Author/s:
Peterson, TJ;Western, AW

Title:
Statistical Interpolation of Groundwater Hydrographs

Date:
2018-07-01

Citation:
Peterson, T. J. & Western, A. W. (2018). Statistical Interpolation of Groundwater Hydrographs. Water Resources Research, 54 (7), pp.4663-4680. https://doi.org/10.1029/2017WR021838.

Persistent Link:
https://hdl.handle.net/11343/284323

Peterson Tim, J. (Orcid ID: 0000-0002-1885-0826)

Western Andrew, William (Orcid ID: 0000-0003-4982-146X)

Statistical interpolation of groundwater hydrographs

Tim. J. Peterson[1] and Andrew W Western[1]

1. Department of Infrastructure Engineering, The University of Melbourne, Victoria, Australia

Abstract

Groundwater observation bores are often monitored irregularly and infrequently. The resulting groundwater hydrographs are consequently less informative for understanding groundwater level trends, seasonality, flow directions, draw-down and recovery. This paper presents an approach to temporally interpolate a groundwater hydrograph that has an irregular observation frequency to daily time-steps. The approach combines nonlinear transfer function noise modeling with temporal kriging of the model residuals to produce an interpolated hydrograph that honors all water level observations input to the modeling and accounts for meteorological forcing between the observations. The reliability of the approach was evaluated using six observation bores having extended periods of daily data and by re-sampling them to six observation frequencies ranging from weekly to annually. The analysis showed that for weekly to monthly re-sampled data, >90% of the observed daily variability can be simulated at 4 of 6 bores. The performance declined with observation step-size, as expected, but even at a biannual time-step the error corrected interpolation can explain >70% of the variance at 3 of 6 bores. Additionally, an application shows that (1) the probability of a water level depth being exceeded can be estimated from quarterly re-sampled data and (2) the median annual water level range can be estimated from monthly re-sampled data. Supplementing less frequent observations with

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2017WR021838

6 and 12 months of daily data was also examined, with the addition of a 12 month period significantly improving interpolation results at 3 of the 4 analyzed bores. The approach has been incorporated into the *HydroSight* toolbox http://peterson-tim-j.github.io/HydroSight/.

1 Introduction

Groundwater level hydrographs are essential for understanding trends and seasonality in aquifer dynamics, assessing the impacts from drivers such as pumping, assessing flow paths and for calibrating numerical groundwater models. However, too often such uses are compromised by inadequate or irregular groundwater level monitoring. To illustrate the irregularity of monitoring, Fig. 1 shows box plots of the yearly monitoring frequency over 45 years at each monitoring bore in Victoria, Australia. It is clear that monitoring frequency is most often monthly or less frequent, and in Victoria the frequency has declined over this period. This paper presents a time-series statistical approach to interpolate irregular and infrequently monitored observation bores to regular time-steps. The approach is systematically evaluated using six bore hydrographs with extended periods of daily observations. The approach requires only the historic daily meteorological forcing and observed groundwater hydrograph and requires no specification of the site hydrogeology. It has been incorporated into the *HydroSight* toolbox http://peterson-tim-j.github.io/HydroSight/, a freely available toolbox for groundwater time-series modeling [Peterson & Western 2014] that also allows decomposition of groundwater hydrographs to the contribution from drivers [Shapoori et al. 2015a,b] and the data-driven estimation of aquifer hydraulic properties [Shapoori et al. 2015c].

Groundwater hydrograph interpolation has received little attention. Our anecdotal understanding is that groundwater hydrographs with irregular or infrequent groundwater observations are utilized by (a) linear interpolation between observations to the date of interest; (b) adoption of the closest

observation to the date of interest; or (c) averaging of all observations over some period, typically a year. These practices can result in averaging of high frequency dynamics and other impacts on the time series characteristics and, when undertaken across a region, are likely to artificially increase the spatial heterogeneity. To our knowledge, the only more sophisticated approach is by Berendrecht and van Geer [2016] who present a multiple site linear transfer function noise (TFN) model approach that simulates a hydrograph. They first built a TFN model for each individual observation bore in a neighborhood and then adjusted the estimates at the hydrograph of interest using the TFN model residuals from neighboring TFN models. Their approach has multiple uses, one of which was hydrograph interpolation. The resulting interpolation did not always honor the observation points and in some cases caused an obvious bias over the interpolation period. The approach relies on one of the neighboring bores being observed at a high frequency commensurate with the interpolated time-step. For the Netherlands, where the application was developed, this requirement is not a problem but many regions of the world are not in this enviable position. Furthermore, even if data-loggers were now installed in many regions of the world, challenges would still remain for historic periods.

Similar, modern interpolation techniques for other hydrological problems have utilized multiple sites [Mwale et al. 2012]. Such problems include soil moisture infilling [Dumedah & Coulibaly 2011], rainfall infilling [Bárdossy & Pegram 2014] and streamflow infilling [Harvey et al. 2012] and each utilized statistical relationships between the study site and *donor* sites. In all such applications, the selection of the donor sites is an important and non-trivial exercise [Harvey et al. 2012] and, despite the development of semi-automated approaches for stream gauge donor selection and testing, most donor selection methods require expertise, fine tuning and knowledge of the sites [Giustarini et al. 2016]. Furthermore, Mwale et al. [2012] raised the issue that the donor sites cannot have similar gaps to the target stream gauge and that when there are long sequences of missing data, hydrological

modeling should be considered; which we assume to mean conceptual rainfall-runoff modeling but are unaware of any such applications.

In the context of groundwater interpolation, these challenges are likely to be amplified. Specifically, sub-surface aquifer heterogeneity and the influence of boundary conditions is likely to make the selection of donor sites even more challenging, and even when there is a locally dense observation network the donor sites should be selected to ensure they are influenced by similar forcing to the study site. Furthermore, even if suitable donor sites can be identified, the paucity of groundwater observations makes it unlikely that the donor sites would have daily observations at each interpolation time point and have sufficient coincidental observations with the study site and the other donor sites to build the required statistical relationships.

In light of these challenges, this paper presents an approach for groundwater interpolation using a single site and aims to produce interpolation results without discontinuities at observations and that account for the variability between observations. The approach combines a nonlinear TFN approach [Peterson & Western 2014], which is an extension of von Asmuth and Bierkens [2005], with universal temporal kriging of the model residuals to produce an interpolation estimate that honors the observations and accounts for forcing between the observations. Other methods could have been adopted instead of the TFN modeling, and notable alternatives include Kalman filter TFN models [Bierkens & Knotters, 1999], stochastic differential equations [Bierkens, 1998] and various autoregressive models [Knotters & Bierkens, 1999]. The kriging of the residuals could foreseeably be combined with each alternative, but our trials of the nonlinear TFN model have shown it to be highly effective at simulating a range of groundwater head dynamics in both water and energy limited climates; hence it is used here.

In this paper, the groundwater interpolation approach is first detailed, followed by the approach for its evaluation at 6 observation bores. Methods are then detailed for sensitivity analysis of the kriging approach, testing of alternative approaches to the kriging and the evaluation of two applications; namely estimating the exceedance probability of a threshold water level and estimating the annual range in groundwater level. The results section then details the outcomes from the TFN modeling and the aforementioned evaluations, after which the benefit from installing a data-logger to improve the interpolation is evaluated. Finally, a discussion and conclusions section critically reviews the findings and the conditions under which the interpolation approach is likely to be reliable.

2 Methods

This section details the groundwater hydrograph interpolation approach and its evaluation. The approach is available as a graphical module within the *HydroSight* toolbox (http://peterson-tim-j.github.io/HydroSight/). The evaluation was undertaken using six groundwater observation bores, each having periods of daily observations. The high frequency periods were sub-sampled to time-steps ranging between weekly and annual and then interpolated to a daily time step. The reliability of the interpolation was then calculated from the difference between the observed daily water level and the interpolated estimates. The approach was then further evaluated by (1) quantifying the sensitivity of an aspect of the interpolation scheme, (2) evaluating the adopted interpolation scheme against two alternatives and (3) applying the approach to two typical applications, specifically estimating the probability of exceeding a water level and estimating the annual water level range.

2.1 Groundwater hydrograph interpolation procedure

The interpolation procedure has three major steps: (1) building and calibrating a nonlinear transfer function noise (TFN) model using groundwater level data; (2) quantifying the temporal correlation in the residuals; and (3) interpolating the residuals to the interpolation time points and adding them to the TFN simulated water level. The interpolation of the residuals was undertaken using universal kriging. Kriging was chosen because, unlike other approaches such as splines or linear interpolation, the estimates are unbiased, local temporal trends can be accounted for and an estimate of the uncertainty is provided.

The TFN model is detailed in Peterson and Western [2014]. In this application the observed groundwater level elevation at each bore, h_t [L] was simulated as a function of daily precipitation and areal potential evapotranspiration (APET) only. The nonlinear partitioning of precipitation into recharge and other components used the following soil moisture model:

$$\frac{dS}{dt} = P_t - E_t\left(\frac{S}{S_{cap}}\right) - k_{sat}\left(\frac{S}{S_{cap}}\right)^{\beta} \quad (1)$$

where S [L] is the soil water storage at time t [T], S_{cap} [L] is a parameter for soil moisture storage capacity; P_t [LT^{-1}] is the rate of precipitation and E_t [LT^{-1}] is the APET at time t, k_{sat} [LT^{-1}] is the soil vertical saturated conductivity and β [-] is a dimensionless parameter controlling the free drainage as a function of wetness. In this form of the soil model all precipitation is assumed to infiltrate until saturation. The free drainage, D_t [L], and phreatic evapotranspiration (ET) demand, G_t [L] , were estimated as follows:

$$D_t = k_{sat}\left(\frac{S}{S_{cap}}\right)^{\beta} \quad (2)$$

$$G_t = E_t\left(1 - \frac{S}{S_{cap}}\right) \quad (3)$$

To transform these meteorological signals to a simulated head, equation 15 from Peterson and Western [2014] was adopted with the landuse change parameter, f_L [-], set to zero. The model was calibrated to the observed hydrograph using the same objective function as Peterson and Western [2014] and the SP-UCI global calibration scheme [Chu et al. 2011], with 5 complexes per parameter.

The residuals, ε_t [L], were calculated as:

$$\varepsilon_t = h_t - h_t^* - d \quad (4)$$

where h_t^* [L] is the simulated water level at t and d [L] is the drainage elevation constant derived post calibration to produce a mean bias of zero during the calibration period.

The temporal correlation of the residuals at each bore was quantified against the time-step size by calculating the mean semi-variance, $\hat{\gamma}$ [L^2], at Δt:

$$\hat{\gamma}(\Delta t) = \frac{1}{2N(\Delta t)} \sum_{\alpha=1}^{N} [\varepsilon(t_\alpha) - \varepsilon(t_\alpha + \Delta t)]^2 \quad (5)$$

Here $N(\Delta t)$ is the number of pairs of data points in the Δt variogram bin range, Δt [T] is the variogram bin range defining an upper and lower limit for the absolute time difference between data pairs, $\varepsilon(t_\alpha)$ [L] is the residual data point at t_α and $\varepsilon(t_\alpha + \Delta t)$ [L] is a residual data point Δt away. Twelve increments of Δt were used. The maximum Δt was 365 days and each increment was 365/12 days

An exponential model was fitted to the experimental variogram:

$$\gamma(\Delta t) = C_0 - C_1 \left(1 - exp\left(\frac{-\Delta t}{a}\right)\right) \quad (6)$$

where C_0 [L^2] is the variogram nugget, C_1 [L^2] is the variogram sill and a [T] is the variogram range [T]. The exponential variogram model was adopted because, unlike a spherical variogram, the sill asymptotes toward the maximum and so time-points beyond the model range have an influence that decays with the time difference and because, unlike a Gaussian variogram, the model nugget can equal zero without causing numerical stability issues. For further details of model variogram see Goovaerts [1997].

The Nelder-Mead simplex algorithm [Lagarias et al. 1999], as implemented within the MatLab *fminsearchbnd* function [D'Errico, 2006], was used to perform a least-squares fit of the model variogram. Lower and upper bounds on each of the three variogram parameters were zero and infinity. The algorithm was adopted because it is a robust and efficient method for low dimensional problems (herein 3 dimensions). It is, however, a local optimization method and so is not guaranteed to identify the lowest error variogram solution (i.e. the global optima). Schwanghart's [2010] code was used for both variogram calculation and fitting.

With the serial correlation quantified by the fitted variogram, the final step was to use the model variogram to interpolate the hydrograph residuals to the required interpolation time points, $\boldsymbol{t_{sim}}$, using the model residuals from observed time points (not other bores) and then add the result to the simulated water level. To interpolate the residuals, a local universal kriging approach was adopted. The local kriging scheme involved the use of only a sub-set of data points nearest to each time point being kriged. The mean residual is therefore estimated from the sub-set of data, which allows it to change with time. This results in estimates being consistent with the recent mean error rather than the global error. The universal kriging aspect of the scheme allowed for a locally varying linear trend in the residuals, which was most useful when the residuals have a clear trend or in applications where the hydrograph is extrapolated.

To estimate a single simulation time point, t_{sim}, observations in the local neighborhood were first identified. This was undertaken using an approach analogous to the quadrant search approach of most geostatistical packages [Goovaerts 1997; Deutsch & Journel 1998; Pebesma 2004]. Specifically, 24 observations were sought (note, analogous to spatial kriging, all points must be at unique time points). They were found by first identifying the closest 25% of the required number of observations prior to t_{sim} and 25% after t_{sim}. This ensures simulation points, say, immediately before a densely observed period also include observations from lower density prior periods. The remaining 50% of the required observations were the closest observations not already selected and there was no maximum time difference for the selection of observations. Combined, this somewhat complicated search strategy produced a smoother universal kriging trend than simply taking the closest required number of points. For details see the *HydroSight* function *interpolateData()* within the file *HydroSightModel.m*.

With the local observations identified, the next step was to estimate the residual at t_{sim} using universal kriging. Following Goovaerts [1997], the set of universal kriging linear equations were:

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} & \cdots & C_{1n} & 1 & t_1 - t_{\text{sim}} \\ C_{21} & C_{22} & C_{23} & \cdots & C_{2n} & 1 & t_2 - t_{\text{sim}} \\ C_{31} & C_{32} & C_{33} & \cdots & C_{3n} & 1 & t_3 - t_{\text{sim}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & 1 & \vdots \\ C_{n1} & C_{n2} & C_{n3} & \cdots & C_{nn} & 1 & t_n - t_{\text{sim}} \\ 1 & 1 & 1 & \cdots & 1 & 0 & 0 \\ t_1 - t_{\text{sim}} & t_2 - t_{\text{sim}} & t_3 - t_{\text{sim}} & \cdots & t_n - t_{\text{sim}} & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_n \\ \mu_0 \\ \mu_1 \end{bmatrix} = \begin{bmatrix} C_{10} \\ C_{20} \\ C_{30} \\ \vdots \\ C_{n0} \\ 1 \\ 0 \end{bmatrix} \quad (7)$$

where $t_1 - t_{\text{sim}} \ldots t_n - t_{\text{sim}}$ [T] is the time difference in units of days between the observation data point $t_1 \ldots t_n$ and the simulation time point; $\lambda_1 \ldots \lambda_n$ [-] are the kriging weights obtained by solving the linear equations; μ_0 and μ_1 are coefficients for the local search neighborhood mean and temporal trend respectively; $C_{11} \ldots C_{nn}$ [L^2] are the temporal covariance between the re-sampled data points;

and $C_{10}\ldots C_{n0}$ [L^2] are the temporal covariance between a re-sampled data point and the simulation time point. The covariance terms were calculated from a rearrangement of the variogram:

$$C_{ij} = \begin{cases} C_0 + C_1 & \text{if } |t_i - t_j| = 0 \\ C_1 exp\left(\frac{-\Delta t}{a}\right) & \text{if } |t_i - t_j| > 0 \end{cases} \quad (8)$$

Solving for the kriging weights and coefficients, the interpolated residual, $\epsilon_{\widetilde{\text{sim}}}$ [L], and kriging variance, σ^2_{sim} [L^2], are:

$$\epsilon_{\widetilde{\text{sim}}} = \sum_{i=i}^{n} \lambda_i * \varepsilon_i \quad (9)$$

$$\sigma^2_{\text{sim}} = C_0 + C_1 - \sum_{i=i}^{n} \lambda_i * C_{i0} - \mu_0 \quad (10)$$

Repeating the interpolation for all values of t_{sim} produces a vector of interpolated residuals, $\epsilon_{\widetilde{\text{sim}}}$. To produce the final interpolated head time-series, the calibrated TFN model is first used to estimate the head at each interpolation time point, h_{sim}. Next, the simulated head and residuals are summed to produce an interpolated hydrograph, $h^{\sim}_{\text{sim}}$, that honor all observations within the calibration period:

$$h^{\sim}_{\text{sim}} = h_{\text{sim}} + \epsilon_{\widetilde{\text{sim}}} \quad (11)$$

In closing, four aspects of the interpolation warrant discussion. Firstly, the inclusion of a nugget in the variogram model assumes there is some variability in the residuals not explained by temporal correlation within the residuals. The implications are that simulation time points infinitesimally prior to or after an observation may notably depart from the observation (when the observation is further from the local mean than the simulation), with the estimate at the observation appearing as a water level spike.

Secondly, the kriging interpolates the residuals as the sum of a local mean that smoothly varies with the local search neighborhood and a stationary temporally correlated random variable conditioned on the neighboring observational data. Ideally, the varying local mean would be known a priori and the variogram would be derived from the difference between the model residuals and the mean. However, in most applications this is not possible because the trend varies over time; and only in rare cases do the residuals have non-zero linear trend over the whole record that could have been estimated a priori. When there is a trend over time, the variogram continues to rise with the time difference, rather than flatted out. This was most apparent for 124669 (see SI 2 Fig. 1). The implications of this are that the variogram range is over estimated and, hence, observations further away from the point of estimation will have a greater influence causing the interpolated hydrograph to appear smoother than it should.

Thirdly, fitting the variogram model using local calibration, while being efficient, may not identify the global optima. Global calibration schemes could be implemented, as could more advanced variogram objective functions that account for the data-pairs, both of which may improve interpolation performance. The increased computation time would, however, inhibit application to TFN model calibration and uncertainty analysis schemes that produce 10,000+ plausible parameter sets; such as Markov Chain Monte Carlo schemes [Shapoori et al. 2015c].

Fourthly, the interpolation is bi-directional in time. Observations after a given simulation time point influence the interpolation, and therefore the interpolation does not adhere to causality.

2.2 Evaluation of interpolation procedure

To evaluate the interpolation approach, groundwater observation bores were sought with at least 3 consecutive months of sub-daily water level observations. Additionally, the minimum record length was 4 years and bores were sought from confined and unconfined aquifers. To identify these bores,

the observation count from each of the 33,030 Victorian (Australia) groundwater hydrographs was calculated and 47 bores were identified with at least 3,000 observations. The hydrographs from these 47 bores were then manually examined to identify those that appear reliable (i.e. no displacements and no obvious major anomalies). From this process, six bores were identified. Note that these rigorous criteria were necessary to enable evaluation of the approach and do not reflect the data requirements for application.

For the six bores, Table 1 summarizes the location, screen depth and length of the total record and of each period of sub-daily observations. Importantly, the six bores reflect a diverse sample of hydrogeological considerations. Two bores monitor semi-confined aquifers, three bores are screened at greater than 60 m, one at less than 3 m and two bores have a groundwater level trend that rises during a prolonged drought. Considering the TFN model assumes the water level has a positive correlation with precipitation, and negative with APET, the latter two bores represent an example where the interpolation is anticipated to be poor and hence inform the bounds of performance, i.e. a worst-case scenario. All bores were modeled using an identical method.

To assess the approach, the following procedure was applied to each bore for hydrograph interpolation:

1. Identify and remove erroneous and outlier water level observations using the automatic procedure from Peterson et al. (2017). Erroneous observations were identified using heuristics, e.g. maximum acceptable level change per day and the maximum acceptable duration of periods of constant water level. Outliers were identified by fitting a double exponential smoothed curve to the observed hydrograph and then identifying observations that are more than three standard deviations of the estimated noise away from the smoothed estimate.

2. Remove all sub daily observations except the last observation of the day. This was undertaken because the TFN model minimum time step is daily.
3. Re-sample each period of daily observations to the required "observation" period. Observation periods were weekly (Sunday observation), fortnightly (every second Sunday), monthly (end of the month), seasonal (end of February, May, August, November), biannual (end of June and December) and annual interpolation (end of December).
4. Calibrate the TFN time-series model to the re-sampled data, and then fit the model variogram to model residuals. The variogram model was fitted to a sample variogram based on the re-sampled data plus data outside the periods of daily monitoring. Use the TFN model and variogram to interpolate the hydrograph to the observations removed in the re-sampling.
5. Calculate the difference between the interpolation and observed water level data and then derive the following measures of performance:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n}\left(h_{\widetilde{\text{sim}}_i} - h_{\text{obs}_i}\right) \quad (12)$$

$$CoE = 1 - \frac{\text{m}}{n}\frac{\sum_{i=1}^{n}\left(h_{\widetilde{\text{sim}}_i} - h_{\text{daily obs}_i}\right)^2}{\sum_{j=1}^{m}\left(h_{\text{daily obs}_j} - \overline{h_{\text{daily obs}}}\right)^2} \quad (13)$$

$$CoE_{\text{unbiased}} = 1 - \frac{\text{m}}{n}\frac{\sum_{i=1}^{n}\left(h_{\widetilde{\text{sim}}_i} - h_{\text{daily obs}_i} - \bar{X}\right)^2}{\sum_{j=1}^{m}\left(h_{\text{daily obs}_j} - \overline{h_{\text{daily obs}}}\right)^2} \quad (14)$$

where $\bar{X}$ [L] is the mean bias, CoE [-] is the coefficient of efficiency. *CoE*=1 equates to a perfect interpolation and, in this formulation of *CoE*, less than zero approximately equates to an estimate worse than estimating using the observed mean (see red dots on Fig. 4C for the actual *CoE* value where the mean observed head is used as the prediction). CoE_{unbiased} [-] is the unbiased coefficient of efficiency. $h_{\widetilde{\text{sim}}_i}$ [L] is the interpolated water level at a period of daily monitoring that was removed from the modeling, $h_{\text{daily obs}_i}$ [L] is an observed water level at a period of daily monitoring that was removed from the modeling, and $\overline{h_{\text{daily obs}}}$[L] is the mean water level derived from all daily water

level monitoring, n [-] is the number of daily water level observations removed from the modeling and m [-] is the number of all daily water level observations. Note, the denominator in E and E_{unbiased} was derived using all daily observations. This is an unusual formulation of *CoE* but it was used so that the coefficient of efficiency is relative to the observed daily variability and is comparable across the re-sampling time-steps. That is, for a given bore, if the interpolation of weekly and monthly re-sampled data both produced a *CoE* of 0.9 then this indicates that both were able to explain 90% of the daily water level variability. The ratio m/n accounts for different numbers of data points in each summation. Additionally, CoE_{unbiased} was included in the analysis to discriminate between a poor CoE result that occurs because the model is biased but simulates the observed dynamics well and a model that does not simulate the observed dynamics.

2.3 Evaluation of the kriging procedure

To further evaluate the approach, two aspects of the interpolation scheme were evaluated. Firstly, the significance of the somewhat arbitrary decisions to use 24 data points in the kriging was assessed by repeating the interpolation using 12 and 36 data points and calculating the change in the *CoE* relative to using 24 data points.

Secondly, alternative interpolation methods were assessed by replacing the kriging step with (1) linear interpolation between each pair of residuals and (2) a cubic spline fitted over the entire record of residuals. This analysis was repeated for all bores and re-sampling time steps and the results are presented as a change in the interpolation *CoE* relative to kriging.

2.4 Evaluation of example applications

The interpolated hydrograph was also evaluated for two typical applications: (1) estimating the probability of a threshold water level being exceeded; and (2) estimating the groundwater level

seasonality. The evaluation was undertaken for all re-sampling time steps. The criteria for selecting bores for this evaluation were (1) that the daily hydrograph had sufficient within year fluctuations to ensure frequent exceedance of the threshold and (2) that the daily water level observations span most of the record so that the estimations could be evaluated against a known metric. By these criteria (primarily the first), only bore 128044 was selected.

To estimate the probability of exceeding a threshold water level, a given re-sampled time step hydrograph was interpolated to a daily time-step and then the fraction of days with the water table depth less than 1.2m was calculated. This was then evaluated against the equivalent fraction estimated from the observed re-sampled data for the same period (i.e. with no simulation or interpolation) to assess the benefits from using the interpolation approach. Both estimates were also evaluated against the probability derived from all of the observed daily data.

To estimate the annual water level range, a re-sampled time step hydrograph was interpolated to a daily time-step and then the range in each calendar year was estimated as the year's maximum water level elevation minus the year's minimum water level elevation. The variation in the annual range was then presented using a box-plot. The estimate was compared against (1) the median range from all years of observed re-sampled data and (2) a box-plot of the annual range derived from the observed daily data.

3 Results

3.1 Interpolation models

The first step in the hydrograph interpolation is the calibration of the TFN time-series model to the re-sampled hydrograph. Fig. 2 shows the calibration coefficient of efficiency (*CoE*) for each bore and re-

sampled time-step. It shows that for three bores the *CoE* was >0.5 for all re-sampled time-steps. Comparable results were achieved for 58764 and 128044, but the performance degraded at the biannual time-step for both and at the weekly time-step for 58764. The only consistently poor calibration was for 60628. Interestingly, the calibration performance shown in Fig. 2 does not appear to be related to the screen depth, the aquifer being confined or unconfined, the record length or the sign or magnitude of the response to the severe meteorological drought in south eastern Australia from the mid 1990's to 2010; known as the Millennium Drought [van Dijk et al. 2013]. The only discernible pattern is that the *CoE* appears to decline when the daily water level monitoring spans much of the total record (see 60628 and 128044, SI Fig. 1 and 2 respectively). When these data were re-sampled to quarterly, biannual or annual time-steps minimal data remains below the re-sampled time-step, inhibiting the ability of the TFN time-series model to capture within-year dynamics.

The next step in the interpolation is the fitting of an exponential variogram model to the TFN time-series model residuals. For completeness, Fig. 3 shows the model variogram parameters, namely nugget, total sill, normalized sill and range (see SI 2 Fig. 1 for variogram plots). Fig. 3A and Fig. 3C show that the nugget is very low for all bores and time-steps except for 83229 and fortnightly time-steps for 124669; which at a relative nugget of <0.17 is still reasonably low. Fig. 3D shows that the variogram range is greater than zero for all bores and time-steps. Combined with the sill also being greater than zero, this indicates that the residuals from each bore and time-step are serially correlated. Bores 63995 and 128044 had a range of approximately one month for all time-steps while for the four other bores the range was approximately one year (note, at 3 times the range the variogram is at 95% of the sill). Overall, the low nugget indicates that residual variance is predominantly explained by temporal correlation rather than uncorrelated randomness. Incidentally, this indicates that some information in the hydrographs is not captured by the TFN modeling and also that random noise in the measurements is low.

3.2 Interpolation performance

Fig. 4 summarizes the interpolation performance for each re-sampling time-step at each bore. The left column shows results with the variogram interpolation, henceforth referred to as *error corrected*, and the right column shows results without residual interpolation and error correction, henceforth referred to as *non-error corrected*. Note, all 36 interpolated hydrographs summarized in Fig. 4 are shown in SI3 and SI4. SI3 shows the interpolation over a period having daily observations and SI4 shows the interpolation over the entire record length.

Fig. 4A and B show the normalized mean bias when predicting the observed daily water level data. Fig. 4A shows the normalized bias is <0.05 for weekly to monthly time-steps at 5 of the 6 bores and that the bias increases for quarterly to annual time-steps. Comparing Fig. 4A to 4B shows that the error correction did reduce the bias at most bores for a biannual or finer observation time-step. At the annual observation time-step, the improvements to the bias from the error correction were inconclusive.

Fig. 4C and D summarize the coefficient of efficiency. Fig. 4C shows that for weekly to monthly time-steps >90% of the observed variability can be predicted at 4 bores, with only one bore performing worse than simply interpolating using the sample mean water level. The performance declined with observation step-size, as expected, but even at a biannual time-step the error corrected interpolation can explain >70% of the variance at 3 bores. In comparison to the non-error corrected data (Fig. 4D), performance is markedly improved at weekly to month time-steps at 5 of the 6 bores and at all time steps at 4 of the 6 bores. Additionally, to assess if the improved *CoE* from the error correction is due to the bias, Fig. 4E and Fig. 4F show the unbiased *CoE*. Where the error correction is used, the unbiased CoE (Fig. 4E) is very similar to the *CoE* (Fig. 4C). Fig. 4F shows the unbiased

CoE improved relative to Fig. 4D at those bores with a large bias (58764 and 60628). Overall, the improvements from using the error correction are consistent with those from Fig. 4C and D.

3.3 Evaluation of the temporal kriging

The kriging of the residuals at each time point was undertaken using 24 data points (or all points if there were less than 24 data points). This is a somewhat arbitrary choice and warrants evaluation. Fig. 5 shows the change in the interpolation *CoE* when a maximum of 12 and 36 data points are used. It shows that for each bore, the sensitivity generally increases with the re-sampling time-step size (i.e. declines with more frequent data). More importantly it shows that halving or doubling the number of data points typically changes the *CoE* by less than ~10^{-2}, and only 63995 in Fig. 5B shows a consistent improvement in *CoE* with 36 points used in the interpolation. The variable sensitivity to the number of data points, particularly at a biannual and annual time-step, suggests that the interpolation performance could potentially be slightly improved by tuning the number of kriging data points for each specific application; however, the sensitivity is typically low.

The use of kriging to interpolate the residuals is but one possible interpolation method. To explore alternative methods, Fig. 6A shows the change in interpolation *CoE* when using linear interpolation and Fig. 6B shows the change when using cubic spline interpolation. Fig. 6A shows that linear interpolation clearly outperformed kriging in only two trials (biannual at 58764, annual at 124669). For eleven other trials kriging clearly outperformed linear interpolation, typically for quarterly or less frequent time steps. The remaining trials show almost identical *CoE* for linear interpolation and kriging. Fig. 6B shows greater improvement from kriging over cubic spline interpolation than for the linear interpolation comparison and the improvement increases with the time-step. Overall, the

evaluation of the kriging approach against two alternatives indicates that kriging is the most reliable of the three approaches investigated.

3.4 Evaluation of example applications

Interpolation of a groundwater hydrograph is likely to be undertaken for some other purpose, such as quantifying some measure of natural variability. To evaluate some potential applications, Fig 7A shows the probability of the depth to water level at bore 128044 exceeding 1.2m when calculated from the interpolated hydrograph, the re-sampled data set and the daily data. It shows that for observation time-steps up to quarterly, the error corrected data provides a better estimate of the probability than that derived using only the re-sampled data. Specifically, at a quarterly time-step the error in the probability derived using only the re-sampled data was 32% and interpolating the quarterly data reduced the error to 3.5%. Conversely, at annual and biannual time steps, the error in the probability was lowest when the re-sampled data was used directly.

To evaluate the estimation of the annual water level range), on the left of Fig. 6B is a box plot of the annual range derived from the daily observed data and to its right are box plots of the annual range derived from the interpolated hydrographs. Overlain is also the median annual range derived from the re-sampled data. The plot shows that the observed median annual range is most reliably estimated at a monthly or more frequent time-step. Specifically, at a monthly time-step the error in the median annual range using only the re-sampled data is 12% and interpolation reduces the error to 2.8%. At less frequent time steps, the interpolation consistently under estimated the observed median annual range but did consistently outperform estimates derived using only the re-sampled data. Specifically, and at a quarterly time step the error in the median annual range using only the re-sampled data is 35% and interpolation reduces the error to 18%, and similarly at a bi-annual timestep the error reduced from 94% to 53%.

3.5 Evaluation of the benefits of installing a data-logger

To further explore the performance of the interpolation, Fig. 8 shows the bore 128044 hydrograph (pre-2000 only) and the error corrected interpolation using monthly and quarterly time-step data. Fig. 8 also shows the non-interpolated TFN estimate and the interpolation uncertainty (calculated by scaling the TFN model total error, σ_n, by the normalized kriging variance: $1.6499\sigma_n \frac{\sigma_{\text{sim}}^2}{C_0+C_1}$) . The bore was chosen to illustrate the approach for two reasons. (i) The high frequency observations extend over much of the total record, which when re-sampled is representative of the scenario where, say, only monthly or quarterly observations exist. (ii) Also the observed hydrograph has a high within-year variability, relative to the inter-annual trend, and so presents a particularly challenging test of the interpolation relative to smoother hydrographs such as 124669 and 58764. Additionally, the observation time-steps presented where chosen to illustrate high and low *CoE* performance; specifically 0.91 and 0.73 for the monthly and quarterly time-step respectively. This difference in performance is clearly reflected in Fig. 8A and Fig. 8B. While the plots show that all observation points are honored, Fig. 8A shows that some of the observed sub-monthly water level dynamics were simulated. The Fig. 8A insert further illustrates this by showing that the interpolated water level can be above or below the adjacent points; which results from the TFN model simulated dynamics. Fig. 8A does however show numerous time-points where the interpolated water level is dampened, with recharge peaks being lower and discharge points being higher than observed, and other periods (e.g. 1999) where the interpolation is overly responsive at sub-monthly time-steps in other periods. At the quarterly time step, Fig. 8B shows a counter-example where the interpolated hydrograph is significantly smoother than the observed hydrograph.

The poor performance of the quarterly interpolation in Fig. 8B is not surprising because there are no higher frequency data available to estimate the TFN model parameters controlling the sub-quarterly

dynamics. This suggests that if higher frequency data were available for even a short period then the interpolation may be improved. To investigate the potential improvements from installation of a data logger for only 6 months, 128044 monthly and quarterly time-step data was supplemented with two such periods (1 January to 30 July 1998 and 1 July to 31 December 1998) and the interpolation was repeated. Fig. 9 shows the interpolated hydrographs using each additional period of data. Somewhat surprisingly, Fig. 9A shows that the monthly time-step results did not notably improve (c.f. Fig. 8A) despite sub-monthly water level dynamics existing during each period of additional data. Conversely, Fig. 9B shows that the quarterly time-step interpolation clearly improved for both trials (c.f. Fig. 8B). The corresponding monthly time-step *CoE* changed from 0.91 without the daily logger data to 0.93 and 0.90 for the respective periods with the logger data. For the quarterly time-step the *CoE* changed from 0.73 without the daily logger data to 0.79 and 0.77 with the addition of the respective periods of the logger data. This is a promising improvement, and suggests that the installation of a water level data logger for a limited period at an infrequently monitored bore can improve the interpolation during prior and subsequent monitoring periods.

To more rigorously explore the potential benefits from installation of a data logger, analysis was undertaken simulating a 6-month data logger period and a 12-month data logger period with 5 bores and 4 bores respectively having sufficient duration daily data for the analysis. The remaining daily data were used to assess each re-sampling time-step.

A complicating factor is that when a data logger is physically installed for, say, 6 months it cannot be known a priori if the water level over the period will be highly dynamic (e.g. multiple recharge events) and have the potential to inform the high frequency components of the TFN model or not. Therefore, in this analysis the period over which daily data was inserted was randomly selected and ten trials were undertaken for each bore and re-sampling time-step. The interpolation performance

was assessed using the *CoE* derived from only the re-sampled period. Fig. 10 shows box plots of the *CoE* results and, for comparison, the *CoE* without the inserted daily data is denoted by a black star (data from Fig. 4C). Fig. 10A shows that *installation* of a 6-month data-logger clearly improved the interpolation performance at two bores (58764, 83229) at quarterly and biannual monitoring time-steps while at the annual time-step the improvement were less clear. At 124669 only ~50% of the 10 high frequency sampling trials led to improvements and at 60628 the performance degraded with the inclusion of daily data. As expected, at the weekly to monthly monitoring time-step the improvement in performance were modest with only two bores showing a slight improvement (58764 and 83229). Fig. 10B shows that *installation* of a 12-month data-logger clearly improved the interpolation performance, relative to the 6-month data-logger, at all re-sampling time-steps at all four bores. Most notably, the interpolation of biannual and annual data clearly improved at bores 58764, 83229 and 128044 and more subtly so at 124669.

4 Discussion and Conclusion

This paper presents an approach to interpolating an observed groundwater hydrograph to daily time-steps. The approach applies the *HydroSight* nonlinear transfer function noise (TFN) model [Peterson & Western 2014] and then undertakes temporal universal kriging of the model residuals and adds the results to the TFN estimates. The interpolated hydrograph honors all observations and aims to represent the effect of forcing dynamics between the water level observations. To evaluate the interpolation, six observation bores with extended periods of daily data were identified and then the daily periods of each hydrograph were re-sampled to “observation” frequencies of weekly, fortnightly, monthly, seasonally, biannual and annual time-steps. Interpolated hydrographs were then produced for each bore and re-resampled time-step resulting in daily time-step hydrographs. The interpolation error was then quantified over the period of daily observations. Subsequent evaluation

showed the kriging scheme is generally insensitive to the number of data points used in the local kriging (Fig. 5). It also outperformed two alternative interpolation schemes at most bores and re-sampled time steps (Fig. 6); namely linear interpolation between data pairs and cubic splines.

The analysis found that, for weekly to monthly time-steps, >90% of the observed daily variability could be reproduced by the interpolation at 4 of the 6 bores. At quarterly to annual time-steps the performance was mixed, but the kriging of residuals improved performance at 4 of the 6 bores compared with the TFN alone (Fig 4C c.f Fig 4D). Similarly, the scaled bias was reduced to near zero at the weekly to monthly time-steps for 5 of the 6 bores and was substantially reduced in nearly all cases.

There were some difference in results between bores. At four bores (60628, 83229, 124669, 128044), the TFN with interpolated residuals (Fig 4C) generally performed very well (*CoE* > 0.75) at time steps up to quarterly and showed clear improvement over the TFN alone. For biannual and annual data the residual interpolation improved performance at these sites, but not always to useful levels of explanatory power. Performance improvement was better where the observational record was shorter (see Fig S1 and S2 60628 and 124669 c.f. 83229 and 128044), which may just be an artefact of record length.

At bores 58764 and 63995, the TFN only simulation performed poorly. With residual interpolation, the performance improved strongly for 58764 but not for 63995. Both these bores behave counter intuitively in that water levels rose during a long drought period from the mid 1990's to 2010 experienced in south eastern Australia [van Dijk et al. 2013]. The improvement at 58764 with interpolated residuals likely results from local bias correction coupled with the relatively large temporal scale variability during the high frequency monitoring period. At 63995, short timescale variability appears to dominate the high frequency monitoring period, implying a low signal-to-noise

ratio in the resampled residuals input to the interpolation, which would limit the ability of the interpolation to improve performance.

In general, the signal-to-noise ratio in the residuals is determined by the ratio of the sampling time step to the temporal correlation timescales of the residuals at the particular site. This implies that the improvements from residual interpolation will be site and sampling time-step dependent and will inevitably decline as sampling intervals increase [Western and Bloschl, 1999].

Overall, the utility of the interpolation approach is greatest at sites with a monthly observation frequency or better, with significant declines for quarterly sampling in some cases. However, the median observation frequency throughout Victoria, Australia, is quarterly (Fig 1), and assuming Victoria is representative of other monitoring programs, it is worth considering if the degraded performance at the quarterly time-step is primarily due to a lack of information on high frequency water level dynamics. If this were the case, the interpolation might be improved with an additional 6 or 12 months of daily data if it improved the high frequency behavior of the TFN model. In practice, this could be achieved by installing a data-logger and it is simulated here by including periods of daily data when calibrating the TFN (Fig 10).

The simulated *installation* of a 6-month data logger clearly improved interpolation performance at two of five bores (58764, 83229) at quarterly and biannual monitoring time-steps while at a third (124669) the improvements were modest, most likely because the error correction interpolation *CoE* was already high at >0.8 at all time-steps. At 60628 and 128044 there was no clear improvement at the sub-annual time-steps. The simulated *installation* of a logger for 12-months led to much stronger improvements in the interpolation performance at all re-sampling time steps, including biannual and annual, which clearly improved at three of the four analyzed bores. At the fourth bore (124669), the improvements were small. At the biannual time-step the improvements resulted in a median

CoE >0.75 at all four bores; however, at an annual time-step the median *CoE* at two bores (58464, 128044) was <0.4. This suggests that if bores are monitored infrequently (quarterly to biannually) the installation of a water level data logger for 12 months and subsequent interpolation of the entire water level record is likely to significantly improve the estimates of daily water level dynamics over the entire hydrograph.

The utility of the interpolation approach, no matter the observation frequency, will be dependent upon the available forcing data and the forcing data being able to explain the observed water level dynamics. Two examples where the water level dynamics were inconsistent with the meteorological forcing were investigated (58764 and 63995). In such cases, if application of the scheme produces a poor calibration *CoE* then the bounds of possible interpolation performance can be informed by the TFN simulated water level variability between the observations. That is, if the simulated water level variability is very low then the worst-case scenario will be an approximately linearly interpolated hydrograph. Conversely, if the TFN model water level variability is high then the interpolation performance can produce a very poor interpolated hydrograph. For the latter scenario, the interpolation scheme is not applicable.

Additionally, if unobserved forcing influences the groundwater level at a time scale at or greater than the monitoring frequency, for example land cover change that produces very slow water level change, then the water level monitoring should capture the impacts from the unobserved forcing and hence there would be minimal degradation of the interpolation. For more high-frequency unobserved forcing (compared with the monitoring frequency), the obvious being unmetered seasonal groundwater pumping, the interpolation scheme will be unable to simulate, say, the drawdown and recovery. For such scenarios, the monitoring frequency should be sufficient to capturing the drawdown and

recovery. Alternatively, meters could be installed on the pumps and then more complex TFN models accounting for pumping could be adopted [Shapoori et al. 2015a].

This study has shown that, when the historic forcing has been measured and the groundwater water level response is consistent with the forcing, groundwater hydrographs of monthly or better observation frequency can be reliably interpolated to a daily time step. The definition of 'reliable' is obviously case specific, but where it is deemed reliable the interpolated hydrograph could be used to identify and quantify recharge events; quantify minimum or maximum historic or seasonal groundwater levels; quantify the occurrence or probability of a threshold water level being exceeded. The latter two applications were evaluated for 128044 (Fig. 7) and it was shown that the probability of threshold exceedance could be reliability estimated from quarterly data and seasonality could be reliably estimated from monthly data, but the estimation of seasonality at all time steps was clearly improved using the interpolation scheme. Clearly further evaluation of such applications is required, but these results suggest that the application of groundwater management rules based on water level thresholds [Currell 2016] or seasonality could be strengthened by the interpolation approach when the observation frequency is low.

Our interpolation approach, no matter its strengths, is no replacement for observations. Furthermore, our conclusions are based on just 6 bores, which despite efforts to select bores of considerable diversity, remain a small subset from a very large set of hydrogeological and meteorological scenarios. Numerical issues also remain, namely the relative important of monitoring frequency versus record length, with the latter recently being shown to be most important to TFN modeling [van der Spek & Bakker 2017], and quantifying the interpolation uncertainty. On the latter, some progress has been made using DREAM [Vrugt 2016] and a prototype version is available within *HydroSight* (http://peterson-tim-j.github.io/HydroSight/).

5 Acknowledgements

This research was funded by the Australian Research Council Linkage Project LP130100958 and funding partners, Bureau of Meteorology (Australia), Department of Environment, Land, Water and Planning (Vic., Australia), Department of Economic Development, Jobs, Transport and Resources (Vic., Australia) and Power and Water Corporation (N.T., Australia). The data used are specified in Table 1 and are available at http://data.water.vic.gov.au/monitoring.htm.

Table. 1: Observation bore locations and summary of the total record, high frequency observation periods and the number of data points when the daily data was re-sampled to weekly to annual time-steps.

Fig. 1: Box plots of the annual number of water level observations from all observation bores in Victoria, Australia (n=30,303). The blue box denotes the 25^{th} and 75^{th} percentiles, the red line the 50^{th} percentile, red crosses denote outliers, which are identified as points beyond ±1.5 times the box size and the whiskers extend to the most extreme value that is not an outlier. The plot shows that the 25^{th}, 50^{th} and 75^{th} percentile observation frequencies have declined with notable reductions in the mid-1990's and late-2000's.

Fig. 2: TFN time-series model calibration performance (expressed as coefficient of efficiency, *CoE*) for each re-sampled time-step at each observation bore.

Fig. 3: Residual exponential model variogram parameters where (A) shows the variogram nugget, (B) shows the total sill, (C) shows the relative nugget and (D) shows the range.

Fig. 4: Summary of hydrograph interpolation performance during the periods of daily observation data. The left column shows results using the kriging of residuals and the right shows results without the kriging of residuals. (A, B) shows the mean bias normalised by the standard deviation of the daily observed data, (C,D) shows the coefficient of efficiency and (E,F) shows the unbiased coefficient of efficiency. The red dots in (C) show the coefficient of efficiency for a model predicting the mean observed resampled head.

Fig. 5: Sensitivity of the interpolation *CoE* to the number of data points used in the kriging search neighbourhood. (A) shows the change in *CoE* when 12 data points are used. (B) shows the change in *CoE* when 36 data points are used.

Fig. 6: Change in the interpolation *CoE* when the interpolation of the residuals is changed from using local universal kriging to (A) linear interpolation of data pairs and (B) a cubic spline.

Fig. 7: Evaluation of the interpolation approach to quantify the (A) the probability of the water level being shallower than 1.2m and (B) the water level seasonal range. Each analysis was undertaken for bore 128044 and for all 6 re-sample time-steps. (A) shows the estimated probability using the interpolated hydrograph (grey bars), the observed probability calculated from the daily data (black line) and the probability calculated from the re-sampled data (black star). (B) shows on the left the observed range in annual water level seasonality calculated from the daily data (box-plot), to the right is the interpolation estimate of the range in annual water level seasonality (box-plot) and overlain is the median seasonality calculated from the re-sampled data (black star).

Fig. 8: Hydrograph interpolation examples for bore 128044 over the period of daily observation data prior to 2000. (A) shows the interpolation using re-sampled monthly observation data, and the insert shows the interpolation from the second half of 1997 honouring the monthly observation points. (B) shows the interpolation using re-sampled quarterly data as an example of poor interpolation performance.

Fig. 9: Hydrograph interpolation examples for bore 128044 with two 6 month periods of daily data simulating installation of a data-logger. (A) shows two interpolation results, each using one of the 6 month periods of daily data and the remaining re-sampled monthly observation data. (B) as per (A) but using re-sampled quarterly data.

Fig. 10: Box plots of the hydrograph interpolation coefficient of efficiency at weekly to annual re-sampling time-steps and with the inclusion of a randomly chosen (A) 6- month period of daily data and (B) 12-month period of daily data to simulate the installation of a data logger (note, w: weekly, F:

fortnightly, M: monthly, Q:quarterly, B: biannual, A:annual). For (A), 5 bores had sufficient duration periods of daily data to analyze while (B) 4 bores had sufficient duration daily data. For reference, the coefficient of efficiency from Fig. 4C is presented as a black star in (A) and (B). Ten periods of the data logger were analysed for each data logger duration, re-sampling time-step and bore. The blue box denotes the 25^{th} and 75^{th} percentiles, the red line the 50^{th} percentile, red crosses denote outliers, which are identified as points beyond ±1.5 times the box size and the whiskers extend to the most extreme value that is not an outlier.

References

von Asmuth, J. R. and Bierkens, M. F. P. (2005). Modeling irregularly spaced residual series as a continuous stochastic process, Water Resour. Res. 41 10.1029/2004WR003726.

Bárdossy, A. and Pegram, G. (2014). Infilling missing precipitation records – A comparison of a new copula-based method with other techniques., *Journal of Hydrology* 519 : 1162 - 1170 .

Berendrecht, W. and van Geer, F. (2016). A dynamic factor modeling framework for analyzing multiple groundwater head series simultaneously., *Journal of Hydrology* 536 : 50 - 60 .

Bierkens, M. F. P. (1998), Modeling water table fluctuations by means of a stochastic differential equation, Water Resour. Res., 34 10.1029/98WR02298.

Bierkens, M. F. P., M. Knotters, and F. C. vanGeer (1999), Calibration of transfer function–noise models to sparsely or irregularly observed time series, Water Resour. Res., 35 10.1029/1999WR900083.

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

Currell, M. J. (2016). Drawdown “Triggers”: A Misguided Strategy for Protecting Groundwater-Fed Streams and Springs, *Groundwater* 54 : 619-622 10.1111/gwat.12425.

D'Errico, J. (2006). *fminsearchbnd*. https://au.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd--fminsearchcon

Deutsch, C. V. and Journel, A. G., 1998. *GSLIB Geostatistical Software Library and User's Guide*. Oxford University Press, New York, .

van Dijk, A.; Beck, H.; Crosbie, R.; de Jeu, R.; Liu, Y.; Podger, G.; Timbal, B. and Viney, N. (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 10.1002/wrcr.20123.

Dumedah, G. and Coulibaly, P. (2011). Evaluation of statistical methods for infilling missing values in high-resolution soil moisture data., *Journal of Hydrology* 400 : 95 - 102 .

Giustarini, L.; Parisot, O.; Ghoniem, M.; Hostache, R.; Trebs, I. and Otjacques, B. (2016). A user-driven case-based reasoning tool for infilling missing values in daily mean river flow records., *Environmental Modelling & Software* 82 : 308 - 320 .

Goovaerts, P., 1997. *Geostatistics for natural resource evaluation*. Oxford University Press, New York, .

Harvey, C. L.; Dixon, H. and Hannaford, J. (2012). An appraisal of the performance of data-infilling methods for application to daily mean river flow records in the UK., *Hydrology Research* 43 : 618 - 636 10.2166/nh.2012.110.

Knotters, M., and M. F. P. Bierkens (2000), Physical basis of time series models for water table depths, Water Resour. Res., 36 10.1029/1999WR900288.

Lagarias, J. C.; Reeds, J. A.; Wright, M. H. and Wright, P. E. (1999). Convergence properties of the Nelder-Mead simplex method in low dimensions., *SIAM Journal on Optimization* 9 : 112 .

Mwale, F.; Adeloye, A. and Rustum, R. (2012). Infilling of missing rainfall and streamflow data in the Shire River basin, Malawi – A self organizing map approach., *Physics and Chemistry of the Earth* 50-52 : 34 - 43 .

Pebesma, E. (2004). Multivariable geostatistics in S: The gstat package, *Computers and Geosciences* 30 : 683-691 .

Peterson T. J. and Western, A. W. (2014). Nonlinear time-series modeling of unconfined groundwater head, *Water Resour. Res.* 50 : 8330–8355 10.1002/2013WR014800.

Peterson, T. J.; Western, A. W. and Cheng, X. (in-press). The good, the bad and the outliers: automated detection of errors and outliers from groundwater hydrographs, *Hydrogeology Journal*. 10.1007/s10040-017-1660-7

Schwanghart, W. (2010). *variogramfit*. https://au.mathworks.com/matlabcentral/fileexchange/25948-variogramfit

Schwanghart, W. (2011). *Experimental (Semi-) Variogram*. https://au.mathworks.com/matlabcentral/fileexchange/20355-experimental--semi---variogram

Shapoori, V.; Peterson, T.; Western, A. and Costelloe, J. (2015c). Estimating aquifer properties using groundwater hydrograph modelling, *Hydrological Processes* 29 : 5424-5437 10.1002/hyp.10583.

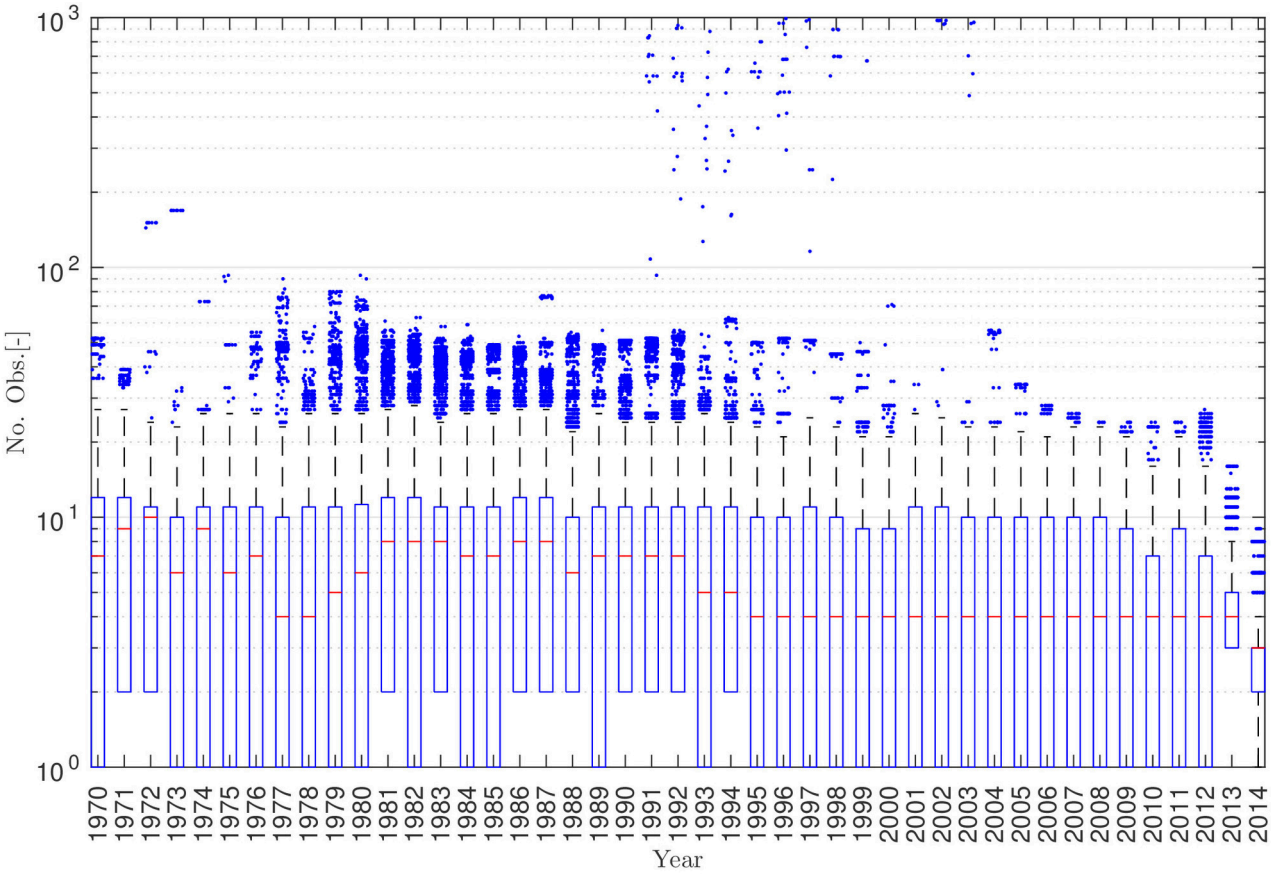
Shapoori, V.; Peterson, T. J.; Western, A. W. and Costelloe, J. F. (2015b). Decomposing groundwater head variations into meteorological and pumping components: a synthetic study, *Hydrogeology Journal* 23 : 1431-1448 10.1007/s10040-015-1269-7.

Shapoori, V.; Peterson, T. J.; Western, A. W. and Costelloe, J. F. (2015a). Top-down groundwater hydrograph time-series modeling for climate-pumping decomposition, *Hydrogeology Journal* 10.1007/s10040-014-1223-0.

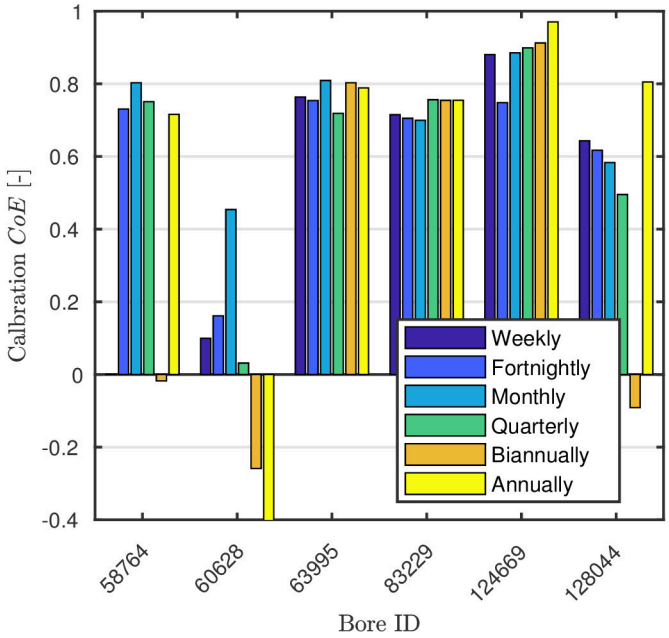
van der Spek, J. E. and Bakker, M. (2017). The influence of the length of the calibration period and observation frequency on predictive uncertainty in time series modeling of groundwater dynamics, *Water Resources Research* 53 : 2294-2311 10.1002/2016WR019704.

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation , *Environmental Modelling & Software* 75 : 273 - 316 10.1016/j.envsoft.2015.08.013

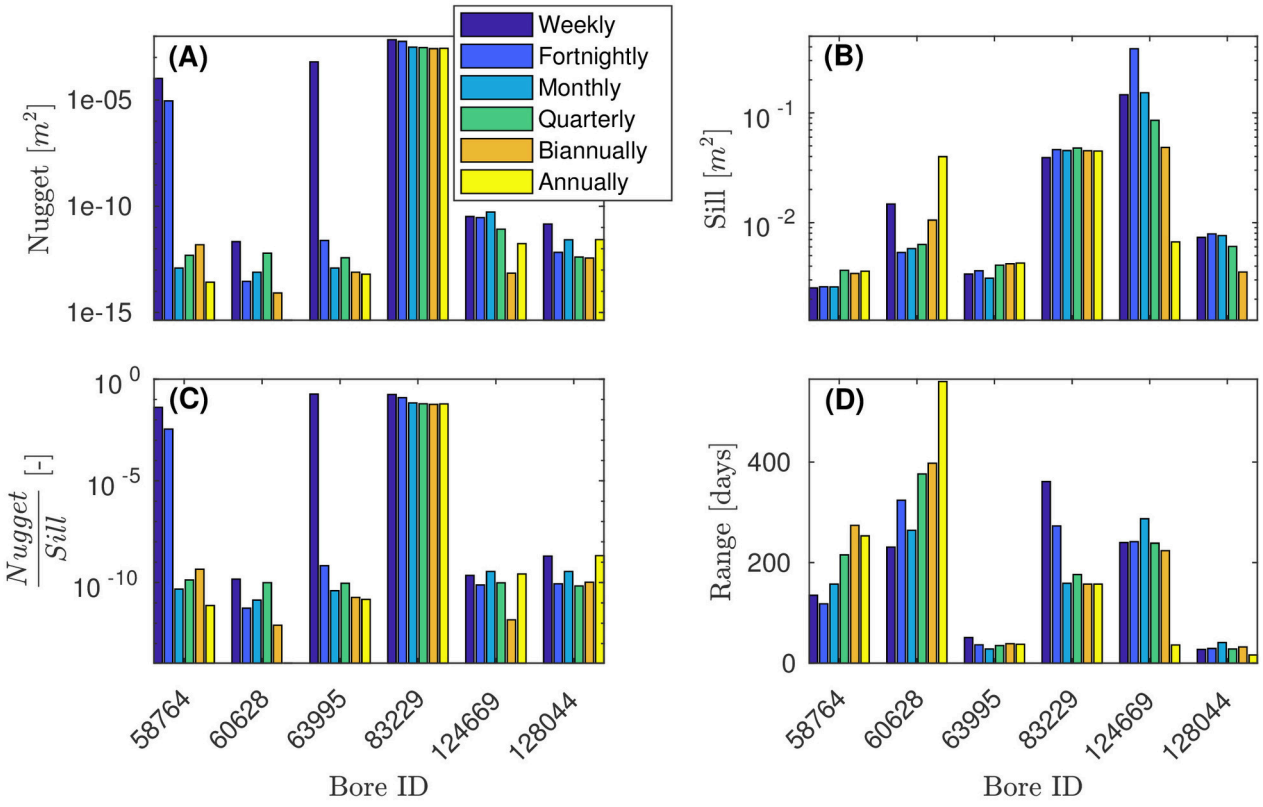
Western, A.W., and Blöschl, G., 1999. On the spatial scaling of soil moisture. Journal of Hydrology, 217: 203-224.



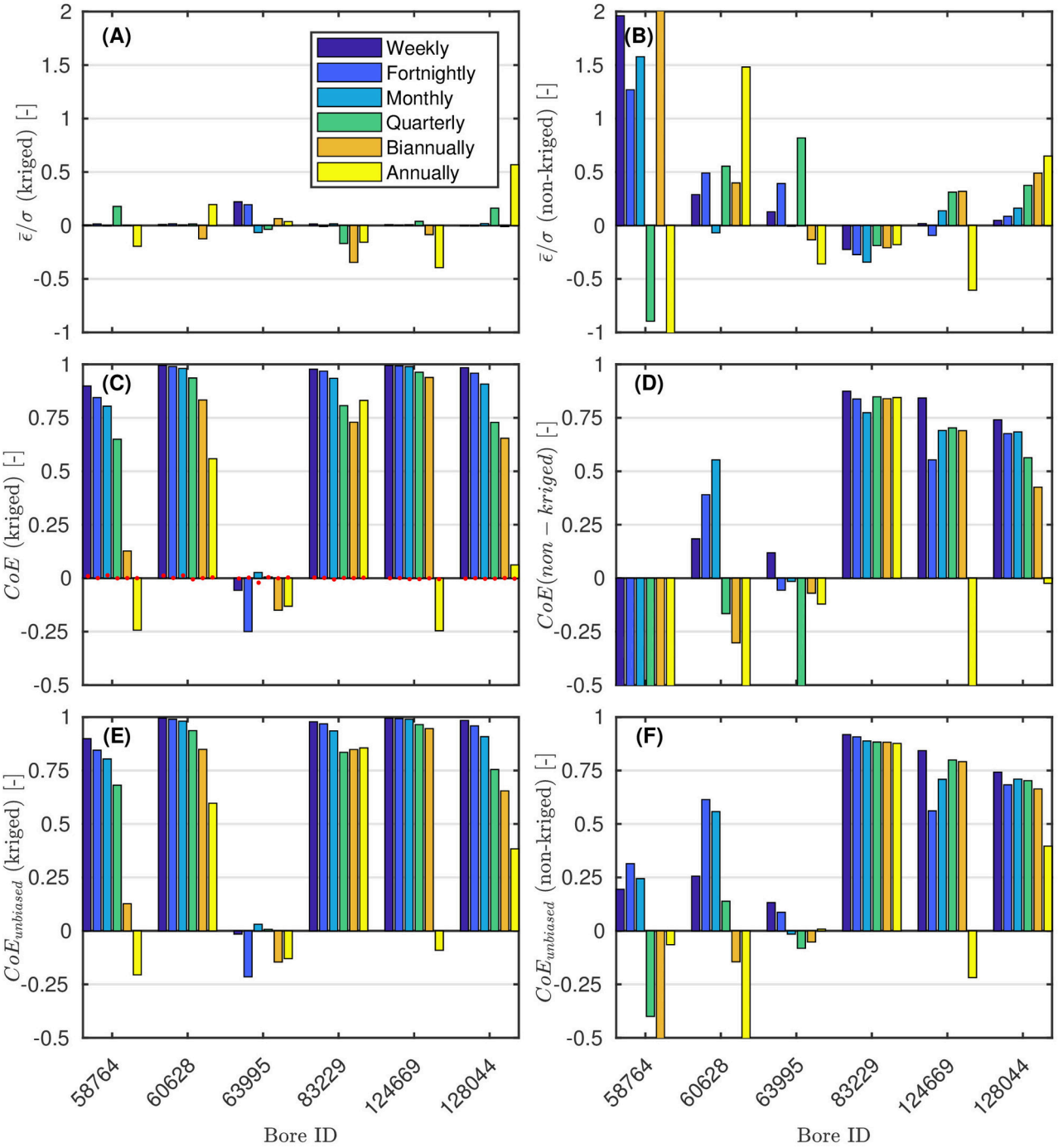
2017wr021838-f01-z-.eps



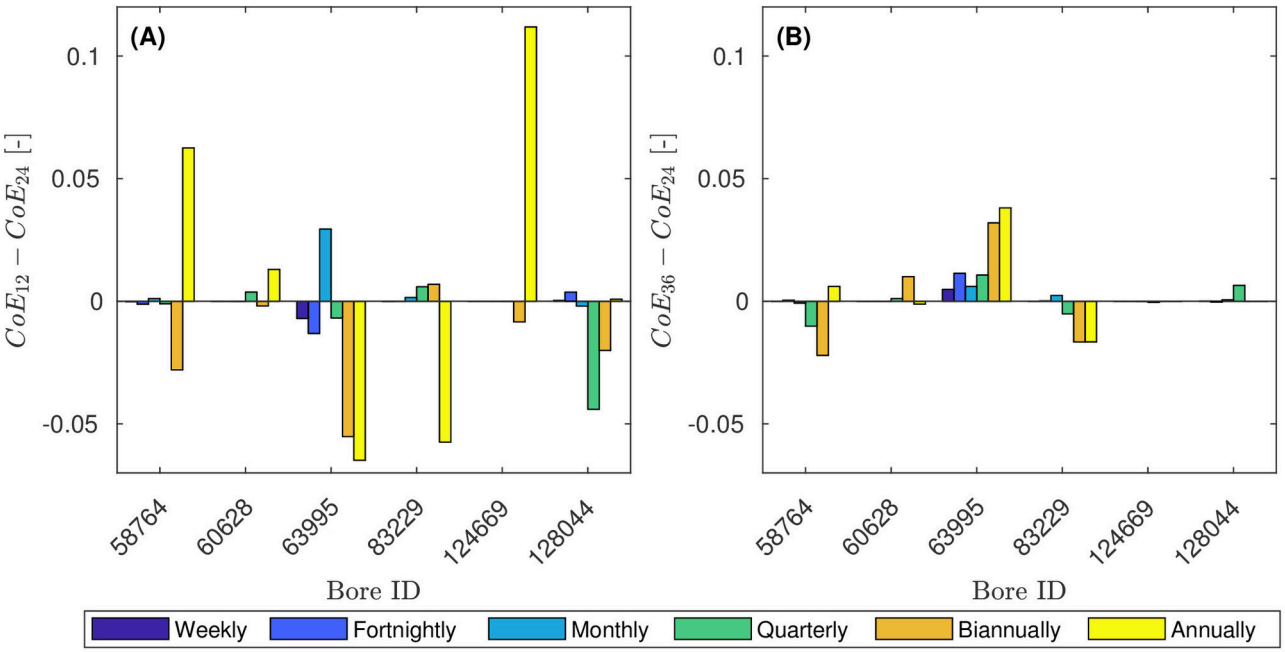
2017wr021838-f02-z-.eps



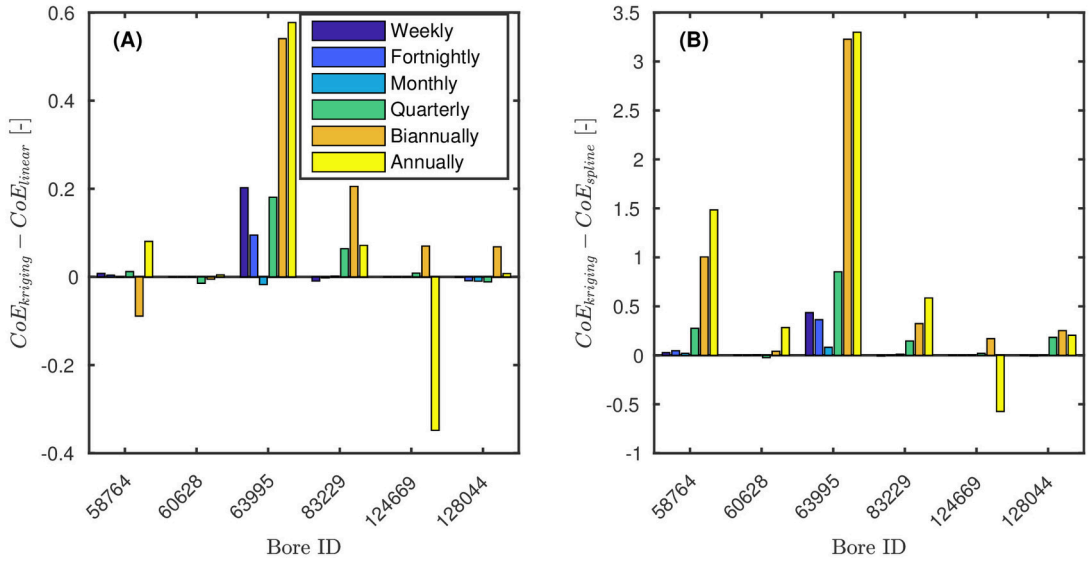
2017wr021838-f03-z-.eps



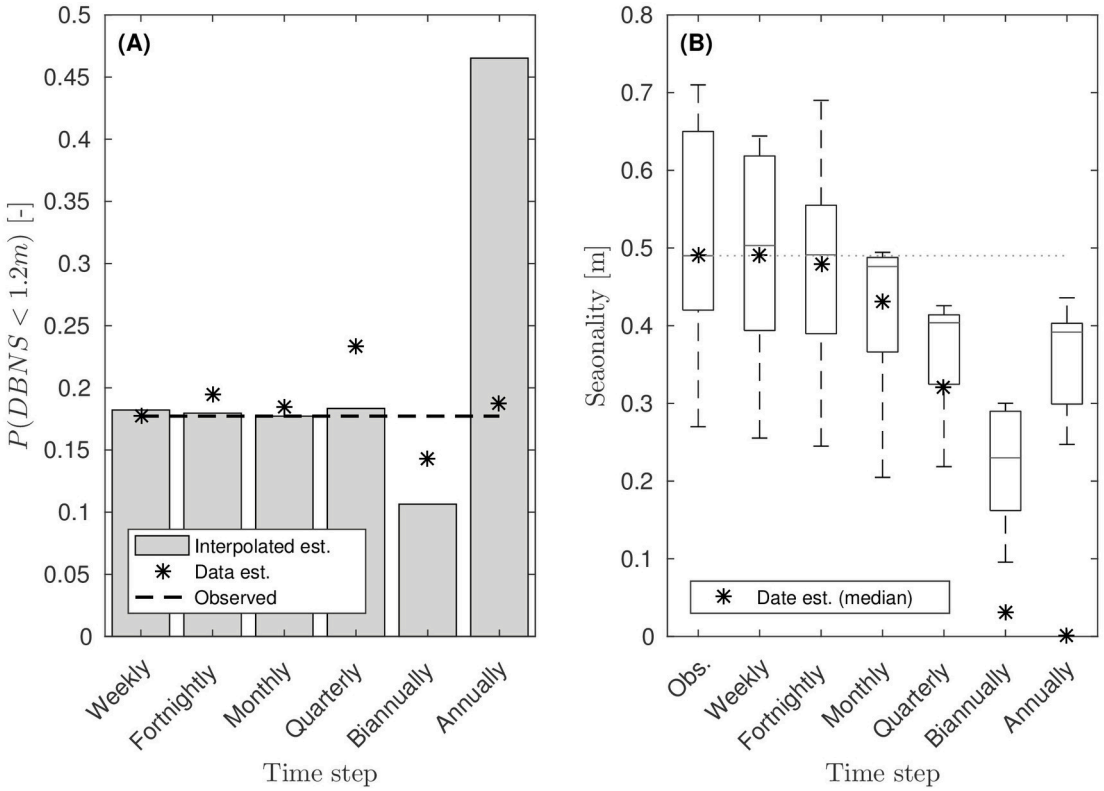
2017wr021838-f04-z-.eps



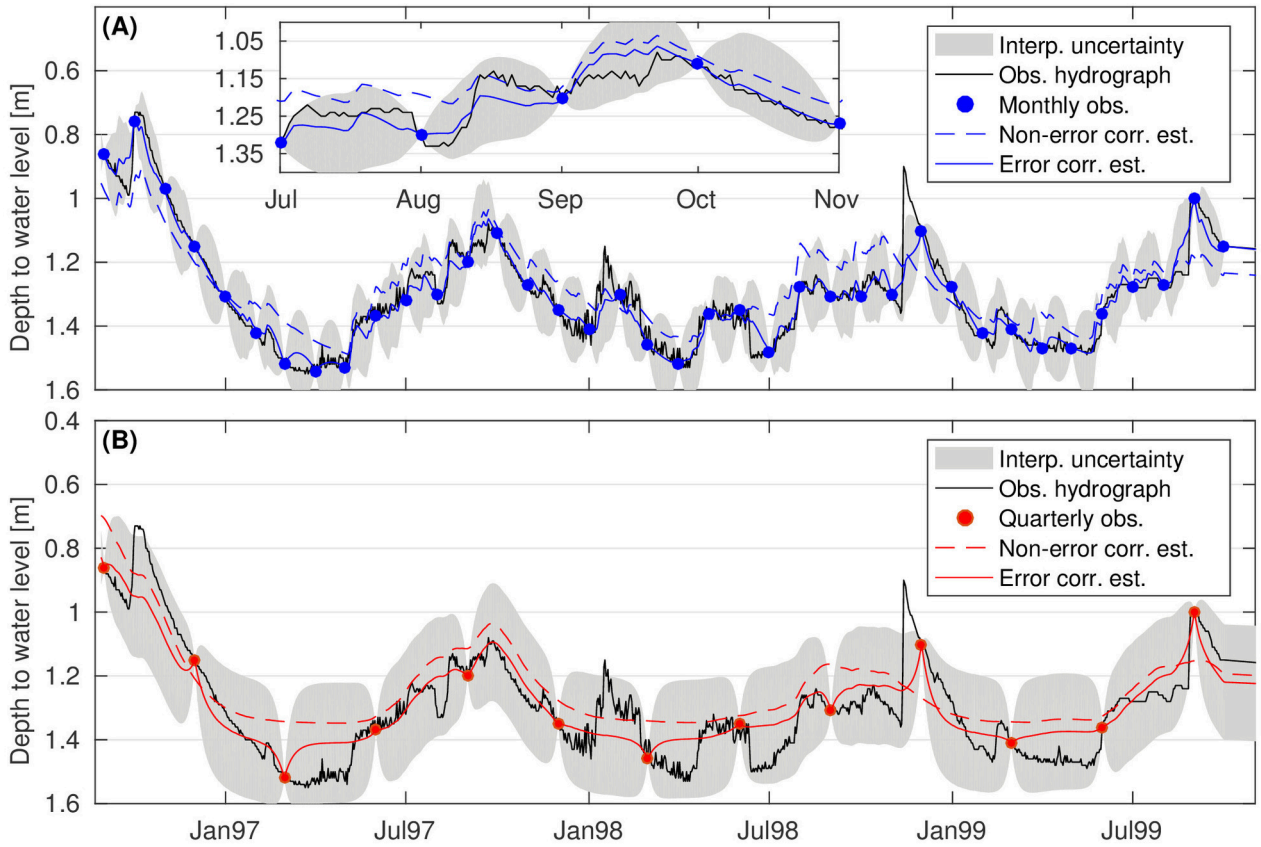
2017wr021838-f05-z-.eps



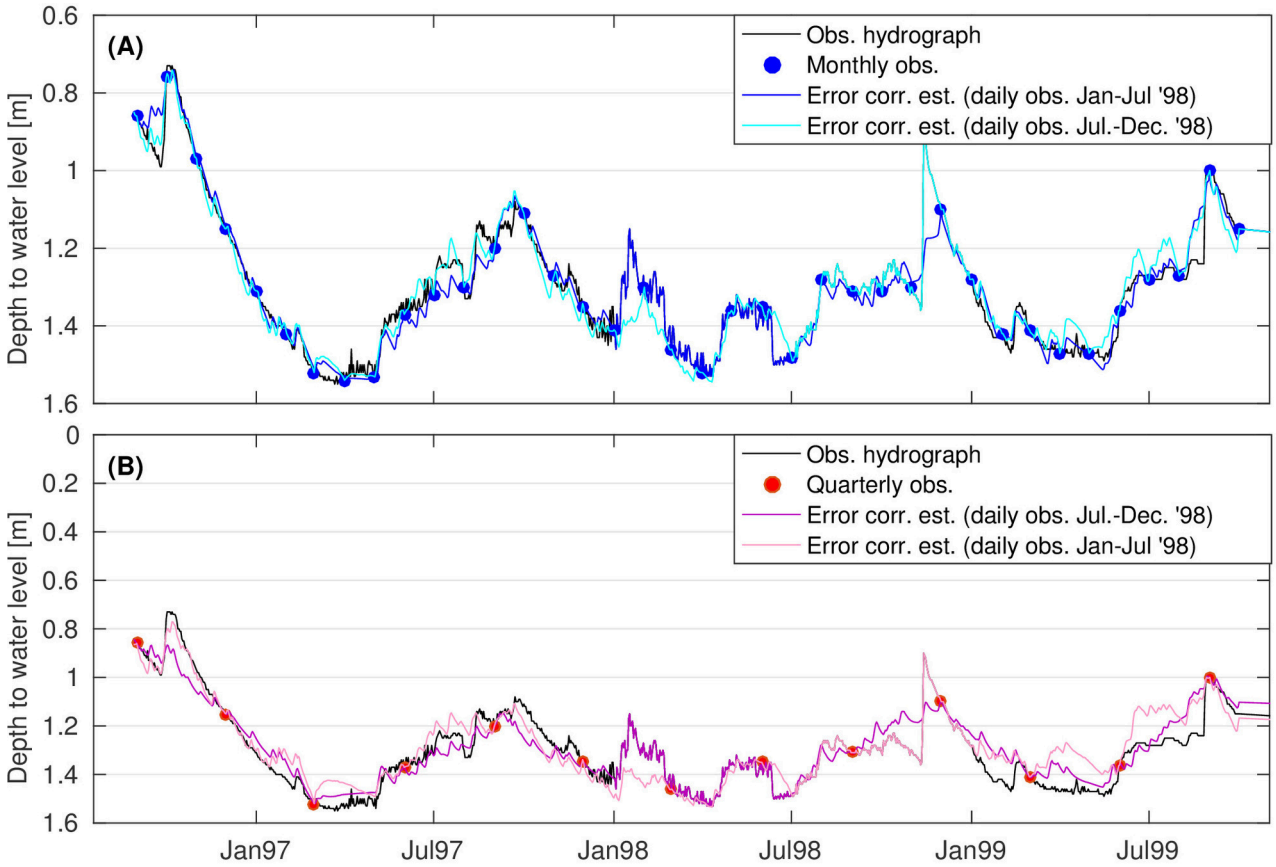
2017wr021838-f06-z-.eps



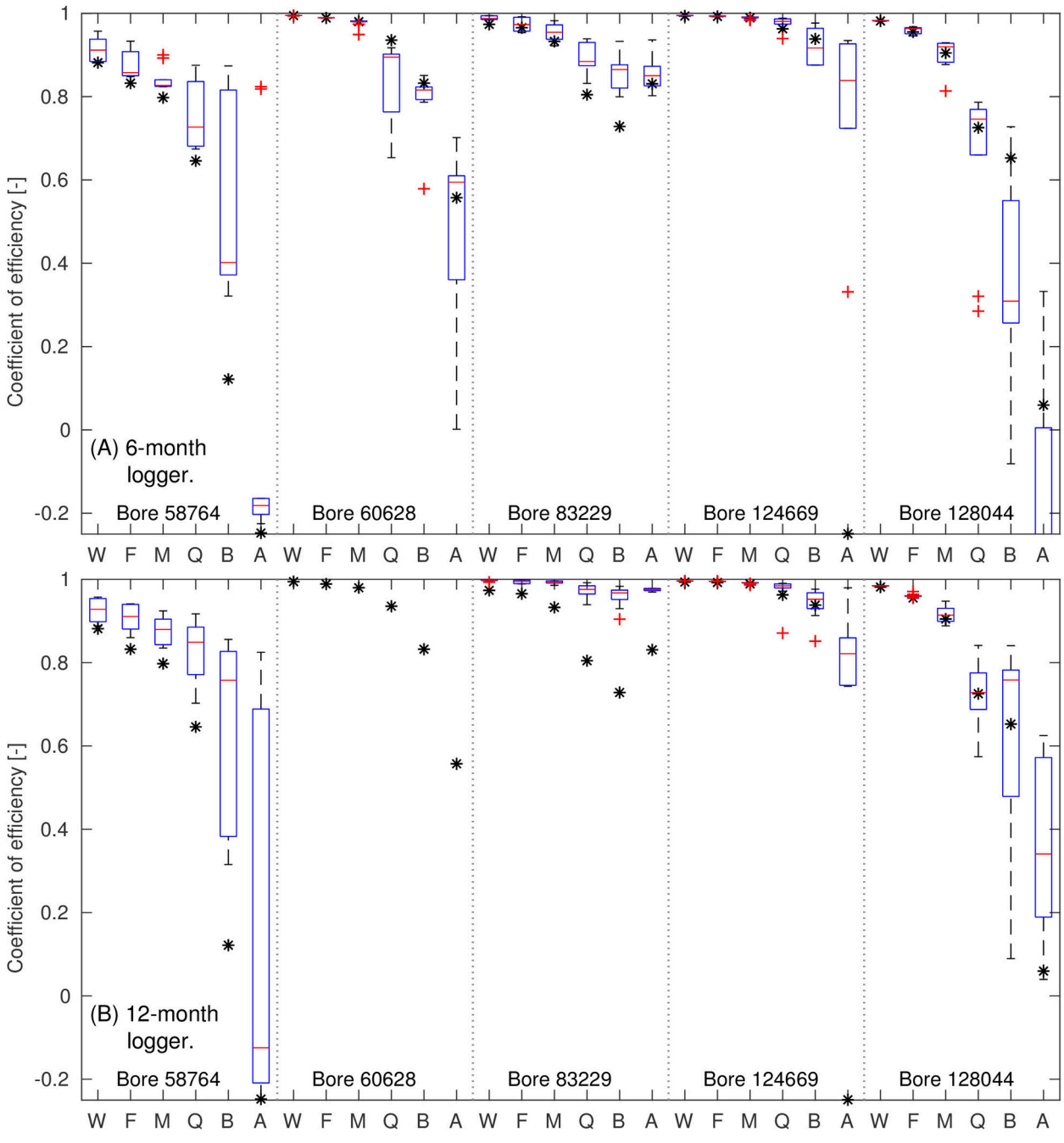
2017wr021838-f07-z-bw.eps



2017wr021838-f08-z-.eps



2017wr021838-f09-z-.eps



2017wr021838-f10-z-.eps