recharge events in the late 1990s and the subsequent decadal decline. However, at the upper-most bore, Figure 6e shows that the nonlinear TFN model simulation of the 1990s recharge event to have deteriorated relative to the nonlinear TFN simulation of lower catchment bores. During the evaluation period, the prediction was exceptionally good at bore 124676, while at bores 124676 and 124672 the prediction was biased but the timing of the 2007 recharge event was well simulated (Figures 6d and 6e). Neither of the two linear TFN models performed well on the

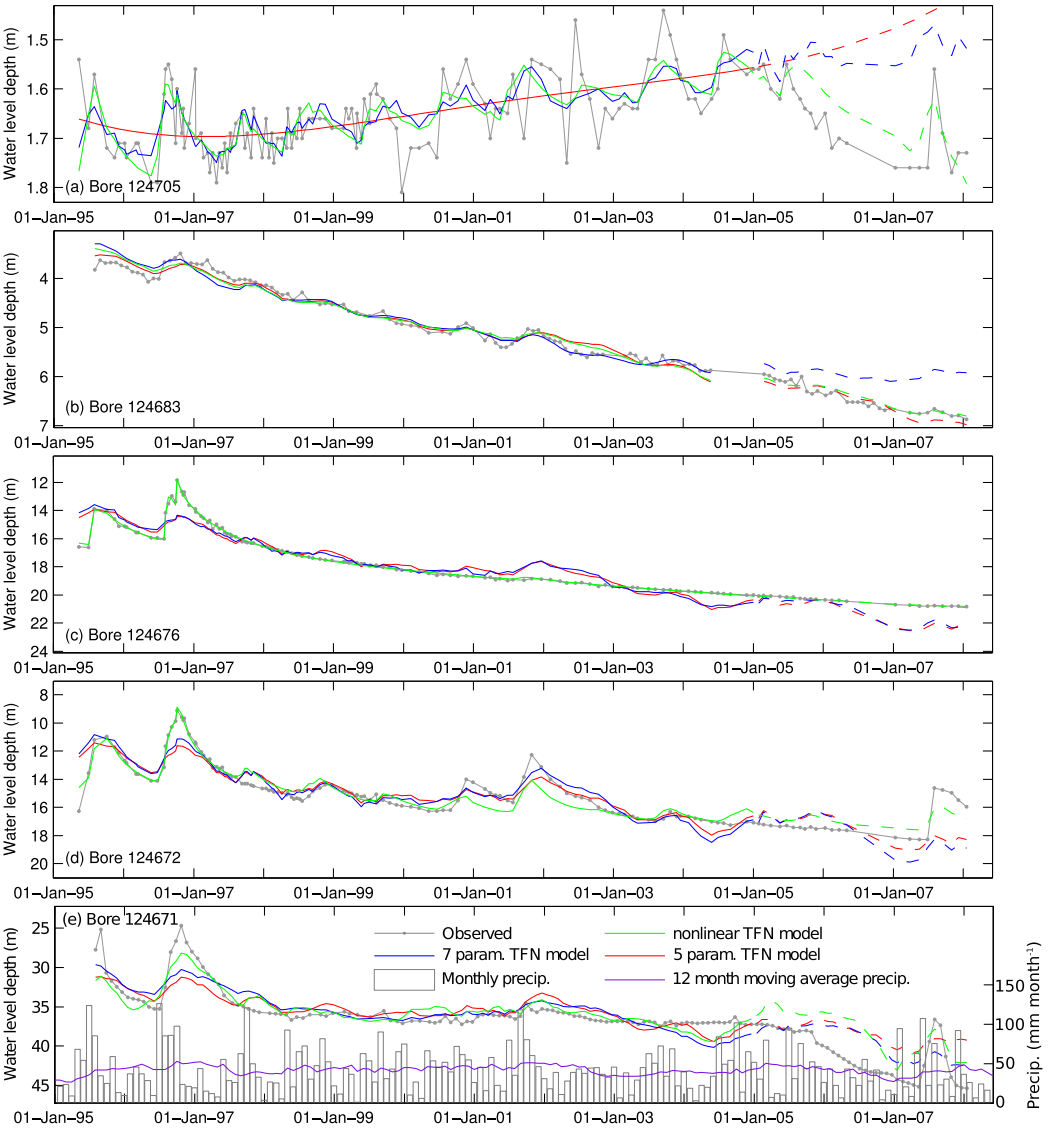

Figure 6. Control catchment observed and modeled hydrographs for bores (a) 124705, (b) 124683, (c) 124676, (d) 124672, and (e) 124671. The solid lines for the modeled estimates denote the calibration period and the dashed line the evaluation period.

upper slopes. Both failed to simulate the large recharge events in the 1990s and, during the subsequent decadal decline, both incorrectly simulated groundwater dynamics comparable to those simulated during the large recharge events in the 1990s. With regard to the lower slopes of the control catchment, Figures 6a and 6b show less distinction between the models (excluding the five parameter model in Figure 6a). For bore 124683, during the calibration period, all three models were comparable. However, during the evaluation period, the seven parameter linear TFN model had a considerable bias. Finally, for bore 124705 the nonlinear TFN model and the seven parameter linear TFN model did simulate the annual dynamics and the gentle rise from 1995 to 2002. In contrast, the five parameter linear TFN model did not simulate the annual dynamics but did simulate the long-term trend during the calibration period. During the evaluation period, only the nonlinear TFN model simulated the gentle decline from 2004 onward and the 2007 recharge event.

For the upper slopes of the treated catchment, Figures 7c–7f show the nonlinear TFN model to again have achieved exceptionally good reproduction of the large recharge events in the late 1990s and the subsequent decadal decline. The performance during the evaluation period was also exceptional; however, the evaluation period was not a difficult test for the models as it contained minimal variance from a near linear decline. With regard to the linear TFN models, they again failed to simulate the large recharge events and,

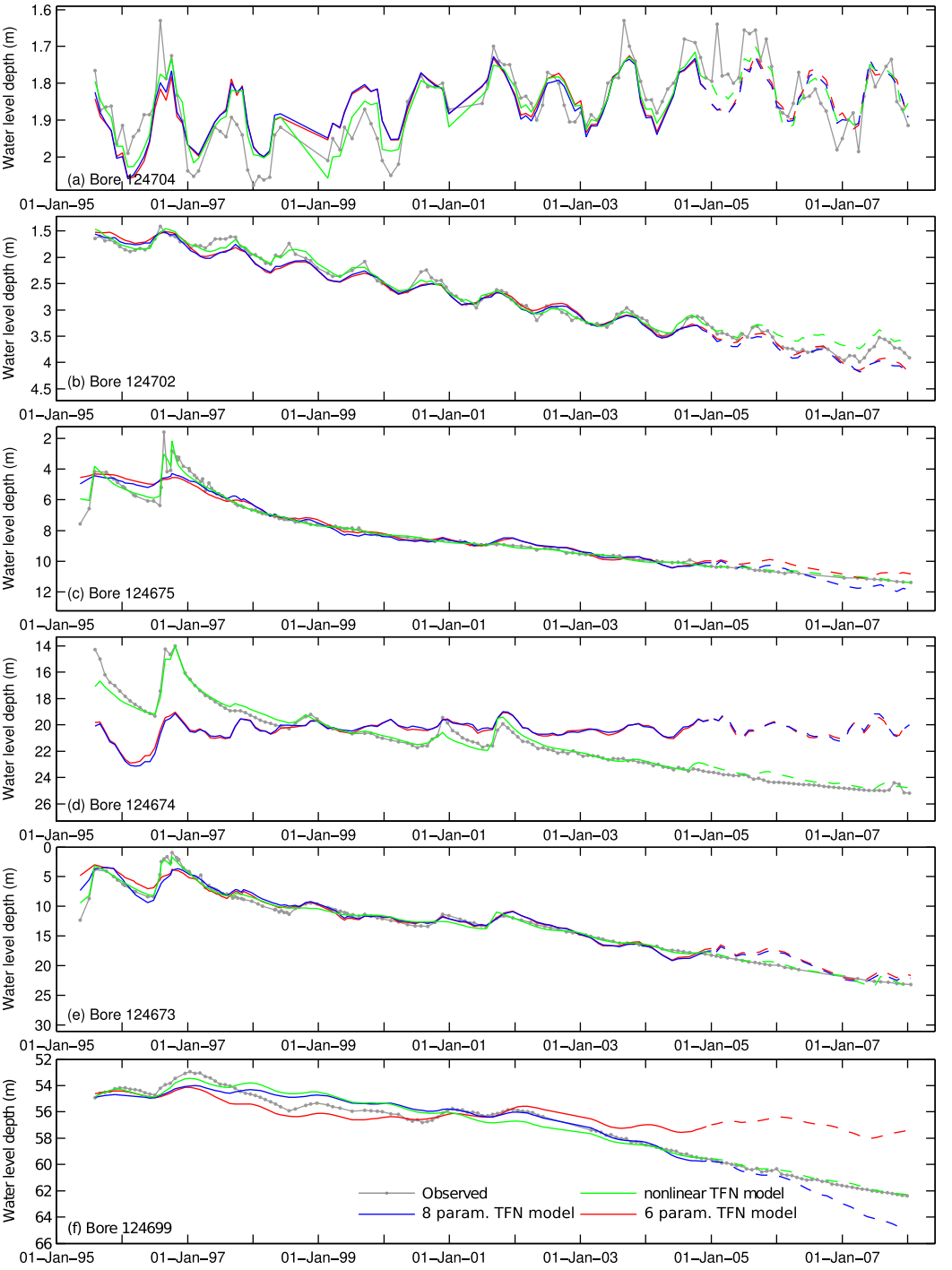

Figure 7. Treated catchment observed and modeled hydrographs with the inclusion of land-use change for bores (a) 124704, (b) 124702, (c) 124675, (d) 124674, (e) 124673, and (f) 124699. The solid lines for the modeled estimates denote the calibration period and the dashed line the evaluation period.

for bore 124674, it also failed to simulate the subsequent long-term decline. For the lower slopes of the treated catchment, Figures 7a and 7b showed less distinction between the models. All three models adequately simulated the calibration and evaluation period dynamics with minimal bias.

To objectively assess the model performance, and to compare forms of the nonlinear TFN soil moisture models, Table 2 summarizes the calibration and evaluation CoE, unbiased CoE and the bias from the best nonlinear TFN model and the five and seven parameter linear TFN models (six and eight parameters

Table 2. Model Performance Statistics During the Calibration and Evaluation Periods for Selected Models[a]

	Control Catchment					Treated Catchment					Treated Catchment
Bore ID	124705	124683	124676	124672	124671	124704	124702	124675	124674	124673	124699
Nonlinear TFN											
Model Number	23	56	31	24	17	29	63	31	63	57	27
Input form[b]	InfSeL	InfCwNl	DrNl	InfSeL	InfCwNl	InfCwL	DrNl	DrNl	DrNl	DrNl	DrNl
Output form[c]	GeL	GeL	GeL	GeL	GeL	GeL	GeNl	GeL	GeNl	GeL	GeL
No. parameters	10	8	10	7	9	10	11	10	11	11	9
CoE-Cal.	0.513	0.967	0.997	0.906	0.784	0.772	0.986	0.979	0.955	0.981	0.888
UnbiasCoEEval.	0.559	0.882	0.981	0.467	0.670	0.605	0.782	0.982	0.791	0.931	0.986
Bias-Eval(m)	−0.031	−0.037	0.026	−0.243	−3.001	0.010	−0.158	−0.066	−0.316	−0.116	−0.155
Linear TFN—Seven Parameters (Eight for Treated Catchment)											
CoE-Cal.	0.476	0.962	0.873	0.841	0.662	0.622	0.962	0.917	−0.222	0.928	0.911
UnbiasCoEEval.	−0.158	0.272	−3.991	−1.618	0.457	0.414	0.590	−0.112	−1.403	0.766	0.022
Bias-Eval(m)	−0.143	−0.500	0.710	0.572	−1.737	0.019	0.175	0.083	−4.205	−1.145	1.008
Linear TFN—Five Parameters (Six for Treated Catchment)											
CoE-Cal.	0.338	0.972	0.875	0.823	0.626	0.606	0.956	0.900	−0.150	0.908	0.725
UnbiasCoEEval.	−1.370	0.795	−3.791	−1.083	0.436	0.382	0.558	0.648	−1.196	0.760	0.586
Bias-Evaluation (m)	−0.171	0.038	0.797	0.344	−2.774	0.015	0.125	−0.447	−4.166	−1.493	−4.018

[a]Note: the model number refers to the model structures detailed in the supporting information. Also, for the nonlinear models the input and output form refers to the soil flux adopted to replace precipitation and PET respectively.

[b]DrNl: free drainage—nonlinear (i.e., calibrated β). InfSeL: infiltration via saturation excess—linear (i.e., fixed $\alpha = 0$). InfGwL: infiltration via catchment wetness—linear (i.e., fixed $\alpha = 1$). InfGwNl: infiltration via catchment wetness—nonlinear (i.e., calibrated α).

[c]GeL: groundwater evaporative potential—linear (e.g., fixed $\gamma = 1$). GeNl: groundwater evaporative potential—nonlinear (e.g., calibrated γ).

respectively for the treated catchment). The best performing nonlinear model was that having the maximum evaluation period unbiased CoE. During the calibration period, the nonlinear TFN model had the highest CoE at 9 of the 11 bores. During the evaluation period, the nonlinear TFN model had the highest unbiased CoE at all 11 bores and, at 9 bores, it had the lowest bias. Overall, the nonlinear TFN models performed consistently better than the linear TFN models during both the calibration and evaluation periods.

With regard to which type of nonlinear TFN model performed best, Table 2 is less conclusive. For the evaporative flux input to the nonlinear TFN model, the best performance was achieved at all bores by models using the groundwater evaporation potential and at two bores this was nonlinear. For the rainfall flux input to the nonlinear TFN model, the best predictive performance at treated catchment bores on the upper slopes was achieved using the nonlinear free drainage from the soil layer. However, for the control catchment the free drainage was the best option only at bore 124676. At the other four control catchment bores the infiltration rate was the best option. Considering that recharge is driven by soil drainage, and not infiltration, this result is physically questionable.

To further explore evidence for adopting the infiltration rate, Figure 8 compares the calibration and evaluation period CoE from the best infiltration model against the best free drainage model (the best model was selected by the maximum CoE during the evaluation period). It shows that, at bores where the infiltration model is the best model during either period, the drainage model performance is very comparable to that from the infiltration, with the only exception being bores 124671 and 124704 during the calibration period. Furthermore, if the best model was instead chosen by the maximum CoE during the calibration period, then the best nonlinear TFN model structure is more consistent with all but the shallowest bore in each catchment (bores 124704 and 124705) being best simulated using the free drainage rate while at the shallowest bore adoption of free drainage would reduce the CoE by only 0.03 and 0.008 respectively (results not presented). Overall, this suggests that the nonlinear TFN modeling should use the groundwater evaporation and the more physically realistic option of nonlinear free drainage.

With regard to the best soil model structure, there was considerable variability across the 11 bores with the number of parameters ranging from one to four. However some similarities did arise whereby six bores simulated infiltration as a function of soil moisture ($\alpha \neq 0$), ten bores simulated nonlinear free drainage ($\hat{\beta} > 1$) and nine simulated $\hat{k_{sat}} > 0$ (see section 2.2.2 for a discussion of simulating free drainage with $\hat{k_{sat}} = 0$) while nine bores simulated linear evaporation ($\gamma = 1$). So while optimal model performance is

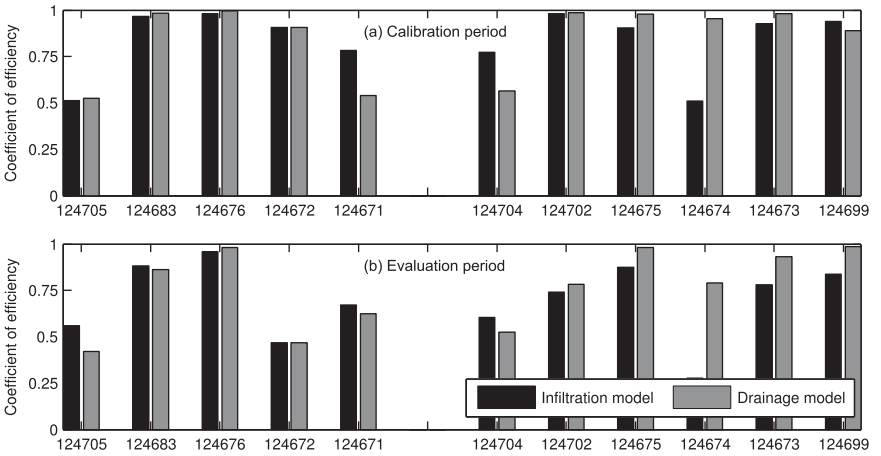

Figure 8. Comparison of best nonlinear TFN models using the soil infiltration rate versus the soil free drainage rate during the (a) calibration and (b) evaluation period.

dependent upon the soil model structure, the following soil models appear to be viable options (note, model numbers 31 and 63 respectively):

$$\frac{dS}{dt} = P\left(1 - \frac{S}{\hat{S_{cap}}}\right) - \hat{k_{sat}}\left(\frac{S}{\hat{S_{cap}}}\right)^{\hat{\beta}} - E_t\left(\frac{S}{\hat{S_{cap}}}\right) \tag{23}$$

$$\frac{dS}{dt} = P - \hat{k_{sat}}\left(\frac{S}{\hat{S_{cap}}}\right)^{\hat{\beta}} - E_t\left(\frac{S}{\hat{S_{cap}}}\right) \tag{24}$$

3.3. Estimation of Land-Use Change Impacts

To account for the revegetation within the treated catchment, and to quantify its impact on the groundwater level, each model included an integral simulating the groundwater uptake from the revegetation. For each model the integral was a rescaling of the ET integral weighted by the fraction of land use change. To summarize the estimated impact from the revegetation, Table 3 shows the total change in head on the final day of simulation for each treated catchment bore and the three types of models. Taking note that a negative value denotes a decline in head, Table 3 shows that the nonlinear TFN model estimates a decline in head on the midslopes (bores 124702, 124675 and 124673) and an increase in head at the lower and upper slopes (bores 124704, 124674 and 124699). This inconsistent impact of the revegetation also arose using the six and eight parameter linear TFN models. However, all models estimated a comparable decline at bores 124702 and 124675 and a comparable rise at bore 124704.

To further explore the validity of including the revegetation in the models, Table 4 shows the change in the calibration and evaluation CoE when the revegetation integral was added to each model (note, a positive value denotes an improved performance with the revegetation integral). It shows the nonlinear TFN to have a consistent improvement in evaluation period unbiased CoE when revegetation was added. For the calibration period, the changes were very modest and notably improved only at bore 124702. For the eight parameter linear TFN bores, the inclusion of revegetation improved evaluation period performance at four bores, while for the six parameter linear TFN bores the evaluation period performance improved only at two bores.

In evaluating the model estimates of the impacts from revegetation, an independent estimate is required. However, no clear impact is observable when comparing the hydrographs from the treated catchment (Figures 7) to those at similar elevation within the control catchment (Figures 6). All but the lowest elevation bore within both catchments show a long-term declining head and no obvious difference in response to recharge events. Considering that from the late 1990s to 2010 southeast Australia experienced the driest conditions since 1900 [*van Dijk et al.*, 2013], and that the observed dynamics within both catchments is typical of drought conditions, it is not surprising that the revegetation appears to have had minimal impact on groundwater levels. However to quantitatively investigate, Table 3 also presents the difference in the

Table 3. Groundwater Elevation Impact (meters) From Revegetation Within the Treated Catchment in the Final Year of Simulation[a]

Bore ID	124704	124702	124675	124674	124673	124699
Nonlinear TFN	0.213	−1.709	−1.008	0.1460	−6.524	1.538
Linear TFN—eight parameters	0.229	−2.280	−3.007	2.0955	1.9062	0.2703
Linear TFN—six parameters	0.287	−2.034	−4.254	2.1414	0.0	−1.825
Estimated observed impact		0.766	−0.290	−1.840		

[a]Note: a negative impact indicates that the water table became deeper.

observed long-term trend at treated catchment bores compared to the control catchment bores having a similar elevation multiplied by the duration to the end of the record. This independent estimate shows an inconsistent impact from the revegetation where, compared with the control catchment bores, the water level is estimated to have risen at bore 124702 and declined at bores 124675 and 124674. Furthermore, the total decline in head from the revegetation appears to be small. Hence, it is not surprising that adding revegetation to the models produced modest performance improvements and small and inconsistent estimates of the impacts from revegetation.

3.4. Temporal Decomposition of Hydrographs

A strength of the linear TFN models is that the transfer functions provide insight into the response times of the aquifer. For example, in Figure 3a the skewed Gaussian-like curves represent a lagged response to climate while Figure 3b shows exponential-like curves with no lag. However, for the nonlinear TFN models the climate data is processed through the soil moisture model before input to the transfer function. Hence, the transfer functions cannot provide direct insight to the climate lags and response times. To address this limitation, the soil model and the climate transfer functions were used to decompose the simulated hydrograph into the contributions from climate periods of 0–1, 1–2, 2–5, 5–10, and 10–20 years ago. To illustrate, Figure 9 shows the decomposition for treated catchment bores 124704 and 124674 (note, in interpreting these decomposition plots, the magnitude of the fluctuations and trends is most informative and not the mean value over time). For bore 124704, Figure 9a shows that the climate 0–1 years ago drove most of the water level fluctuation while the climate from 1 to 2 years ago caused fluctuations of approximately 0.05 m.

Regarding bore 124674, Figure 9b shows that climate 0–1 years ago drove the recharge events while climate 2–5 and 5–10 years ago drove longer term trends; with the latter periods appearing to explain the observed long-term water level decline shown in Figure 7d. Considering that at the study area the Australian Millennium Drought [*van Dijk et al*., 2013] started in 1994 [see *Peterson*, 2009, Figure 7.5] and that the downward trend in Figure 7d is only apparent 4 years later in 1998, a 2–10 year lag is plausible. This finding, however, must be viewed in the context of the problematic estimation of the revegetation impacts.

4. Discussion and Conclusions

This paper presented nonlinear groundwater hydrograph transfer function noise (TFN) models that extend *von Asmuth and Bierkens* [2005] to account for the nonlinear processes of runoff and recharge and long-term groundwater level decline. To achieve efficient calibration, and hence a useful model, the impulse response functions from *von Asmuth and Bierkens* [2005] was rederived to minimize the parameter covariance and the CMA-ES global calibration scheme was adopted. The model nonlinearity arose from the transformation of daily precipitation and PET using a highly flexible soil moisture model. The precipitation was transformed to free drainage or to the surface infiltration rate while the PET was transformed to the

Table 4. Change in the CoE as a Result of Including a Land Use Change Component Within a Treated Catchment Time-Series Model[a]

Bore ID		124704	124702	124675	124674	124673	124699
Nonlinear TFN	ΔCoE calibration	0.006	0.772	−0.009	−0.011	0.0170	−0.030
	ΔCoE evaluation	0.232	0.300	0.603	0.609	0.228	0.072
Linear TFN 8 param.	ΔCoE calibration	0.159	0.792	0.003	−1.08	0.660	0.001
	ΔCoE evaluation	−0.092	0.135	3.944	1.150	0.509	−0.818
Linear TFN 6 param.	ΔCoE calibration	0.149	0.905	0.027	−0.256	8.257	0.370
	ΔCoE calibration	−0.002	0.246	1.692	−0.233	4E-08	−0.001

[a]Note: a negative value indicates that the model performance declined when the land use change component was added to the model.

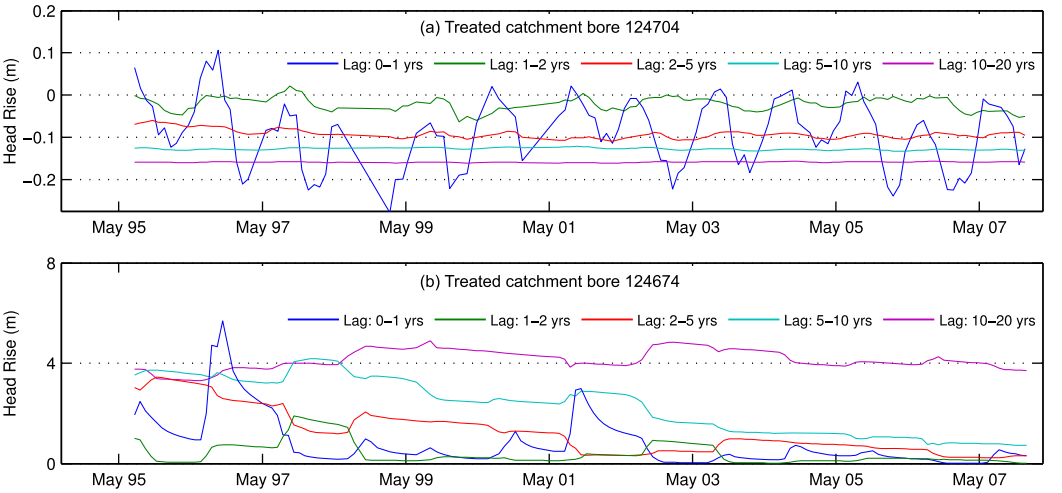

Figure 9. Temporal decomposition of bores (a) 124704 and (b) 124674 within the treated catchment.

groundwater evaporation potential; that is, the evaporative potential after soil ET was subtracted from the PET. For each combination of these fluxes, at least 16 soil moisture model structures were derived which resulted in 84 different nonlinear TFN models.

To evaluate the linear and nonlinear models, two linear models and all 84 nonlinear models were applied to the 11 piezometers within the Great Western paired catchment in Victoria, Australia. However, before exploring

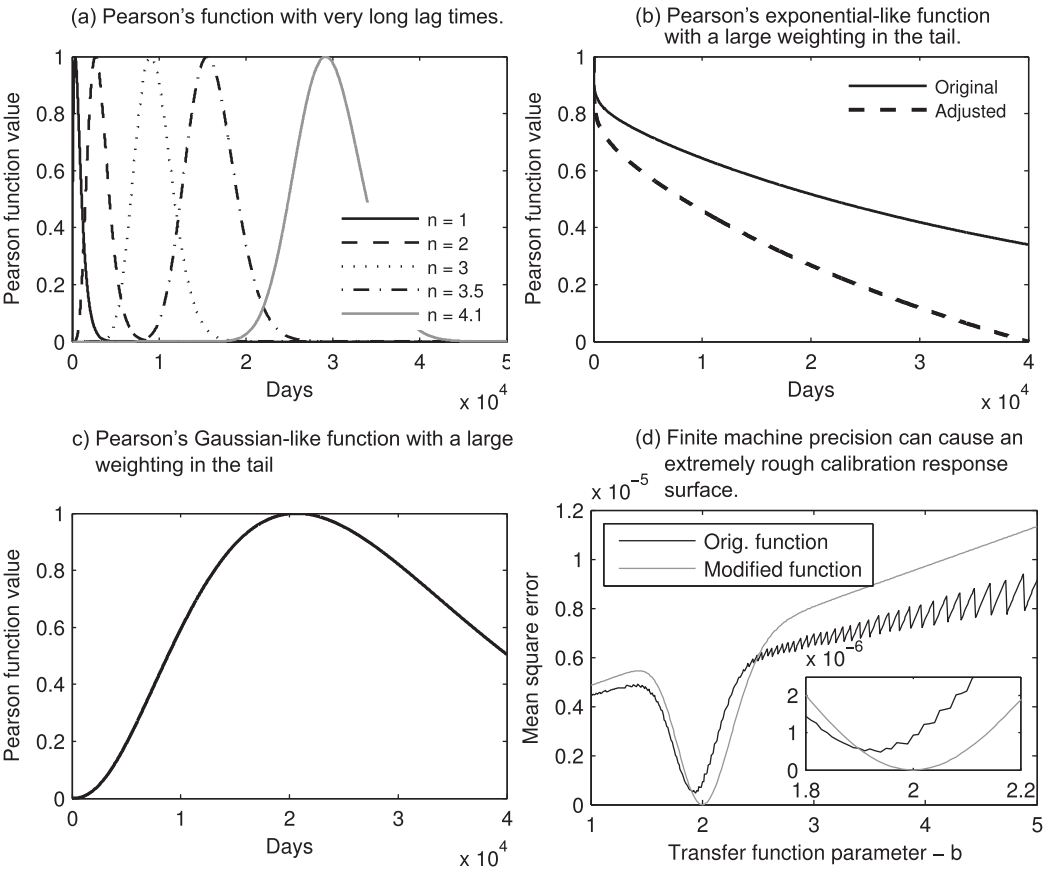

Figure 10. Four weaknesses of Pearson's function and some solutions. (a) Pearson's a function with very long lag times. (b) Pearson's *exponential-like* function with a large weighting in the tail. (c) Pearson's *Gaussian-like* function with a large weighting in the tail. (d) Rounding errors from finite machine precision can cause an extremely rough calibration response surface.

the performance of individual models, the reproducibility of the CMA-ES calibration scheme to identify the global optima was assessed. This was undertaken by repeating the calibration 10 times at four bores within the control catchment and assessing the variability of the estimated global optima. For a five parameter linear TFN model, the calibration was 100% reproducible at three of the four bores. For a nine parameter nonlinear TFN model, the calibration was reproducible at each of the four bores but only to within two to three decimal places in the coefficient of efficiency (CoE). This outcome, while not ideal, is consistent with *Arsenault et al*. [2014] and highlights the challenge of global calibration and the probabilistic nature of this, and most, global calibration schemes. That is, for all but the most trivial applications, one cannot be sure that the global optima has been identified. It also highlights that the interaction of the model structure with the observation and forcing data often produces multiple local optima that challenge even a robust global calibration scheme [*Langdon and Poli*, 2007] and that render the calibration of the TFN models potentially suboptimal.

The lower reproducibility for the more complex models may have caused some of the 84 nonlinear models applied to each bore to be calibrated to a local optima. Despite these calibration challenges, the nonlinear models performed significantly better than both linear TFN models during both calibration and evaluation periods. Only the nonlinear TFN model was able to simulate major recharge events and the observed long-term decline. Unexpectedly, however, the best performing nonlinear TFN model at five of eleven bores used the soil infiltration rate in place of the precipitation rather than the more physically plausible soil drainage rate. Closer examination of the 84 nonlinear models showed that adopting the drainage rate at these five bores only slightly reduced model performance compared with using the infiltration rate. Combined with the variability in the best soil model structure (see Table 2) this suggests that optimal performance can be sensitivity to the soil model structure and the fluxes taken from it. Nevertheless, reasonable performance can likely be achieved using the suggested overall best performing model structure (see section 3.2). Furthermore, the best nonlinear TFN model for a bore is dependent upon how the best model is selected. In this paper, the best nonlinear TFN model was selected by the evaluation period CoE. Often this identified a robust model that performed well during the calibration and evaluation periods. However, occasionally the calibration period CoE was notably lower than if the best model had been selected using the calibration period CoE (e.g., for bore 124671 selection by the evaluation period CoE gave a calibration period CoE of 0.67 while selection by the calibration period CoE gave a calibration period CoE of 0.96). Combined with the variable model structure, this highlights that when building nonlinear TFN models carefully consideration should be given to the purpose of the application (e.g., interpolation or extrapolation) and the importance of identifying the optimal model structure.

In using the calibrated linear and nonlinear TFN models, for selected bores, the simulated hydrographs were decomposed into the contribution from each stressor and, for the nonlinear models, to periods over time. For the decomposition to the stressors, both the linear and nonlinear TFN model estimates of the impacts of revegetation were unreliable, although the nonlinear model results were slightly more consistent. This was probably due to both the Millennium drought in southeast Australia reducing the impacts from the revegetation and by the simplistic representation of the land use change as a binary change at the planting date. Potentially, a more reliable estimate of the impacts from revegetation could be achieved by using a time series of vegetation data, such as the *Normalized Difference Vegetation Index*. The temporal decomposition provided an interesting output that potentially explains the climatic time-lag driving long-term water level trends without resorting to simple fixed trend estimates derived either from the hydrograph or use of a simple regression model such as HARTT [*Ferdowsian et al*., 2001]. However, the outputs do require future evaluation against independent measures.

Overall, the proposed optimal nonlinear TFN model, while performing best in this application, should be adopted with some caution. It was derived for only one climate type, geology and land use and the predictive period at most bores had a declining trend of low variance. The latter does not present the most challenging test to the predictive performance but, unlike stream flow, an aquifer effectively has no minimum water level (i.e., if the head is far above the basement) so a long-term decline does give opportunity for the prediction to be very biased. Despite this, it is unlikely that the identified optimal model structure would be appropriate for all climates and geologies. To aid the exploration of different nonlinear TFN structures the source code is open-source (see section 5) and, because of its object oriented structure, is highly flexible. New TFN model structures can easily be built and tested with different climate transformations and transfer functions or entirely new types of models can be built and linked into the overall modeling framework. In

summary, the nonlinear transformation of climate forcing is essential for simulating large infrequent recharge events and long-term head declines. We do not expect the model structures presented herein to be optimal everywhere and invite others to use the groundwater time-series modeling framework to develop and test new models.

Appendix A: Revised Pearson's Type III Function

Five weaknesses were identified with the Pearson type III function adopted for the impulse response function. In the following, each weakness is detailed and modifications to the Pearson type III function are presented.

The first modification was to minimize the parameter covariance. To begin with, the term b^n within equation (3) was removed. This simply caused the scaling effect of b^n to be shifted to parameter A. Similarly, the term $\Gamma(n)$ also produces a scalar result and was removed:

$$\theta(t)=At^{n-1}\exp(-bt) \tag{A1}$$

Equation (A1) does, however, still have a covariance between parameters A, b and n when n is greater than one; that is when it has a Gaussian-like shape. To address this, when n is greater than one, the terms $t^{n-1}\exp(-bt)$ were divided by the peak value of the function. By simple calculus the peak was determined as follows:

$$f=t^{n-1}\exp(-bt) \tag{A2}$$

$$\frac{\partial f}{\partial t}=(n-1)t^{n-2}\exp(-bt)-bt^{n-1}\exp(-bt) \tag{A3}$$

To determine the timing of the peak, equation (A3) was equated to zero and rearranged to an explicit equation for t:

$$0=(n-1)t^{n-2}\exp(-bt)-bt^{n-1}\exp(-bt) \tag{A4}$$

$$(n-1)t^{n-2}\exp(-bt)=bt^{n-1}\exp(-bt) \tag{A5}$$

$$(n-1)t^{n-2}=bt^{n-1} \tag{A6}$$

$$\frac{t^{n-2}}{t^{n-1}}=\frac{b}{n-1} \tag{A7}$$

$$t^{n-2-n+1}=\frac{b}{n-1} \tag{A8}$$

$$t=\frac{n-1}{b} \tag{A9}$$

Next, the value of f at the peak was determined by substituting equation (A9) into equation (A2):

$$f_{peak}=\left(\frac{n-1}{b}\right)^{n-1}\exp(1-n) \tag{A10}$$

Finally, dividing equation (A1) by equation (A10) gave the following impulse response function for when parameter n is greater than one:

$$\theta(t)=A\frac{t^{n-1}\exp(-bt)}{\left(\frac{n-1}{b}\right)^{n-1}\exp(1-n)} \tag{A11}$$

The second modification addressed parameter transformations. In trialling the linear (and nonlinear) TFN models, the parameter n (see equation (A11)) postcalibration ranged over five orders of magnitude from less than 0.001 one to over three hundred (similar range occurred for parameters A and b). To illustrate, Figure 10a shows the Pearson's Type III function curves for n ranging from 1 to 1000. It shows that larger values of n shift the peak back in time; hence simulating a larger time-lag between a climate event and a head response. Considering that global calibration methods often sample the parameter space from within user-set upper and

lower bounds, such a large parameter range for n would reduce the sampling density; which would cause either a lower reliability in locating the global optima or require an excessive number of calibration iterations. To minimize these calibration problems, parameters A, b and n within equation (A11) were antilog transformed. This effectively back-transformed the parameters from a *compressed* parameter space during the calibration to a value of the same form as within equation (A11). Hence, equation (A11) was modified to:

$$\theta(t)=\hat{A}\frac{t^{\hat{n}-1}\exp\left(-\hat{b}t\right)}{\left(\frac{\hat{n}-1}{\hat{b}}\right)^{\hat{n}-1}\exp\left(1-\hat{n}\right)} \tag{A12}$$

where $\hat{A}$ is the antilog transform of A and $\hat{A}=10^{A}$; $\hat{b}$ is the antilog transform of b and $\hat{b}=10^{b}$; and $\hat{n}$ is the antilog transform of n and $\hat{n}=10^{n}$.

The third modification was undertaken for when $\hat{n}\leq 1$ in order to reduce the impulse response function value at the start of the climate record. To illustrate, Figure 10b shows that, when the function has an exponential-like form, very low values of parameter b can result in a large value thousands of days prior. When such an impulse response function is multiplied by a very long daily precipitation record and integrated, precipitation from many years past will be estimated to produce a significant head displacement from that of the drainage elevation (that is, the d constant within equation (4)). This problem did occur and caused d, which is derived postcalibration, to have an unrealistically low elevation. To fix this problem, the parameter A was first omitted and the function value at 100 years prior to the first climate observation was calculated. This was subtracted from the function and, to minimize parameter covariance, was then used to normalize the impulse response function. The final impulse response function for $\hat{n}\leq 1$ was:

$$\theta(t)=\hat{A}\frac{t^{\hat{n}-1}\exp\left(-\hat{b}t\right)-f_{limit}}{1-f_{limit}} \tag{A13}$$

where:

$$f_{limit}=t_{limit}^{\hat{n}-1}\exp\left(-\hat{b}t_{limit}\right) \tag{A14}$$

$$t_{limit}=\min\left(t\right)-100\times 365 \tag{A15}$$

Additionally, Figure 10c shows that the Gaussian-like form of the function can also have a very large value at 10,000 days in the past. A challenge in fixing this form of the problem is that the impulse response function does not monotonically decrease. No solution could be derived for this problem and it remains a weakness of the impulse response function based on the Pearson's type III function.

The fourth modification, and one related to the third, was the integration of the function from the first climate observation to negative infinity. This was undertaken because, without the tail integration, spurious parameter estimates resulted from the calibration. To undertake the modification, the function was integrated from the first climate observation, t_0 [T], to negative infinity and multiplied by the mean daily climate forcing rate:

$$\int_{\tau=-\infty}^{t_0} P_\tau\theta(t-\tau)d\tau\approx\bar{P}\int_{\tau=-\infty}^{t_0}\theta(t-\tau)\partial\tau \tag{A16}$$

For the Gaussian-like function in equation (A12), an analytical solution to the integral was derived by firstly moving the constants to outside the integral:

$$\begin{aligned}&\int_{\tau=-\infty}^{t_0}\hat{A}\frac{(t-\tau)^{\hat{n}-1}\exp\left(-\hat{b}(t-\tau)\right)}{\left(\frac{\hat{n}-1}{\hat{b}}\right)^{\hat{n}-1}\exp\left(\hat{n}-1\right)}\partial\tau\\&=\frac{\hat{A}}{\left(\frac{\hat{n}-1}{\hat{b}}\right)^{\hat{n}-1}\exp\left(1-\hat{n}\right)}\int_{\tau=-\infty}^{t_0}(t-\tau)^{\hat{n}-1}\exp\left(-\hat{b}(t-\tau)\right)\partial\tau\end{aligned} \tag{A17}$$

Note that the integral is of a similar form to the Gamma function:

$$\Gamma(a)=\int_{s=0}^{\infty} s^{a-1}e^{-s}ds \tag{A18}$$

To rearrange the integral into this form, u was substituted for $\hat{b}(t-\tau)$:

$$\int_{\tau=-\infty}^{t_0}(t-\tau)^{\hat{n}-1}\exp\left(-\hat{b}(t-\tau)\right)\partial\tau=\int_{u=-\infty}^{0}\frac{1}{\hat{b}^{\hat{n}-1}}u^{\hat{n}-1}\exp(-u)\frac{\partial u}{\hat{b}} \tag{A19}$$

$$=\frac{1}{\hat{b}^{\hat{n}}}\int_{u=-\infty}^{0}u^{\hat{n}-1}\exp(-u)\partial u \tag{A20}$$

$$=-\frac{\Gamma_i(\hat{n},u)}{\hat{b}^{\hat{n}}} \tag{A21}$$

where $\Gamma_i(\eta,u)$ is the incomplete Gamma function integrated from zero to u. Integrating from $-\infty$ to t_0, replacing $\hat{n}$ for η and reinserting the constants from equation (A17) gives the following equation for the integration of equation (A12) from the first climate observation to negative infinity:

$$\bar{P}\int_{\tau=-\infty}^{t_0}\theta(t-\tau)\partial\tau=\frac{\bar{P}\hat{A}\Gamma(\hat{n})}{\hat{b}^{\hat{n}}\left(\frac{\hat{n}-1}{\hat{b}}\right)^{\hat{n}-1}\exp(1-\hat{n})}\left[\Gamma_i\left(\hat{n},\hat{b}t_0\right)-1\right] \tag{A22}$$

For the exponential-like function in equation (A13), a similar approach was adopted. However, to integrate from t_{limit} and not from $-\infty$, the third modification of subtracting the function value at the start of the climate record was incorporated and the first modification of scaling by the peak was omitted. The integration of equation (A13) from the first climate observation to t_{limit} was:

$$\bar{P}\int_{\tau=t_{limit}}^{t_0}\theta(t-\tau)\partial\tau=\frac{\bar{P}\hat{A}}{1-f_{limit}}\left(\frac{\Gamma(\hat{n})}{\hat{b}^{\hat{n}}}\left[\Gamma_i\left(\hat{n},\hat{b}t_0\right)-1\right]-f_{limit}(t_{limit}-t_0)\right) \tag{A23}$$

In implementing the fourth modification, the mean daily forcing rate became a function of the calibration period. That is, if a time-series model was calibrated from, say, years 1990 to 2000 and the input climate date extended from 1950 to 2000, then the mean daily precipitation and PET was calculated for 1950 to 2000 and all subsequent simulations used these mean rates.

The fifth modification was undertaken to minimize numerical rounding in the convolution of the transfer functions with daily climate data. Within equation (A12), the term $t^{\hat{n}-1}$ can become very large when $n \gg 1$ and a very long daily climate record is input.

When numerically approximating the TFN integrals they can, and did, exceed machine precision and rounded to ∞. To address this problem the order of operations was modified when $\hat{n} \gg 1$. Initially, the time points rounding to ∞ were removed from the convolution sum. However, as shown within Figure 10d, this produced an extremely rough response surface with hundreds of local optima. To overcome this problem, the order of operation within equation (A12) was modified by shifting the power term $\hat{n}-1$ applied to t (see equation (A12)) to the outer of the equation:

$$\theta(t)=\hat{A}\left[\frac{t\exp\left(\frac{-\hat{b}t}{\hat{n}-1}\right)}{\left(\frac{\hat{n}-1}{\hat{b}}\right)^{\frac{\hat{n}-1}{\hat{n}-1}}\exp\left(\frac{1-\hat{n}}{\hat{n}-1}\right)}\right]^{\hat{n}-1} \tag{A24}$$

$$=\hat{A}\left[\frac{t\exp\left(\frac{-\hat{b}t}{\hat{n}-1}\right)}{\left(\frac{\hat{n}-1}{\hat{b}}\right)\exp(-1)}\right]^{\hat{n}-1} \tag{A25}$$

$$= \hat{A}\left[t\exp\left(\frac{-\hat{b}t}{\hat{n}-1}\right)\frac{\hat{b}}{\hat{n}-1}\exp(1)\right]^{\hat{n}-1} \quad \text{(A26)}$$

Importantly, in equation (A26) was only adopted if $\hat{n}$ was sufficiently large that the convolution with the daily forcing rate rounded to ∞. This was because as $\hat{n}$ approaches 1 and $\hat{b}$ is large (e.g., 0.01), the term $\exp\left(\frac{-\hat{b}t}{\hat{n}-1}\right)$ within equation (A26) exceeds machine precision and rounds to zero, and hence produces a discontinuity in the response surface.

Lastly, the integration of the tail to $-\infty$ (equation (A22)) was similarly modified when rounding errors occurred:

$$\bar{P}\int_{\tau=-\infty}^{t_0}\theta(t-\tau)\partial\tau = \frac{\bar{P}\hat{A}\Gamma(\hat{n})}{\hat{b}^{\hat{n}}\left(\frac{\hat{n}-1}{\hat{b}}\exp(-1)\right)^{\hat{n}-1}}\left[\Gamma_i\left(\hat{n},\hat{b}t_0\right)-1\right] \quad \text{(A27)}$$

Acknowledgments

The authors are grateful for the financial support provided by the Australian Research Council (grants LP0991280 and LP130100958), the Department of Environment and Primary Industries, Victoria, Australia; and the Bureau of Meteorology, Australia. The authors also thank Vahid Shapoori and Lionel Siriwardena of the University of Melbourne for the their testing of the models. Lastly, the source code and example input data are available as supporting information and the latest version of the code is available at https://github.com/peterson-tim-j/Groundwater-Statistics-Toolbox.

References

Arsenault, R., A. Poulin, P. Côté, and F. Brissette (2014), A comparison of stochastic optimization algorithms in hydrological model calibration, *J. Hydrol. Eng.*, *19*, 1374–1384, doi: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 Resour. Res.*, *44*, W04420, doi:10.1029/2007WR006239.

Box, G. E., G. M. Jenkins, and G. C. Reinsel (2008), *Time Series Analysis: Forecasting and Control*, 4th ed., 746 pp., John Wiley, N. J.

CSIRO (2009), Groundwater yields in south-west Western Australia: A report to the Australian Government from the CSIRO south-west Western Australia sustainable yields project, technical report number 2, CSIRO Water for a Healthy Country Flagship, Australia.

Das, S., and P. Suganthan (2011), Differential evolution: A survey of the state-of-the-art, *IEEE Trans. Evol. Comput.*, *15*(1), 4–31, doi:10.1109/TEVC.2010.2059031.

Duan, Q. Y., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031, doi:10.1029/91WR02985.

Fabbri, P., C. Gaetan, and P. Zangheri (2011), Transfer function-noise modelling of an aquifer system in NE Italy, *Hydrol. Processes*, *25*(2), 194–206, doi:10.1002/hyp.7832.

Fan, J., and J. Pan (2006), Convergence properties of a self-adaptive Levenberg-Marquardt algorithm under local error bound condition, *Comput. Optim. Appl.*, *34*(1), 47–62, doi:10.1007/s10589-005-3074-z.

Ferdowsian, R., and D. Pannell (2009), Explaining long-term trends in groundwater hydrographs, in *18th IMACS World Congress and MODSIM09 International Congress on Modelling and Simulation*, edited by R. Anderssen, R. Braddock, and L. Newham, pp. 3109–3115, Modell. and Simul. Soc. of Aust. and N. Z. and Int. Assoc. for Math. and Comput. in Simul., Cairns, Queensland, Australia.

Ferdowsian, R., D. J. Pannell, C. McCarron, A. Ryder, and L. Crossing (2001), Explaining groundwater hydrographs: Separating atypical rainfall events from time trends, *Aust. J. Soil Res.*, *39*(4), 861–875, doi:10.1071/SR00037.

Ferdowsian, R., A. Ryder, R. George, G. Bee, and R. Smart (2002), Groundwater level reductions under lucerne depend on the landform and groundwater flow systems (local or intermediate), *Aust. J. Soil Res.*, *40*(3), 381–396, doi:10.1071/SR01014.

Guswa, A. J., M. A. Celia, and I. Rodriguez-Iturbe (2002), Models of soil moisture dynamics in ecohydrology: A comparative study, *Water Resour. Res.*, *38*(9), 1166, doi:10.1029/2001WR000826.

Hansen, N. (2006), The CMA evolution strategy: A comparing review, in *Towards a new evolutionary computation. Advances on estimation of distribution algorithms, Studies in Fuzziness and Soft Computing*, vol. 192, edited by J. Lozano et al., pp. 75–102, Springer, Netherlands.

Hansen, N., S. Muller, and P. Koumoutsakos (2003), Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES), *Evol. Comput.*, *11*(1), 1–18, doi:10.1162/106365603321828970.

Harbaugh, A. W., E. R. Banta, M. C. Hill, and M. G. McDonald (2000), MODFLOW-2000, the U.S. Geological Survey modular ground-water model—User guide to modularization concepts and the ground-water flow process, *Tech. Rep. 00–92*, U.S. Geol. Surv., Reston, Va.

HydroTechnology (1995), Salinity control option trial at Great Western: Drilling of observation bores, *Tech. Rep. MC/44070.050/1*, Melbourne, Australia.

Jeffrey, S. J., J. O. Carter, K. B. Moodie, and A. R. Beswick (2001), Using spatial interpolation to construct a comprehensive archive of Australian climate data, *Environ. Modell. Software*, *16*(4), 309–330, doi:10.1016/S1364-8152(01)00008-1.

Kavetski, D., and M. Clark (2010), Ancient numerical daemons of conceptual hydrological modeling: 2. impact of time stepping schemes on model analysis and prediction, *Water Resour. Res.*, *46*, W10511, doi:10.1029/2009WR008896.

Kavetski, D., G. Kuczera, and S. W. Franks (2006), Calibration of conceptual hydrological models revisited: 1. overcoming numerical artefacts, *J. Hydrol.*, *320*(1-2), 173–186, doi:10.1016/j.jhydrol.2005.07.012.

Klemes, V. (1986), Operational testing of hydrological simulation models, *Hydrol. Sci. J.*, *31*(1), 13–24.

Langdon, W. B., and R. Poli (2007), Evolving problems to learn about particle swarm optimizers and other search algorithms, *IEEE Trans. Evol. Comput.*, *11*(5), 561–578, doi:10.1109/TEVC.2006.886448.

Lehsten, D., J. von Asmuth, and M. Kleyer (2011), Simulation of water level fluctuations in kettle holes using a time series model, *Wetlands*, *31*(3), 511–520, doi:10.1007/s13157-011-0174-7.

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

Manzione, R., M. Knotters, G. Heuvelink, J. Asmuth, and G. Camara (2010), Transfer function-noise modeling and spatial interpolation to evaluate the risk of extreme (shallow) water-table levels in the Brazilian Cerrados, *Hydrogeol. J.*, *18*(8), 1927–1937, doi:10.1007/s10040-010-0654-5.

Marquardt, D. W. (1963), An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.*, *11*(2), 431–441.

McIntyre, J. (2003), Implementation of salinity control options: Salinity control option trial Great Western 1996–2003, technical report, Dep. of Primary Ind., Horsham, Victoria, Australia.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I—A discussion of principles, *J. Hydrol.*, *10*(3), 282–290, doi:10.1016/0022-1694(70)90255-6.

Obergfell, C., M. Bakker, W. Zaadnoordijk, and K. Maas (2013), Deriving hydrogeological parameters through time series analysis of groundwater head fluctuations around well fields, *Hydrogeol. J.*, *21*(5), 987–999, doi:10.1007/s10040-013-0973-4.

Peel, M. C., B. L. Finlayson, and T. A. McMahon (2007), Updated world map of the Köppen-Geiger climate classification, *Hydrol. Earth Syst. Sci.*, *11*(5), 1633–1644, doi:10.5194/hess-11-1633-2007.

Peterson, T. J. (2009), Multiple hydrological steady states and resilience, PhD thesis, Dep. of Civ. and Environ. Eng., Univ. of Melbourne, Parkville, Australia. [Available at http://repository.unimelb.edu.au/10187/8540.]

Peterson, T. J., and A. W. Western (2011), Time-series modelling of groundwater head and its de-composition to historic climate periods, in *34th IAHR World Congress—Balance and Uncertainty*, pp. 1677–1684, Eng. Aust., Brisbane, Australia.

Shapoori, V., T. J. Peterson, A. W. Western, and J. Costelloe (2011), Quantifying the impact of pumping on groundwater heads using observation data and advanced time series analysis, in *MODSIM 2011 International Congress on Modelling and Simulation*, edited by F. Chan and D. Marinova, pp. 3966–3972, Modell. and Simul. Soc. of Aust. and N. Z. [Available at http://www.mssanz.org.au/modsim2011/index.htm.]

Sinclair Knight Merz (1996), Salinity control option trial at Great Western: Site selection and recharge mapping, *Tech. Rep. WC00152*, Melbourne, Australia.

Sinclair Knight Merz (1997), Great Western salinity control option trial: Comparison of groundwater and surface water regimes in the trial sub-catchments, *Tech. Rep. WC00436:R12AKHGW*, Melbourne, Australia.

Sinclair Knight Merz (2005), Five year technical review of the Neuarpur WSPA groundwater management plan, *Tech. Rep. WC02865*, Melbourne, Australia.

Sinclair Knight Merz (2006a), National Action Plan for Salinity and Water—Benchmarking regional water table and trends, *Tech. Rep. WC02661*, Melbourne, Australia.

Sinclair Knight Merz (2006b), Great Western salinity control option trial: Update of monitoring data: 2002–2006, *Tech. Rep. WC00436*, Melbourne, Australia.

Tu, W., and R. Mayne (2002), Studies of multi-start clustering for global optimization, *Int. J. Numer. Methods Eng.*, *53*(9), 2239–2252, doi:10.1002/nme.400.

van Dijk, A., H. Beck, R. Crosbie, R. de Jeu, Y. Liu, G. Podger, B. Timbal, and N. Viney (2013), The millennium drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society, *Water Resour. Res.*, *49*, 1040–1057, doi:10.1002/wrcr.20123.

von Asmuth, J., K. Maas, M. Knotters, M. Bierkens, M. Bakker, T. Olsthoorn, D. Cirkel, I. Leunk, F. Schaars, and D. von Asmuth (2012), Software for hydrogeologic time series analysis, interfacing data with physical insight, *Environ. Modell. Software*, *38*, 178–190, doi:10.1016/j.envsoft.2012.06.003.

von Asmuth, J. R., and M. F. P. Bierkens (2005), Modeling irregularly spaced residual series as a continuous stochastic process, *Water Resour. Res.*, *41*, W12404, doi: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 Resour. Res.*, *38*(12), 1287, doi:10.1029/2001WR001136.

von Asmuth, J. R., K. Maas, M. Bakker, and J. Petersen (2008), Modeling time series of ground water head fluctuations subjected to multiple stresses, *Groundwater*, *46*(1), 30–40, doi:10.1111/j.1745-6584.2007.00382.x.

Yi, M., and K. Lee (2004), Transfer function-noise modelling of irregularly observed groundwater heads using precipitation data, *J. Hydrol.*, *288*(3-4), 272–287, doi:10.1016/j.jhydrol.2003.10.020.

Yihdego, Y., and J. Webb (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*(4), 877–887, doi:10.1007/s10040-011-0726-1.

Zhou, Y. (2009), A critical review of groundwater budget myth, safe yield and sustainability, *J. Hydrol.*, *370*(1-4), 207–213, doi:10.1016/j.jhydrol.2009.03.009.