

A state-wide assessment of optimal groundwater hydrograph time series models

L. Siriwardena^a, T.J. Peterson^a, A.W. Western^a

^a Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia.
Email: timjp@unimelb.edu.au

Abstract: A number of groundwater hydrograph time series models have been proposed over recent years but, to our knowledge, there has been no systematic review of their performance and thus no means of selecting a model to suit the prevailing conditions. This paper presents an evaluation of the new groundwater hydrograph time series models presented in Peterson and Western (2011) against existing models on 620 bore hydrographs distributed throughout Victoria. Bores that monitor water level under natural conditions and having at least 20 years of data were used. The aim of this study is to rigorously demonstrate the strength of the Peterson and Western (2011) models (hence referred to as soil moisture store – transfer function noise model, or SMS-TFN) and ascertain which forms of the various soil moisture components within the model perform best and under what conditions. To assess the relative performance, the widely used HARTT model (Ferdowsian et al. 2001, Ferdowsian et al. 2002) and the standard transfer function noise model (von Asmuth et al., 2002) were also investigated.

This investigation into the groundwater head time series modelling was assessed by evaluating the performance of eleven model variants of three classes of models (SMS-TFN, TFN, HARTT) using the Coefficient of Efficiency (CoE) and the Akaike Information Criterion (AIC) as the performance measures for the calibration and evaluation periods. The results showed that the SMS-TFN model (Peterson and Western, 2011), significantly improves the predictive model performance compared to the performance of the traditional TFN model. The SMS-TFN model with ground water recharge as the forcing component shows better model calibration and predictive performances than models with infiltration as the forcing component. These model variants produced the best median calibration period CoE of 0.655 (where 1.0 is a perfect fit) and the best evaluation period unbiased CoE of 0.270 (see Figure 1).

The predictive performance of the HARTT model was shown to be highly variable and inconsistent across the bore hydrographs tested. If a sustainable linear time trend exists in the bore hydrograph, the model produced good results as indicated by performance measures during both calibration and validation periods. However, in the absence of such trends, the model performed poorly. This illustrates the potential risk in assuming a non-climatic time trend which may or may not exist in the bore hydrographs. More importantly, the SMS-TFN model with ground water recharge as the forcing component was shown to be the most robust model which can explain most of the bore hydrographs from climate data alone.

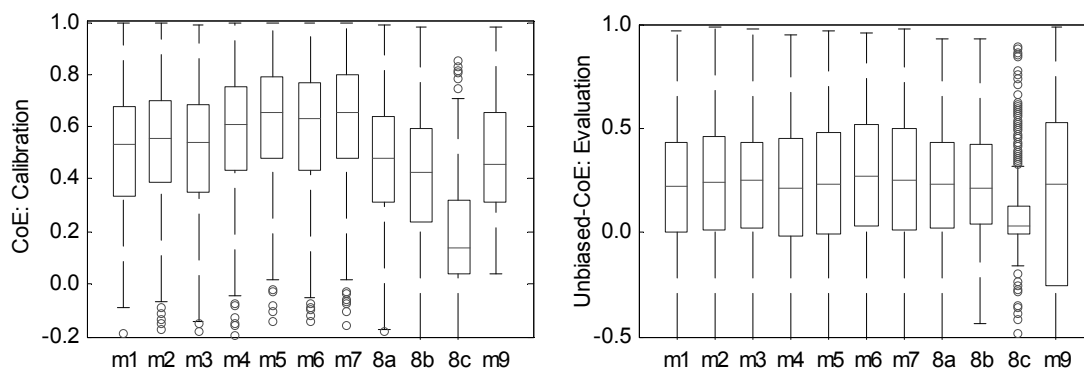


Figure 1. Box plots of CoE for calibration period and unbiased CoE for evaluation period

Keywords: Time series modelling, Groundwater hydrograph, Climate forcing, HARTT model

1. INTRODUCTION

Groundwater models are valuable tools for the catchment managers to take actions on the management of groundwater resources. A good model should be able to capture the essence of the catchment dynamics and the interaction of climatic processes on the groundwater behaviour and should be good in the predictive capability. Testing and improvements to existing groundwater hydrograph models and development of new models are of vital importance. The HARTT model (Ferdowsian *et al.* 2001, Ferdowsian *et al.* 2002) has been widely used in Australia for groundwater time-series modelling (CSIRO 2009, Sinclair Knight Merz 2005, Sinclair Knight Merz 2006). The model assumes an underline linear time trend in groundwater level over time and the effect of rainfall fluctuations is added to it. Rainfall is represented as an accumulation of deviations from the average rainfall. The original model has been subsequently improved to include the effect of evaporation (Sinclair Knight Merz 2006) and to model non-linear time trends in groundwater (Ferdowsian and Pannell 2009). Limitations of this model in groundwater hydrograph modelling are highlighted in Peterson and Western (2011). A more robust model, the transfer function noise (TFN) model (von Asmuth *et al.* 2002, von Asmuth and Bierkens 2005, von Asmuth *et al.* 2008) is an auto-regressive moving average externally forced time series model. The model can handle multiple stresses by assigning a separate transfer function to each and summing the effects of all stresses. It requires no assumption of first or second order climate stationarity, does not require uniformly spaced observations and explains the observed hydrograph without adoption of an arbitrary non-climatic temporal trend (as adopted by the HARTT model). Applications of this model to Victorian bore data are found in Yihdego and Webb (2011).

Peterson and Western (2011) presented a new, and statistically robust, hydrograph analysis method as an extension of the standard TFN model. In this method, parsimonious vertically lumped soil moisture model is added and a soil moisture state such as infiltration or groundwater recharge is convoluted through transfer functions instead of using climate variables directly. In addition, PET term of the TFN model was replaced by a soil moisture deficit and this was found to be required to capture long term declines. Their investigation into groundwater head time series modelling, using 12 observation bores in upper catchment in Wimmera River, identified that the inclusion of a parsimonious soil moisture store to the standard TFN model significantly increased the model predictive performance. Furthermore, Peterson and Western (2011) has been extended by Shapoori *et al.* (2011) to incorporate groundwater pumping. It has been tested using synthetic ModFlow models and the results are shown to be very encouraging.

This paper presents an evaluation of the new groundwater hydrograph time series models (Peterson and Western 2011) against existing models on 620 water table bore hydrographs distributed throughout Victoria. The aim of this study is to rigorously demonstrate the strength of the Peterson and Western (2011) models (hence referred to as soil moisture store – transfer function noise model, or SMS-TFN) and ascertain which forms of the various soil moisture components within the model perform best. To assess the relative performance, the widely used HARTT model and the standard TFN model (von Asmuth *et al.*, 2002) were also investigated.

2. METHODS

2.1. Model description

Three classes of groundwater time series models with increasing complexities were employed. The salient features of the 11 models associated with these three classes are outlined in Table 1. Models *m1* to *m7* refer to SMS-TFN model variants while *m8* and *m9* refer to TFN and HARTT models respectively. The simplest of the models, the HARTT model (Ferdowsian *et al.*, 2001), simulates the groundwater head as a linear combination of the cumulative monthly rainfall residual (CRR) and a time trend. The model used here is an improvement of the original model in which a monthly areal potential evapotranspiration term, scaled by the sum of two linear sigmoidal terms, is added (Sinclair Knight Merz, 2006). The five parameters of the model are determined by multiple linear regression.

In transfer function noise models (TFN), the contribution of a climate variable on the groundwater head is estimated as the convolution of the historic time series of the variable with the weights from a non-linear transfer function (von Asmuth and Bierkens, 2005; von Asmuth *et al.*, 2002; von Asmuth *et al.*, 2008). Multiple responsive climate variables and external stresses can be handled by assigning a separate transfer function to each variable or scaling a transfer function assigned to one variable for another variable where applicable. The time series of groundwater head is then obtained by summing the separate effects of all

forcing variables. In model *m8a*, separate Pearson type III distribution functions (PIII) are applied for daily precipitation (P_t) and areal potential evapotranspiration (E_t) while in model *m8b* the same TFN used for precipitation is scaled and applied to E_t . In model *m8c*, precipitation is considered solely responsible for the groundwater head. The parameters of the models are estimated by minimizing the variance of the innovations, not residuals, using a weighted least-squares objective function. An exponential noise model is incorporated to simulate the correlation of residuals from which the innovations are calculated.

Table 1. Salient features of groundwater time series models used in the study

Model Class	Model ID	Sat excess	Infiltr excess	ET scaling	+ forcing	- forcing	Number of Parameters			
							Soil	TFN	Noise	Reg
Soil moisture component + non-linear TFN (SMS-TFN)	<i>m1</i>	✓	✗	✗	Infiltr	deficit	1	6	1	-
	<i>m2</i>	✓	✗	✓	Infiltr	deficit	2	6	1	-
	<i>m3</i>	✓	✓	✗	Infiltr	deficit	2	6	1	-
	<i>m4</i>	✓	✓	✓	Infiltr	deficit	3	6	1	-
	<i>m5</i>	✓	✓	✗	Rechar	deficit	3	6	1	-
	<i>m6</i>	✓	✗	✗	Rechar	deficit	2	6	1	-
	<i>m7</i>	✓	✗	✓	Rechar	deficit	3	6	1	-
Non-linear TFN	<i>m8a</i>	-	-	-	Precip	E_t	-	6	1	-
	<i>m8b</i>	-	-	-	Precip	E_t^*	-	4	1	-
	<i>m8c</i>	-	-	-	Precip	E_t	-	3	1	-
HARTT	<i>m9</i>	Multiple linear regression (P_t , E_t and trend)				-	-	-	-	5

* scaled by precip transfer function; all other forcing using independent TFNs

Peterson and Western (2011) improved the traditional transfer function noise model by inclusion of a derived soil moisture state variable (SMS-TFN model) as a surrogate for direct climate forcing. They considered that the groundwater head response to precipitation is dependent upon the variables such as runoff, interception, unsaturated vertical conductivity and soil moisture. A parsimonious vertically lumped soil moisture model was added to estimate the soil moisture state variables. The SMS-TFN model variants listed in Table 1 can be divided into two broader categories. In models *m1* to *m4*, the precipitation in the traditional TFN model is replaced by an infiltration rate while in models *m5* to *m7* it is replaced by a free-drainage recharge estimate. The evaporation from the soil layer is modelled as E_t times the fraction of soil moisture. There are subtle differences in the soil moisture models used across different variants. In models *m3*, *m4* and *m5* precipitation is filtered by a maximum infiltration rate parameter as such the model mimics both infiltration excess and saturation excess and this infiltration parameter is absent in other models. In models *m2*, *m4* and *m7* an additional parameter is included to scale the evaporation from the soil layer. The model is incorporated with smoothing functions to constrain operators within upper and lower bounds without response surface discontinuities as a requirement for the calibration of the model using a gradient based solver. All model variants employ separate transfer functions of Pearson Type III for two forcing components. Factors that contribute to increase the water head are termed as positive forcing and vice versa for negative forcing (Table 1). An important adoption is the scaling of E_t by the soil moisture deficit for the E_t forcing contribution. This was conceptualized as the uptake of groundwater by vegetation during low soil moisture periods and was found to be essential for simulating long term groundwater level declines (Peterson and Western, 2011).

2.2. Groundwater Data

A comprehensive Victorian groundwater database has been established for this study comprising groundwater databases maintained by Victorian Department of Sustainability and Environment (DSE), Department of Primary Industries (DPI) and NSW Department of Water and Energy (DWE). The database contains 343,153 bores and 10.7 million observations. For evaluation of the groundwater hydrograph time series models 920 bore records distributed throughout Victoria were selected from this database. The selected bores are of three types: bores within sedimentary deposits that are monitoring the water table; bores monitoring confined aquifers; and shallow bores within bedrock. This paper is limited to the

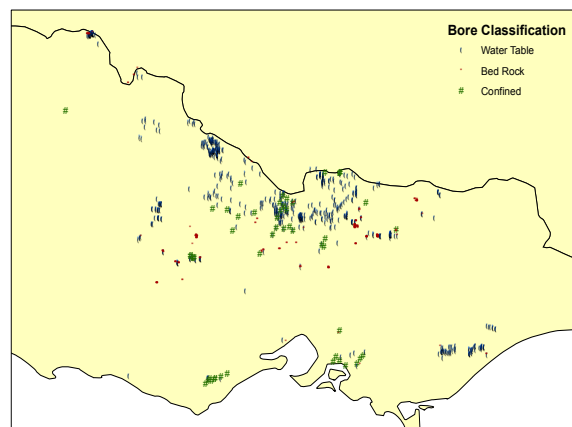


Figure 2. Map of bore locations

evaluation of 620 bore records that measure water table. Figure 2 shows the distribution of selected bores.

The extraction of suitable bore records from this database was subject to pre-processing of data to remove water level observations likely to be erroneous. Water level observations with any quality codes and all multiple observations for a single day were initially removed. The data records were then filtered to be approximately free of outliers using an automated procedure. The moving average was calculated using a two tailed exponential weighting with a window size of 365 days. The residuals were then derived (i.e. the difference between the moving average and the observations) and the standard deviation was calculated. If an observation's residual was further than three standard deviations from the mean it was removed. The moving average process was repeated three times per time series.

The modelling aims to investigate ground water level changes over time under natural conditions influenced only by climate and underlined geological conditions. The bores considered in this paper are representative of bores that monitor water level under natural conditions over long periods and having a good spatial coverage. The identification of bores that monitor the water table, and not the deeper aquifer layers, was based on the comparison of bore depth against the Victorian aquifer stratigraphy elevations (SKM and GHD 2009, Spatial Vision 2009) using an automated algorithm. The bores used for irrigated agriculture and for other intensive uses were not considered. Observation records were guaranteed for a minimum period of 20 years after excluding suspect observations. The data records were also checked for adequate continuity by applying an additional constraint in which at least for 15 years of the record should have a minimum of 6 data observations for each year. Nested bores within close proximity were also excluded if the bore depths are very similar. Bore depth of the 620 bores, which measure the water table, considered in this paper ranged from 1.5 to 100m with a median depth of 14.1m. The record length ranged from 20.0 to 58.2 years with a median record of 28.8 years while number of water level observation for each bore ranged from 149 to 1088 with a median of 284 observations.

2.3. Climate Data

Daily precipitation and potential evapotranspiration (PET) data were used as the input to the soil moisture model of the SMS-TFN model and as the climate forcing contribution through appropriate transfer response functions for other models. Precipitation and other meteorological data required to estimate PET were compiled from the Australian Water Availability Project (AWAP) database in which data for each climate element are provided at 0.05 degree resolution for entire Australia (Raupach *et al.* 2009). PET was estimated from the Morton's (1983) complementary relationship areal model using daily values of maximum and minimum temperature, vapour pressure and net solar radiation. As the daily net solar radiation was available from 1990 onwards the computation of PET values prior to 1990 was based on the mean monthly average net radiation values from 1960 to 1990. Precipitation and PET daily time series over the period 1950-2010 at each bore location was based on the computed values at the nearest grid cell to the bore location.

2.4. Model Implementation

Each of the eleven models was calibrated using all observation data minus the data in the last 10 years irrespective of the length of the bore record. The calibrated models were then evaluated for predictive model performance on the last ten years data. The calibration was undertaken using a multi-start Levenberg-Marquardt Newtonian solver (Hill & Tiedeman, 2007; Levenberg, 1944; Marquardt, 1963). Numerous extensions were implemented to aid efficient computation, most notably the Broyden updating of Jacobian matrix (Broyden, 1965; Davidon, 1991) and parallel computation of multiple starts. Each was run with 8 initial starts, with each start having a maximum of 1500 iterations and the function and parameter convergence criteria set to 1E-8. The models were run on the University of Melbourne *Alfred* Linux cluster using its 40 64bit Opteron nodes (www.hpc.unimelb.edu.au/alfred).

2.5. Model assessment

The performance of the models was assessed by the Coefficient of Efficiency, CoE (Nash and Sutcliffe, 1970) and the Akaike Information Criterion, AIC (Akaike, 1974) for the calibration and evaluation periods. For the evaluation period an unbiased COE was used in which the bias corrected residuals were used in the computation. It masks the effect of model bias on the CoE measure and allows a more effective comparison of model performance. The bias defined as the mean of the residuals, i.e. the observed minus the modelled head, was assessed independently. The AIC is a measure of the relative goodness of fit of a model with an increasing penalty on the number of parameters in the model, which facilitates a comparison of performance of models of differing number of parameters. A lower value indicates a parsimonious model with a good fit.

3. RESULTS

Figure 3 shows box plots of CoE between observed and estimated water head for the calibration period and unbiased CoE for the evaluation period for all 11 models tested. Similarly, Figure 4 shows box plots of AIC for the calibration and evaluation periods. Figure 5 shows box plots of bias in the estimated groundwater head. Table 2 presents the median values of each performance measure shown in figures 3 to 5.

Figure 3 shows that all variants of the SMS-TFN perform considerably better than the TFN model variants and the HARTT model with respect to the CoE measure over the calibration period. It is also evident that the performance of the recharge forced SMS-TFN model variants (*m5*, *m6*, *m7*) is better than that of the infiltration forced SMS-TFN model variants. The models *m5* and *m7* have the highest calibration CoE with median values of 0.655 and 0.654 respectively. These two variants have one additional soil layer parameter than in *m6*. In addition to soil moisture capacity and recession parameters, *m5* has an infiltration parameter and *m7* has a scaled ET parameter. The performance of the standard TFN model with both precipitation and evapotranspiration forcing components included is much greater than the model with only the precipitation term. The performance of the TFN models with the evapotranspiration term (*m8a* and *m8b*) is generally comparable with the performance of the HARTT model. The calibration AIC measure (Figure 3) also shows very similar performance across all the models tested. The models *m5* and *m7* have again shown the best performance with the lowest AIC values although the distinction among the models is not high as with the CoE values. It should be noted that these models performed well regardless of the penalty of the AIC criterion for having more parameters in the model.

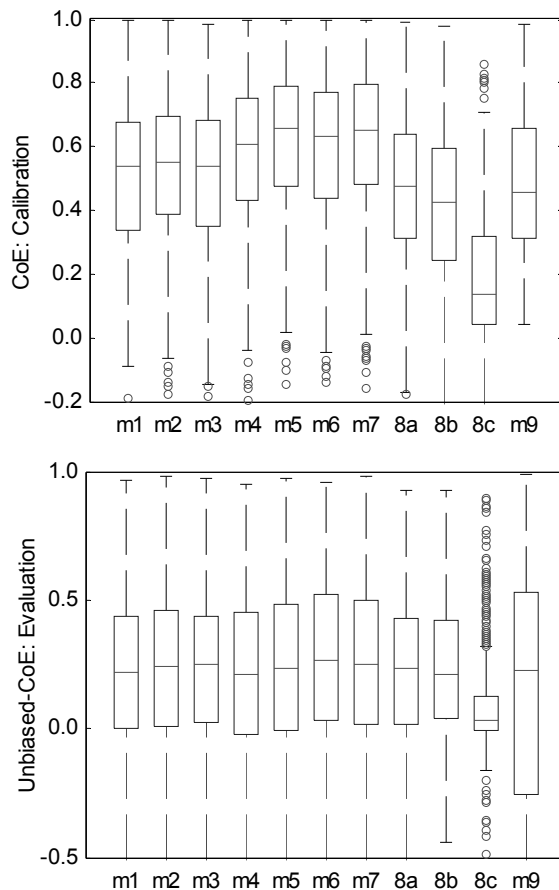


Figure 3. Box plots of CoE for calibration period and unbiased CoE for evaluation period

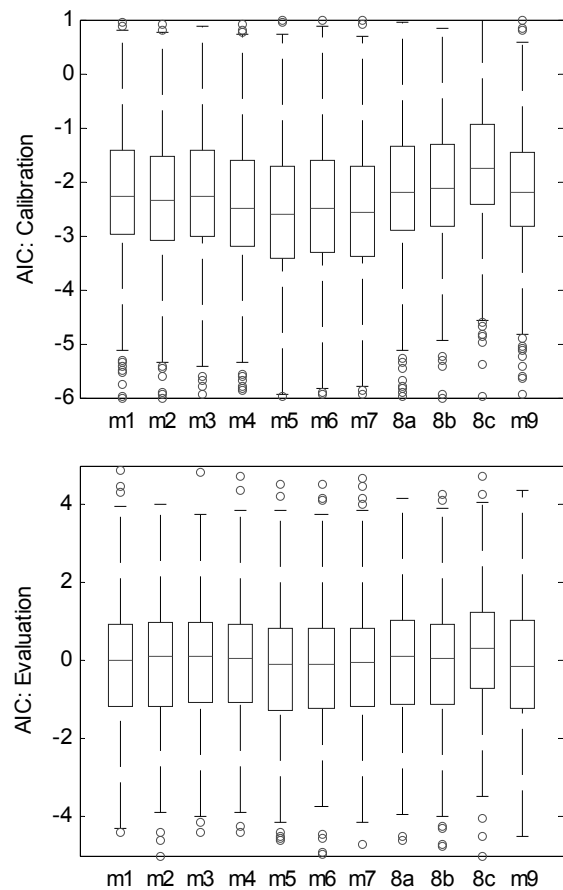


Figure 4. Box plots of AIC for calibration and evaluation periods

For the evaluation period, the performance measures are comparatively less distinctive across the models and significantly lower compared to the calibration period performance. The model *m6* produced the highest median unbiased CoE (0.270) but the performance is only marginally better than that of other SMS-TFN models. With regard to the HARTT model, it produced a comparable median unbiased CoE (0.227) but its inter quartile range was very large. This shows that the HARTT model lacks the ability to produce consistent

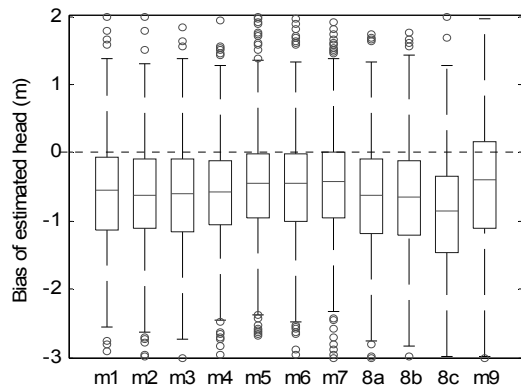


Figure 5. Box plots of the evaluation period bias of the estimated head

results across the whole range of bore records. All the models show some bias in the modelled groundwater head (Figure 5), i.e. the modelled head tends to be higher than the observed head. The least median biases shown in the recharge forced SMS-TFN models and the HARTT model are in the range of 0.4m. The AIC measure over the evaluation period (Figure 4) shows somewhat better performance from models *m5*, *m6*, *m7* and *m9*.

The HARTT model has shown good results when the observed data generally supports a linear time trend but performs poorly when the data do not exhibit such trends. On the other hand, the recharge based SMS-TFN model variants have shown to be the most robust models which can be applied over a wider range of bores in different climatic and geographical regions.

The bore hydrographs from these models could be explained from climate data alone without any time trend assumptions.

Table 2. Median values of performance measures

Statistic	SMS-TFN (Infiltration)				SMS-TFN(Recharge)			TFN			HARTT
	<i>m1</i>	<i>m2</i>	<i>m3</i>	<i>m4</i>	<i>m5</i>	<i>m6</i>	<i>m7</i>	<i>m8a</i>	<i>m8b</i>	<i>m8c</i>	<i>m9</i>
CoE (cal)	0.535	0.554	0.537	0.607	0.655	0.632	0.654	0.476	0.427	0.140	0.453
CoEU (val)	0.217	0.239	0.248	0.214	0.233	0.270	0.254	0.231	0.212	0.033	0.227
Bias (m)	-0.56	-0.62	-0.59	-0.57	-0.43	-0.44	-0.43	-0.62	-0.65	-0.85	-0.42
AIC (cal)	-2.28	-2.34	-2.28	-2.50	-2.60	-2.51	-2.59	-2.19	-2.12	-1.75	-2.19
AIC (val)	-0.01	0.10	0.09	0.07	-0.09	-0.12	-0.06	0.08	0.04	0.33	-0.17

4. DISCUSSION AND CONCLUSIONS

This investigation into the groundwater head time series modelling showed that the SMS-TFN model (Peterson and Western, 2011), with an added parsimonious soil moisture model to the TFN model, significantly improves the predictive model performance compared to the performance of the traditional TFN model. The SMS-TFN model with ground water recharge as the forcing component shows better model calibration and predictive performances than models with infiltration as the forcing component. Conceptually, recharge is the direct link to the groundwater storage and use of this soil moisture state as the forcing component has more relevance and yielded better results. The models *m5* and *m7* produced the highest median calibration CoE while the model *m6* resulted in the highest unbiased CoE for the evaluation period. The model *m7* yielded compromised performance statistics for both calibration and validation periods.

For calibration, the HARTT model performed significantly worse than all SMS-TFN model variants and the P and E_t forced TFN model (*m8a*). Furthermore, the predictive performance of the HARTT model was shown to be highly variable and inconsistent across the bore hydrographs tested. If a sustained linear time trend exists in the bore hydrograph, the model produced good results as indicated by performance measures during both calibration and validation periods. However, in the absence or change in such a trend, the model performed poorly. These results illustrate the considerable risk in adopting the non-climatic time trend of HARTT. More importantly, the SMS-TFN model with an appropriate soil moisture state as the forcing component was shown to produce a robust model that can explain most of the bore hydrographs from climate data alone, thus making the HARTT non-climatic time trend redundant.

Future research will evaluate the performance of various models on bore records from confined and bedrock bores. Future research will also assess under what conditions each model is likely to perform best.. The results presented here are based on the predictive performance of the models evaluated using a validation period of the last 10 years of each record and the predictive performance of various models over shorter periods is under investigation. The results will be published in a future paper.

ACKNOWLEDGMENTS

The Authors are very grateful to the Australian Research Council (LP0991281) for contributing to the funding of this research. The authors are also very grateful to the following research partners for their funding and generous and encouraged support: Dr. Andrew Frost, Mr. Derek Bacon and Dr. Carl Daamen of the Bureau of Meteorology, Mr. Xiang Cheng of the Department of Primary Industries, Victoria and Mr. Chris McAuley of the Department of Sustainability, Victoria.

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Broyden, C.G. (1965). A class of methods for solving nonlinear simultaneous equations, *Mathematics of Computation, American Mathematical Society*, 19(92): 577-593.
- 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, Australia.
- Davidon, W. C. (1991). Variable metric method for minimization, *SIAM Journal on Optimization*, 1(1), 1–17.
- Ferdowsian, R., Pannell, D. J., McCarron, C., Ryder, A. and Crossing, L. (2001). Explaining groundwater hydrographs: Separating atypical rainfall events from time trends, *Australian Journal of Soil Research*, 39(4), 861–875.
- Ferdowsian, R. and Pannell, D. (2009). Explaining long-term trends in groundwater hydrographs, 18th World IMACS / MODSIM Congress, Cairns, Australia, pp. 3109–3115.
- Ferdowsian, R., Ryder, A., George, R., Bee, G. and Smart, R. (2002). Groundwater level reductions under lucerne depend on the landform and groundwater flow systems (local or intermediate), *Australian Journal of Soil Research* 40(3), 381–396.
- Hill, M. C. and Tiedeman, C. R. (2007). Effective groundwater model calibration: with analysis of data sensitivities, predictions, and uncertainty, John Wiley & Sons, New Jersey, USA.
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares, *The Quarterly of Applied Mathematics*, 2, 164–168.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431–441.
- Morton, F.I. (1983). Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology. *Journal of Hydrology*, 66, 1-76.
- Nash, J.E. and Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part 1 – A discussion of principles, *Journal of Hydrology*, 10(3), 282-290.
- Peterson, T.J. and Western, A.W. (2011). Time-series modelling of groundwater head and its de-composition to historic climate periods, 34th IAHR Congress, Brisbane, Australia.
- Raupach, M.R., Briggs, P.R., Haverd, V. King, E.A., Paget, M. and Trudinger, C.M. (2009). Australian Water Availability Project (AWAP): CSIRO Marine and Atmospheric Research Component: Final Report for Phase 3. CAWCR Technical Report No. 013, 67 pp.(<http://www.csiro.au/awap>)
- Shapoori, V., Peterson, T.J, Costelloe and Western, A.W. (2011). Quantifying the impact of pumping on groundwater heads using observation data and advanced time series analysis, MODSIM 2011, Perth, Australia (to be presented).
- Sinclair Knight Merz (2005). Five year technical review of the Neuarpur WSPA groundwater management plan, Technical Report WC02865, Melbourne, Australia.
- Sinclair Knight Merz (2006). National Action Plan for Salinity and Water - Benchmarking regional water table and trends, Technical Report WC02661, Melbourne, Australia.
- SKM and GHD (2009). Hydrogeological mapping of southern Victoria, Sinclair Knight Merz , Armadale, Victoria, SKM Project Number VW04237.
- Spatial Vision (2009). Ground water northern model, Spatial Vision, Melbourne, Victoria, Project Number SV002946.
- von Asmuth, J. R. and Bierkens, M. F. P. (2005). Modeling irregularly spaced residual series as a continuous stochastic process, *Water Resources Research*, 41(12).
- von Asmuth, J. R., Bierkens, M. F. P. and Maas, K. (2002). Transfer function-noise modeling in continuous time using predefined impulse response functions, *Water Resources Research*, 38(12).
- von Asmuth, J. R., Maas, K., Bakker, M. and Petersen, J. (2008). Modeling time series of ground water head fluctuations subjected to multiple stresses, *Groundwater*, 46(1), 30–40.
- Yihdego, Y. and Webb, J.A. (2011). Modeling of bore hydrographs to determine the impact of climate and land-use change in a temperate subhumid region of southeastern Australia, *Hydrogeology Journal*, Hannover, 19(4), 877-887.