Predictability, stationarity, and classification of hydraulic responses to recharge in two karst aquifers

A. J. Long¹ and B. J. Mahler²

¹US Geological Survey, 1608 Mountain View Rd, Rapid City, South Dakota, 57702, USA
²US Geological Survey, 505 Ferguson Lane, Austin, Texas, 78754, USA

Received: 7 August 2012 – Accepted: 8 August 2012 – Published: 16 August 2012

Correspondence to: A. J. Long (ajlong@usgs.gov)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

Karst aquifers, many of which are rapidly filled and depleted, are likely to be highly susceptible to changes in short-term climate variability. Here we explore methods that could be applied to model site-specific hydraulic responses, with the intent of simulating these responses to different climate scenarios from high-resolution climate models. We compare hydraulic responses (spring flow, groundwater level, and stream base flow) at several sites in two karst aquifers: the Edwards aquifer (Texas, USA) and the Madison aquifer (South Dakota, USA). A one-dimensional, lumped-parameter model simulates nonstationary soil moisture changes for estimation of recharge, and a nonstationary convolution model simulates the aquifer response to this recharge. Model fit to data was 4% better for calibration periods than for validation periods. We use metrics that describe the shapes of the impulse-response functions (IRFs) obtained from convolution modeling to make comparisons in the distribution of response times among sites and among aquifers. Combined principal component analysis and cluster analysis of metrics describing the shapes of the IRFs separated those sites with IRFs having a large ratio of the mean response time to the system memory from those with large skewness and kurtosis. Classification of the IRF metrics indicate that there is a range of IRF characteristics for different site types (i.e., spring flow, groundwater level, base flow) within a karst system. Further, similar site types did not necessarily display similar IRFs. Results indicate that the differences existing within either aquifer are larger than the differences between the two aquifers and that the two aquifers are similar according to this classification. The use of multiple metrics to describe the IRFs provides a novel way to characterize and compare the way in which multiple sites respond to recharge. As convolution models are developed for additional aquifers, they could contribute to an IRF database and a general classification system for karst aquifers.
1 Introduction

An understanding of how key hydrologic variables, such as spring flow and groundwater levels, are likely to respond to potential future climate scenarios is critical for water management. Karst aquifers are likely to be particularly susceptible to changes in short-term climate variability because the cavernous porosity of these aquifers allows rapid replenishment by focused recharge through streambeds and sinkholes (White, 1988), the amount and timing of which is tightly linked to precipitation and antecedent moisture conditions (e.g., Long, 2009; Jukić and Denić-Jukić, 2011). The karstic Edwards Balcones Fault Zone aquifer has been identified as being particularly vulnerable to climate-change effects because of high use, strong links to climatic inputs, large variability in precipitation and multi-year droughts, and dependent endangered species (Loáiciga et al., 2000).

High-resolution weather-forecast models have been adapted to simulate regional climate change based on boundary conditions taken from coarser-resolution general circulation models (e.g., Mearns et al., 2009; Hostetler et al., 2011). With regional climate models continually improving, convolution modeling is a promising approach to estimate how hydrologic systems will respond to local-scale climate scenarios. Convolution has been widely used in rainfall-runoff models to simulate streamflow in response to infiltration on a watershed (Singh, 1988; Jakeman and Hornberger, 1993; Jeannin, 2001; Pinault et al., 2001; Simoni et al., 2011). Convolution also has been applied, to a lesser extent, to groundwater response (Hall and Moench, 1972; Long and Dericksen, 1999; Jukić and Denić-Jukić, 2006). Convolution is particularly useful for simulating karst hydrologic systems, which respond rapidly to changes in precipitation but for which site-specific response is difficult if not impossible to simulate with physically based numerical models. Unsteady and nonuniform flow in variably saturated conduits and pressurized flow in filled conduits, as described by Reimann et al. (2011), are spatially and temporally variable conditions that can be simulated by convolution, even if the conduit network is not defined physically. There is interest in classifying different
types of karst aquifers (e.g., Smart and Worthington, 2004), but only a few of the proposed approaches are quantitative, as in Covington et al. (2009, 2012) and Labat et al. (2011). Convolution modeling is well suited to quantification of characteristic aquifer metrics that describe hydraulic response types.

Here we explore methods that could be applied to model site-specific hydraulic responses with the intent of application to projected climate simulations from high-resolution dynamical models, as these models continually improve. We compare hydraulic response characteristics at several sites in two karst aquifers: the Edwards aquifer (Texas, USA) and the Madison aquifer (South Dakota, USA). We describe the application of a one-dimensional, lumped-parameter model on a site-specific basis and quantification of the predictive accuracy of model output. We examine and quantify temporal nonstationarity, a condition common in karst systems especially when epiphreatic conduits exist (Jeannin, 2001). The model simulates nonstationary soil moisture changes for estimation of recharge and uses a nonstationary convolution process to simulate the aquifer response to this recharge. We used metrics that describe the shapes of the IRFs obtained from convolution modeling to make comparisons between sites and between two aquifers in a novel approach that could be useful for karst aquifer classification.

2 Methods

2.1 Estimating recharge

Daily recharge is estimated by first determining a daily soil moisture index $s$, which is weighted by a backward-in-time exponential decay function that operates on the past daily rainfall record, as described by Jakeman and Hornberger (1993):
\[ s_i = cr_i + \left(1 - \kappa_i^{-1}\right) s_{i-1} \]  
\[ = c \left[ r_i + \left(1 - \kappa_i^{-1}\right) s_{i-1} + \left(1 - \kappa_i^{-1}\right)^2 s_{i-2} + \ldots \right] \quad i = 0, 1, \ldots, N \quad 0 > s > 1, \]  
\[ \text{where } c \text{ is a normalizing parameter that limits } s \text{ to values between 0 and 1 (dimensionless), } \kappa \text{ adjusts the influence of antecedent conditions and is related to evapotranspiration (dimensionless), } r \text{ is total daily rainfall (cm), and } i \text{ is the time step (days). Air temperature, which influences evapotranspiration rates, is accounted for by adjusting } \kappa \text{ by daily air temperature (Jakeman and Hornberger, 1993):} \]

\[ \kappa_i = \alpha \exp[(20 - T_i)f] \quad f > 0, \]  
\[ \text{where } \alpha \text{ is scaling coefficient (dimensionless), } T \text{ is air temperature (°C), and } f \text{ is a temperature modulation factor (°C}^{-1}). \text{ Equation (3) has the primary effect of increasing the value of } s \text{ during cool periods } (0 < T < 20°\text{C}) \text{ when evapotranspiration is low. Daily recharge, or effective precipitation, } u_i \text{ (cm) is then calculated by} \]

\[ u_i = r_is_i. \]  

Only precipitation that occurs either as rain or melting snow was included in calculation of \( s \). A method was established to estimate the occurrence of snow precipitation and melting for future periods on the basis of simulated air temperature. To determine the form of precipitation for each day, an air temperature threshold value \( T_s = 0 \) was set, below which precipitation was assumed to occur as snow. To determine days when melting occurs, a melting threshold value \( T_m \) was estimated. If daily snow-depth data are available, this melting threshold can be determined empirically as the average air temperature for days when snow depth decreases to zero from a previous day with a snow depth greater than zero. Sublimation was accounted for by estimating the fraction of snow moisture remaining after sublimation \( (S_i; \text{Long, 2009). Snow precipitation} \)
was multiplied by $S_f$ and summed for the continuous series of snow-precipitation days prior to each snowmelt day. This sum was added to each snowmelt day in the daily rainfall record because snowmelt was assumed to have the same effect as rainfall on the value of $s$.

### 2.2 Convolution

If a system input, after being transported through a medium, results in a system output that is dispersed in time according to a characteristic waveform, then this system can be simulated by convolution. In karst settings, hydrologic responses that are suitable for simulation by convolution are hydraulic head (e.g., water level in a well) and flow (e.g., from a spring), which were simulated on a daily time step.

Convolution is described by

$$y(t) = \int_0^\infty h(t - \tau)u(\tau)d\tau,$$

(5)

where $y(t)$ is the system response function; $u(\tau)$ is the system input, or forcing function; $h(t - \tau)$ is an impulse-response function (IRF); and $t$ and $\tau$ are time variables, where $t - \tau$ represents the delay time from signal to response (Singh, 1988; Olsthoorn, 2008).

For uniform time steps, the discrete form of Eq. (5) is

$$y_i = \sum_{j=0}^{i} h_{i-j}u_j \quad i = 0, 1, \ldots, N.$$

(6)

The IRF describes the system response function $y$ that results from an instantaneous unit input $u_j$. The length of the IRF quantifies the system memory, or time that the response to the impulse persists. IRFs were approximated by exponential or lognormal curves or a combination of the two. The exponential curve is defined as

$$h(t) = a\lambda \exp(-\lambda t),$$

(7)
where $a$ is a scaling coefficient, and $\lambda$ determines the mean and variance of the system response time $t$ as

$$\mu = \lambda^{-1}$$

(8)

and

$$\sigma^2 = \lambda^{-4},$$

(9)

respectively. The lognormal curve is defined as

$$h(t) = \frac{b}{t \sqrt{2\pi \varepsilon}} \exp \left[ -\frac{1}{2\varepsilon} (\ln t - \omega)^2 \right],$$

(10)

where $b$ is a scaling coefficient, and $\omega$ and $\varepsilon$ determine the mean and variance of $t$ as

$$\mu = \exp (\omega + \varepsilon/2)$$

(11)

and

$$\sigma^2 = \mu^2 [\exp(\varepsilon) - 1],$$

(12)

respectively. Parameters of Eqs. (1)–(12) are listed in Table 1. Because exponential and lognormal curves are asymptotic and thus have infinite length with infinitesimal magnitude after some point in time, system memory is defined herein as time $t_m$ on the IRF time scale at which 95% of the area of the curve is between time $t = 0$ and $t_m$.

Modeling of flow in karst settings is complicated by the existence of quick-flow and slow-flow components; e.g., flow through large conduits as well as small fractures (Pinault et al., 2001). In some cases, a single exponential or lognormal IRF can adequately represent the quick-flow and slow-flow components of karst groundwater flow, where the first part of the curve, including the initial peak, represents quick flow, and the tail of the curve represents slow flow. In other cases, a secondary IRF with a long tail
that represents all or part of the slow-flow component may be useful (e.g., Long, 2009). The resulting compound IRF is the superposition of the primary and secondary IRFs. Several authors have used compound IRFs to represent the quick-flow and slow-flow components in karst settings (e.g., Jakeman et al., 1993; Pinault et al., 2001; Denić-Jukić and Jukić, 2003). In the approach herein, the compound IRF can consist of any combination of two IRFs, either exponential or lognormal (Fig. 1). A compound IRF consisting of an exponential and lognormal curve can result in a bimodal, or double-peak, distribution of response times, which could be a result of two-domain flow (Fig. 1b, Long and Putman, 2006). A compound IRF consisting of two exponential curves is useful when quick flow and slow flow are separated into a sharp initial peak and long tail, respectively (Fig. 1d).

If the scaling coefficients \( a \) and \( b \) are set to unity, the areas under the IRFs also are equal to unity. Adjusting the values of these coefficients allows their use as conversion factors to account for the different dimensions between the forcing function and response function and also allows for unequal response magnitudes between the primary and secondary IRFs. For simulation of groundwater levels, a datum \( h_0 \) at which hydraulic head equals zero must be established. Conceptually, this is the level to which hydraulic head would converge if the local system-input recharge was eliminated. This system input is assumed to be the only source of recharge close enough to cause hydraulic-head fluctuation and the only source that causes hydraulic head to rise above \( h_0 \).

### 2.3 Nonstationarity

Antecedent soil moisture effects and complex geological and morphological structures result in nonstationarity in watershed flow processes (Simoni et al., 2011), which are those whose parameters change when shifted in time. For application to karst watersheds, Pinault et al. (2001) used nonstationary IRFs that were assumed to increase and decrease with hydraulic head. Jukić and Denić-Jukić (2006) used nonstationary IRFs that varied according to antecedent recharge conditions.
In this model, nonstationary response characteristics are simulated at two different stages in the overall process. The first stage is the infiltration of precipitation to become recharge. Recharge rate is determined by $s$ (Eqs. 1–4), which varies on a daily time scale resulting in nonstationarity at this stage. The second stage is the transformation of recharge to a hydraulic response, which is simulated by convolution (Eqs. 5–12). The system response characteristics as described by the IRF are affected by changes in precipitation rates on a scale of years to decades. It therefore was useful to separate the precipitation record into wet and dry periods of 1 yr or greater, as determined on the basis of the annual cumulative departure from the long-term mean precipitation (CDMP). Periods with positive slopes in the CDMP occur when precipitation is consistently above the mean and thus indicate wet periods; negative slopes indicate dry periods. To account for nonstationarity at the second stage, the shape parameters and scaling coefficients of the IRFs were estimated separately for wet and dry periods, with the assumption that the curve types do not change temporally within these periods. This is a method not previously used and has advantages for aquifer classification because wet-period and dry-period metrics are distinct and clearly defined. As many as four IRFs were used to simulate the system response during model calibration: primary IRFs (IRF$_{w1}$ and IRF$_{d1}$) and secondary IRFs (IRF$_{w2}$ and IRF$_{d2}$), where the subscripts, $w$ and $d$, refer to wet and dry periods, respectively.

The IRF curve areas can be different for wet and dry periods and primarily are meaningful in their relative magnitudes for comparison of the responses of these two periods, which may provide insight into system operation. For example, if a wet-period IRF has twice the area of a dry-period IRF, then the wet-period response is twice that of the dry period for the same amount of recharge.

### 2.4 Classification of site responses

Hydraulic-response characteristics can differ among sites, which can represent different flow systems within an aquifer; i.e., different networks of conduits, fractures, and other pores spaces. If site responses are classified for multiple aquifers, then
comparisons can be made among the aquifers. The IRF is suitable for comparison of sites and aquifers in climatically different locations because this comparison does not depend on differences in rainfall frequency, variability, or intensity between locations.

We used metrics, which can be quantified for any parametric or nonparametric IRF, to quantify several characteristics of the IRF shapes (Table 2). Metrics were selected that quantify the IRF shape independently from scale so that comparisons were not weighted by the overall IRF magnitude, which might vary for climatically different locations. To define these metrics, the IRF was assumed to be a frequency distribution of the transit times of the response quantity, either hydraulic head or flow, and basic statistics and other parameters of this distribution were computed. Ratios were used for many of the metrics because these are scale independent. Metrics include skewness (skw), kurtosis (krt), and five ratios: standard-deviation : mean (SDMn), mean : memory (MnMm), mode : memory (MdMm), standard-deviation : memory (SDMm), and peak-height : area (PHA; for bimodal distributions the highest peak is used). These seven metrics were quantified for wet and dry periods separately (Table 2), resulting in 14 metrics. For stationary systems, wet and dry metrics are equal. Finally, the wet : dry area ratio (WDA) was included, which is the ratio of the wet-period to dry-period IRF area (WDA=1 for stationary systems) and results in a total of 15 metrics (Table 2).

Principal component analysis (PCA) and cluster analysis were used to assess similarities, differences, and groupings of sites on the basis of IRF shape, as described by the 15 metrics. PCA is a linear transformation of data in multidimensional space, where the transformed axes, or principal components, align with the greatest variances in the multivariate dataset (Davis, 2002). Each principal component is a new variable that is a linear combination of all the original variables. PCA is helpful for elucidating patterns that would otherwise be obscured in attempting to assess a large number of metrics. Cluster analysis groups sites and metrics according to similar characteristics. The software MATLAB (http://www.mathworks.com) was used for PCA and cluster analysis.
2.5 Modeling considerations

Convolution models provide a convenient way to assess system memory, which is an important consideration in any hydrologic model. At a minimum, input corresponding to a time period equal to the system memory is required prior to the start of a calibration period. This initial time period is referred to as model spin-up, the output from which is considered to be of questionable validity.

The length of the calibration and validation periods also must be considered in light of the system memory. There is less confidence in the predictive strength of a model if the observed record is shorter than the system memory than if it is longer. Ideally, the validation period alone should be longer than the system memory, and if it is several times longer than the system memory, the full range of response times is tested several times over.

The use of secondary IRFs, the choice of curve type, and stationarity were evaluated and selected for each site on the basis of model fit for the validation period. Inclusion of secondary IRFs and simulating a system as nonstationary increase model parameterization and complexity over the simpler options. In some cases the simplest model resulted in a better fit for the validation period than one with added complexity. Although model fit might improve for the calibration period as parameters are added, if the validation period indicates that this added complexity is not helpful, the model is overparameterized. If the overall model fit is poor, this model might not be appropriate, or the input data might not represent the recharge area well.

3 Model application and results

Hydraulic responses at several sites with data for spring flow, groundwater base flow in streams, groundwater level, or cave drip (Table 3) were simulated for two karst aquifers: the Edwards aquifer in south-central Texas and the Madison aquifer in western South Dakota (Fig. 2). Weather station data for precipitation and air temperature
were obtained from the National Climatic Data Center (2012). Generally, the recharge areas for the sites simulated are small, and data from a single weather station was used as model input. The weather station either within or nearest the recharge area for each site was used, and the next nearest station was used to fill periods of missing data if necessary. Equations (1), (3), (4), (6), (7), and (10) were programmed in MATLAB (http://www.mathworks.com). The models were calibrated to observed discharge and water levels at gages (Table 3; US Geological Survey, 2012). Model parameters (Table 1) were optimized by using the “lsqcurvefit” function in MATLAB to minimize the differences (residuals) between simulated and observed temperature values. This is a subspace trust-region method and is based on the interior-reflective Newton method (Coleman and Li, 1994, 1996).

### 3.1 Study areas

The Edwards Balcones Fault Zone aquifer (herein, the Edwards aquifer) in South-Central Texas is a well-developed karst aquifer contained within the Edwards Group. Surface recharge to the Edwards aquifer occurs from direct precipitation and sinking streams that cross onto the outcrop area of the Edwards Group from the west and northwest (Fig. 2). The aquifer dips to the south and southeast below the land surface. Groundwater flow generally is to the east and northeast, and discharge occurs at several large springs. The hydrogeology of the Edwards aquifer is described in detail in Maclay and Small (1983), Small et al. (1996), and Lindgren et al. (2004).

The Madison aquifer in Western South Dakota is a well-developed karst aquifer composed of limestone and dolostone (Greene and Rahn, 1995). It is contained within the regionally extensive Madison Limestone of Mississippian age, referred to locally as the Pahasapa Limestone. This formation is exposed at the land surface on all flanks of the Black Hills and dips radially outward in all directions below the land surface (Fig. 2); the outcrop of the Madison Limestone is the recharge area for the Madison aquifer. The hydrogeology of the Madison aquifer in the Black Hills area is described in detail in Driscoll and Carter (2001).
3.2 Sites simulated

For the Edwards aquifer, water level in six wells and flow from two large springs were simulated (Fig. 2, Table 3). For the Madison aquifer, three wells, one spring, one spring complex, one cave water body, and stream base flow for two Madison Limestone watersheds were simulated (Fig. 2, Table 3). In addition, cave-drip at two sites and streamflow in one fractured-rock watershed located in the Madison study area were simulated for comparison to the aquifer sites (Fig. 2, Table 3). The observed and simulated spring flow for Barton Springs is shown as an example of simulation results (Fig. 3).

Daily precipitation and air temperature were used as model input for all sites, except for the Reptile Gardens well (site RG; Fig. 2, Table 3). This site is located in an area where the primary recharge source is sinking streamflow, which was estimated as described in Hortness and Driscoll (1998) on a daily time step for streamflow gage 06407500 (Fig. 2) and used as model input.

Hydrologic response data for all of the Edwards aquifer sites consisted of direct water-level or spring-flow measurements, but additional description or data manipulation was required for some of the Madison aquifer sites, e.g., because of indirect spring-flow measurements. Observed streamflow at Little Spearfish Creek and Spearfish Creek (sites LScr and SPFcr) was used as an estimate of groundwater discharge, which is 97% and 86%, respectively, of total streamflow (Driscoll and Carter, 2001). Overland runoff in these watersheds rarely occurs because of the highly porous karst terrain (Miller and Driscoll, 1998). Groundwater input to streamflow in the Fall River (site FALr) is about 96%, primarily flowing from a complex of artesian springs (Rahn and Gries, 1973; Back et al., 1983; Driscoll and Carter, 2001). A 1-yr moving average of streamflow was used as a surrogate for total observed spring flow to remove anthropogenic variability resulting from municipal water use and wastewater discharge. Recharge from Beaver Creek (site BEVcr), a sinking stream, was used as a source of recharge for simulation of Windy City Lake (site WCL) in addition to precipitation, and therefore, this model included an additional IRF for streamflow input. Simulated
streamflow at a daily time step was used as model input for the period prior to 1991 because data were not available for that period.

3.3 Calibration and validation of models

The models were validated by (1) calibrating each model to part of the record for the system response, (2) executing the model under those conditions for the remaining period (validation period), and (3) examining the similarity of the simulated and observed system responses for the validation period. Calibration periods ranged from 0.9 to 70 yr and validation periods ranged from 0 to 23.8 yr (Table 4). Model parameters were adjusted by trial and error until the simulated hydrographs were similar to the observed hydrographs. Parameter optimization then was executed for the calibration periods, and the validation periods were evaluated on the basis of the Nash criterion (Nash and Sutcliffe, 1970), which is a measure of the fit between simulated and observed time-series records, hereafter referred to as model fit. The Nash criterion is defined as

\[
C_{Nash} = 1 - \frac{\sum (y_{obs} - y_{sim})^2}{\sum (y_{obs} - y_{mean})^2},
\]

where \(y_{obs}\) and \(y_{sim}\) are daily time-series of the observed and simulated responses, respectively, and \(y_{mean}\) is the mean of \(y_{obs}\). The Nash criterion varies from 0 (poorest fit) to 1.0 (best fit) and is a comparison of the magnitude of residuals (numerator) to the overall amplitude of fluctuation in the observed record (denominator).

The Nash criterion was calculated for the calibration and validation periods separately (\(C_{Nash-c}\) and \(C_{Nash-v}\); Table 4). For comparison of residuals for different periods, the denominator of the second term in Eq. (13) must be consistent across all cases. Therefore, the denominator was calculated on the basis of the full period, and the numerator was calculated on the basis of the period of interest only. This method provides a direct comparison of residuals for different periods, even if these periods have different fluctuation amplitudes, and thus is robust for comparing short periods, where...
fluctuation amplitudes might be small in comparison to the overall record. Because parameters were optimized for the calibration periods only, $C_{Nash-c}$ values were 0.03 higher (4%) on average than $C_{Nash-v}$ values.

Primary and secondary IRFs were included in the initial model calibration for all sites. Trial and error was used to determine whether to use exponential or lognormal IRFs (Table 4). Multiple trials of the calibration process indicated whether or not secondary IRFs were necessary. For example, initial calibration sometimes resulted in a small or negligible curve area for a secondary IRF. In these cases, the minimized IRF was omitted if it did not degrade the $C_{Nash-v}$ value. In some cases, omitting secondary IRFs resulted in increased $C_{Nash-v}$ values, which indicated that adding secondary IRFs resulted in overparameterization. For example, using all four IRFs for the Medina FM1796 well resulted in a $C_{Nash-v}$ value of 0.81, but using only two IRFs resulted in a higher $C_{Nash-v}$ value (0.91; Table 4), even though the two cases fit the calibration period equally well.

All models were assumed to be nonstationary for initial calibration trials. Similar shapes for the wet-period and dry-period IRFs indicated that the system might be stationary. Stationarity for these cases was tested by using the same IRF for wet and dry periods. If $C_{Nash-v}$ decreased by less than 0.02 as a result, then parsimony was preferred, and the system was simulated as stationary. In some cases, model fit for the validation period indicated that a stationary model was best. For example, $C_{Nash-v}$ for the Lovelady well model was 0.69 for the nonstationary case and 0.85 (Table 4) for the stationary case, which indicated that the nonstationary model was overparameterized.

### 3.4 Assessing model fit and predictive strength

The length of the validation period is an indicator of the length of time for which a model can be projected with confidence into the future on the basis of climate-model predictions. The value of $C_{Nash-v}$ is an indicator of the expected model error for the future period, but does not account for error in the output from a climate model that is used as input for the hydraulic-response model. A goal here was to have the longest validation periods possible, while maintaining a minimum, or target, $C_{Nash-v}$. In general,
\( C_{\text{Nash-v}} \) increases with an increasing calibration period, which decreases the length of the validation period. This inverse relation between \( C_{\text{Nash-v}} \) and validation period length results in a compromise that must be considered when setting a target \( C_{\text{Nash-v}} \). Also, for nonstationary systems, the calibration periods needed to include both wet and dry periods. Therefore, the length of the validation period was limited in some cases by the occurrence of wet and dry periods for a particular system.

A target \( C_{\text{Nash-v}} \) value was set at 0.70. Three of the sites had very low \( C_{\text{Nash-v}} \) values (<0.5) for any validation-period length tested, and thus this period was set to zero with no \( C_{\text{Nash-v}} \) value listed (Table 4). Three other sites had \( C_{\text{Nash-v}} \) values that were less than the target (Table 4). These sites also had \( C_{\text{Nash}} \) (full period) values that were less than the target, and thus achieving the target \( C_{\text{Nash-v}} \) was not possible for any validation period for these sites. The \( C_{\text{Nash-v}} \) value for the remaining sites ranged from 0.70 to 0.92 (Table 4).

An additional measure of predictive strength is the memory ratio, defined as the ratio of the system memory to the length of the period of record for the system response. A value of less than unity for the memory ratio is desirable because if the system memory is longer than the data record, then the IRF has a partial tail segment that has no effect on the model output for the period of record. This means that this tail segment has not been calibrated to data, and model predictions are less certain than if the entire IRF were calibrated. Two sites, Windy City Lake and the Medina FM1796 well, had memory ratios greater than unity (Table 4); a memory ratio greater than unity is common for systems such as these with long memories.

Most of the sites were simulated as nonstationary (Table 4), and model spin-up periods were at least as long as system memories for all sites. For Windy City Lake, Medina FM1796 well, and Fall River, data were not available for the first part of the spin-up period. Mean values of precipitation and temperature, calculated from the periods of record, were used as surrogates for the missing data for these three sites. The lengths of these periods in years were 56 (1856–1911), 61 (1840–1900), and 24 (1887–1910) for these three sites, respectively. The effects of using mean values were tested
by replacing the mean values with the subsequent period of data, which changed the $C_{Nash-v}$ values by less than 0.02 for these three sites. This indicates model insensitivity to precipitation and temperature variability in the early parts of the spin-up periods.

### 3.5 Classification of site responses

The eight Edwards aquifer sites and eight Madison aquifer sites were included in PCA and cluster analysis. Characteristics of the IRF shapes were quantified by 15 metrics (Table 2). The values were log transformed and standardized to a mean of zero and standard deviation of unity before applying PCA. Principal components 1 through 4 (PC1–PC4) explain 50, 18, 14, and 10%, respectively (92% total) of the variance in the dataset and represent the primary characteristics quantified by the IRF metrics. The relation of the sites (also known as scores) to the metrics (also known as loadings) were plotted in principal component space (Fig. 4).

Cluster analysis was used to determine similarities between sites and to characterize site clusters according to the metrics (Fig. 4). First, a cluster analysis was applied to determine site clusters, in which the score values from PCA were used as input for the cluster analysis, as described by Suk and Lee (1999). This analysis iteratively assigns each site to a cluster by minimizing the sum of Euclidian distances in principal component space between sites (score values) and the nearest cluster centroid (Seber, 1984; Spath, 1985). A second cluster analysis was applied to the log-transformed and standardized metric values, with three clusters specified for each analysis. The site clusters primarily are separated on the PC1 axis but occupy similar ranges on the PC2 and PC3 axes, which indicates that PC1 represents the differentiation of the site clusters (Fig. 4). The distribution of the plotting positions of sites results from differences in IRF shapes, and the plotting positions of the metrics in relation to the sites reveal specific IRF characteristics that are represented by each principal component.

A comparison of IRF shapes for different clusters illustrates relations between sites and metrics (Fig. 5). Site cluster $A_s$ is typified primarily by a lognormal curve with a delayed peak response, generally combined with an exponential curve that sometimes...
results in a bimodal distribution, and, on average, 68% of the IRF curve area is log-normal (Fig. 5). Site cluster $B_s$ is typified primarily by an immediate peak response (exponential curve sometimes combined with lognormal), and, on average, 71% of the IRF curve area is exponential (Fig. 5). Site cluster $C_s$ contains only one site (HCV), and the response is 100% lognormal, with less delay in the lognormal peak than those of site cluster $A_s$ (Fig. 5).

The plotting positions of the metrics help explain the meaning of the first three principle components (Fig. 4). The three metric clusters each occupy a different range on the PC1 axis and primarily are separated on that axis; on the PC2 and PC3 axes, however, metric clusters $A_m$ and $C_m$ occupy largely coincident ranges (Fig. 4). Metric cluster $B_m$ is widely separated from the other two on the PC3 axis (Fig. 4). Metrics that are most heavily loaded on PC1 are the mean : memory ratios ($MnMm$; negative) and all metrics in cluster $C_m$ (positive). Thus, these metrics have the largest role in differentiating sites along the PC1 axis and range from sites that have large mean : memory ratios on the left side of Fig. 4 to sites with large skewness and kurtosis (skw, krt) on the right side. The four metrics for skewness and kurtosis (cluster $C_m$) plot close to one another (Fig. 4a), which indicates positive correlation of these metrics because highly skewed exponential and lognormal curves generally have high kurtosis as well. The other four metrics in cluster $C_m$ (standard-deviation : memory (SDMm), standard-deviation : mean (SDMn)) plot in a similar range on PC1 to skewness and kurtosis but in a different range on PC2 (Fig. 4a); thus PC1 makes little distinction between metrics in cluster $C_m$, whereas PC2 describes these differences. The most heavily loaded metrics on PC2 are the mode : memory ratios ($MdMm$) and kurtosis (Fig. 4a). The distribution of sites along the PC2 axis range from those with high mode : memory ratios and kurtosis in the upper part to those with high standard-deviation : mean and peak-height : area ratios (PHA) in the lower part (Fig. 4a). Kurtosis is heavily loaded on PC1 and PC2.

The peak-height : area ratios, which have small loadings on the PC1 and PC2 axes, are heavily loaded on PC3 (Fig. 4b), which indicates that PC3 primarily represents
these metrics. The distribution of sites along the PC3 axis range from those with high peak-height : area ratios on the upper side to those with high wet : dry area ratios (WDA) on the lower side (Fig. 4b). For PC3, the manner in which metrics in cluster \( C_m \) are paired by proximity suggests that PC3 differentiates wet- and dry-period IRFs; i.e., some wet-period metrics and dry-period metrics are paired on PC3 (skw-w and krt-w, skw-d and krt-d, SDMn-w and SDMm-w, SDMn-d and SDMm-d; Fig. 4b). This pairing also indicates a separation between the wet and dry counterparts of these metrics; e.g., skw-w and kurt-w are paired, but skw-w and skw-d are separated (Fig. 4b). Further, the wet and dry counterparts of the mode : memory ratio plot farthest apart on PC3 (Fig. 4).

The wet : dry area ratio is the most heavily loaded metric on PC4 (not shown), which indicates that PC4 primarily describes this metric. Therefore, PC4 also represents differences between wet and dry periods, except that PC4 is associated with differences in IRF area rather than shape.

Six of the 16 Edwards and Madison aquifer sites (38 \%) are stationary (Table 4), which caused the wet and dry metric counterparts to plot more closely on Fig. 4 than if all sites were nonstationary. Positively correlated metrics are easily identified by proximity of plotting position, whereas negatively correlated metrics can be identified as those that plot on opposite sides of the origin; e.g., MnMm and SDMn (Fig. 4a).

4 Conclusions

The convolution models are well suited to link climate scenarios to hydrologic responses critical to human water supply and ecosystems. As climate models improve in simulating storm processes, projected precipitation and air-temperature records can be used as input for hydrologic models to investigate how water levels and spring discharge might change. The output from the hydrologic models can, in turn, serve as input for models evaluating vulnerability of karst ecosystems to potential changes in water level and spring discharge. Unlike most physically based models
(e.g., three-dimensional finite difference), convolution models are calibrated to data at short time steps that are well suited to short-term variability characteristic of karst groundwater responses.

The use of multiple metrics to describe the IRFs provides a novel way to characterize and compare the way in which multiple sites respond to recharge. Classification of the IRF metrics and evaluation with PCA and cluster analysis indicate that there is a range of IRF characteristics describing different sites within a karst system. The IRF shape is a result of the aquifer’s pore geometry and connectivity, which is spatially heterogeneous in karst aquifers, and thus similar site types do not necessarily display similar IRFs. Knowing which metrics correspond to the different principle components assists in understanding the meaning of these components hydrologically. This understanding also requires an understanding of how each metric affects the shape of the IRF and what the IRF shapes mean hydrologically.

Two of the site clusters contain sites for both aquifers (cluster $C_s$ contains only one site), and site cluster $B_s$, which contains most of the sites, is equally balanced between the two aquifers. This indicates that the differences that exist within either aquifer are larger than the differences between the two aquifers and that these two aquifers are similar according to this classification. The Madison and Edwards aquifers each have well-developed networks of large conduits, which is consistent with the similarity determined by the classification. Currently, IRFs have been developed for only two karst aquifers, but as convolution models are developed for additional sites in additional aquifers, they could contribute to an IRF database and, moreover, a general classification system for karst aquifers. Of particular interest will be comparison of telogenetic and eogenetic systems, humid and arid systems, and diffuse- and conduit-controlled systems. The availability of long-term records for precipitation, air temperature, water level, and spring flow in many areas would facilitate this effort.
Acknowledgements. The authors would like to thank Johnathan Bumgarner of the US Geological Survey for his helpful review and comments on an early draft of this article. We also thank Jonathan McKaskey, Thomas Sample, and Jennifer Bednar of the US Geological Survey for assistance with obtaining data and creating figures and tables.

References


Hydraulic responses to recharge in two karst aquifers

A. J. Long and B. J. Mahler


Hydraulic responses to recharge in two karst aquifers

A. J. Long and B. J. Mahler


Table 1. Model parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Equation</th>
<th>Estimation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>Soil moisture parameter</td>
<td>1</td>
<td>Optimized</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>Soil moisture parameter</td>
<td>3</td>
<td>Optimized</td>
</tr>
<tr>
<td>$f$</td>
<td>Soil moisture parameter</td>
<td>3</td>
<td>Optimized</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Exponential IRF shape parameter</td>
<td>7</td>
<td>Optimized</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Exponential IRF curve area</td>
<td>7</td>
<td>Optimized</td>
</tr>
<tr>
<td>$\omega$</td>
<td>Lognormal IRF shape parameter</td>
<td>10</td>
<td>Optimized</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>Lognormal IRF shape parameter</td>
<td>10</td>
<td>Optimized</td>
</tr>
<tr>
<td>$b$</td>
<td>Lognormal IRF curve area</td>
<td>10</td>
<td>Optimized</td>
</tr>
<tr>
<td>$S_f$</td>
<td>Sublimation fraction</td>
<td>−</td>
<td>Optimized</td>
</tr>
<tr>
<td>$h_0$</td>
<td>Hydraulic-head datum</td>
<td>−</td>
<td>Optimized</td>
</tr>
<tr>
<td>$T_s$</td>
<td>Snow precipitation threshold</td>
<td>−</td>
<td>Assumed 0 °C</td>
</tr>
<tr>
<td>$T_m$</td>
<td>Snowmelt threshold</td>
<td>−</td>
<td>Empirical</td>
</tr>
</tbody>
</table>

* Not applicable.
**Table 2.** Impulse-response function (IRF) metrics. Metrics were quantified for wet and dry periods separately by adding “-w” or “-d”, respectively, to the abbreviations.

<table>
<thead>
<tr>
<th>IRF Metric</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness</td>
<td>skw</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>krt</td>
</tr>
<tr>
<td>SD : mean ratio</td>
<td>SDMn</td>
</tr>
<tr>
<td>SD : memory ratio</td>
<td>SDMm</td>
</tr>
<tr>
<td>Mean : memory ratio</td>
<td>MnMm</td>
</tr>
<tr>
<td>Mode : memory ratio</td>
<td>MdMm</td>
</tr>
<tr>
<td>Peak-height : area ratio</td>
<td>PHA</td>
</tr>
<tr>
<td>Wet : dry area ratio</td>
<td>WDA*</td>
</tr>
</tbody>
</table>

* Not defined separately for wet and dry periods.
### Table 3. Sites simulated

<table>
<thead>
<tr>
<th>Site label (Fig. 2)</th>
<th>Site name</th>
<th>USGS site number(^a)</th>
<th>Site type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Edwards aquifer sites</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FM1796</td>
<td>Medina FM1796</td>
<td>292618099165901</td>
<td>Well</td>
</tr>
<tr>
<td>Bxr</td>
<td>Bexar Co. Index well</td>
<td>292845098255401</td>
<td>Well</td>
</tr>
<tr>
<td>HCV</td>
<td>Hill Country Village</td>
<td>293522098291201</td>
<td>Well</td>
</tr>
<tr>
<td>Bud</td>
<td>Buda Well</td>
<td>300510097504001</td>
<td>Well</td>
</tr>
<tr>
<td>Dow</td>
<td>Dowell Well</td>
<td>300835097483401</td>
<td>Well</td>
</tr>
<tr>
<td>BARsp</td>
<td>Barton Springs</td>
<td>8155500</td>
<td>Spring</td>
</tr>
<tr>
<td>LVL</td>
<td>Lovelady</td>
<td>8159000</td>
<td>Well</td>
</tr>
<tr>
<td>COMsp</td>
<td>Comal Spring</td>
<td>8168710</td>
<td>Spring</td>
</tr>
<tr>
<td><strong>Madison aquifer sites</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FALr</td>
<td>Fall River</td>
<td>06402000</td>
<td>Spring complex</td>
</tr>
<tr>
<td>RFsp</td>
<td>Rhoads Fork Spring</td>
<td>06408700</td>
<td>Spring</td>
</tr>
<tr>
<td>LScr</td>
<td>Little Spearfish Creek</td>
<td>06430850</td>
<td>Base flow</td>
</tr>
<tr>
<td>SPFcr</td>
<td>Spearfish Creek</td>
<td>06430900</td>
<td>Base flow</td>
</tr>
<tr>
<td>WCL</td>
<td>Windy City Lake</td>
<td>433302103281501(^b)</td>
<td>Cave water body</td>
</tr>
<tr>
<td>RG</td>
<td>Reptile Gardens well(^c)</td>
<td>435916103161801</td>
<td>Well</td>
</tr>
<tr>
<td>Tilf</td>
<td>Tilford well</td>
<td>441759103261202</td>
<td>Well</td>
</tr>
<tr>
<td>LA88C</td>
<td>Spearfish GC well</td>
<td>442854103505602</td>
<td>Well</td>
</tr>
<tr>
<td><strong>Cave drip</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTD</td>
<td>Caving Tour drip</td>
<td>433302103281508(^d)</td>
<td>Cave drip</td>
</tr>
<tr>
<td>RmDr</td>
<td>Room Draculum drip</td>
<td>433302103281509(^d)</td>
<td>Cave drip</td>
</tr>
<tr>
<td><strong>Fractured-rock watershed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEVcr</td>
<td>Beaver Creek</td>
<td>06402430</td>
<td>Watershed</td>
</tr>
</tbody>
</table>


\(^b\) Partial water-level record was estimated from water levels in well Md7-11 (Fig. 2) with data from the South Dakota Department of the Environment and Natural Resources in Pierre, South Dakota ($R^2 = 0.99$ for correlation between the two sites; Davis, 2002).

\(^c\) Streamflow at USGS gage 06407500 (US Geological Survey, 2012) was used to estimate recharge.

\(^d\) Data provided by Wind Cave National Park, written communication, 2012.
Table 4. Summary of modeling results.

<table>
<thead>
<tr>
<th>Site label</th>
<th>Calibration period, in years</th>
<th>Validation period, in years</th>
<th>$C_{\text{Nash-c}}$ (calibration period)$^a$</th>
<th>$C_{\text{Nash-v}}$ (validation period)$^a$</th>
<th>IRF$_{\text{w1}}$</th>
<th>IRF$_{\text{w2}}$</th>
<th>IRF$_d$</th>
<th>Stationary model</th>
<th>Memory ratio$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edwards aquifer sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HCV</td>
<td>32.1</td>
<td>15.7</td>
<td>0.85</td>
<td>0.70</td>
<td>Logn</td>
<td>–</td>
<td>Logn</td>
<td>Yes</td>
<td>0.25</td>
</tr>
<tr>
<td>LVL</td>
<td>4.2</td>
<td>10.4</td>
<td>0.90</td>
<td>0.85</td>
<td>Logn</td>
<td>–</td>
<td>Logn</td>
<td>Yes</td>
<td>0.12</td>
</tr>
<tr>
<td>FM1796</td>
<td>12.0</td>
<td>23.8$^c$</td>
<td>0.97</td>
<td>0.91</td>
<td>Logn</td>
<td>–</td>
<td>Exp</td>
<td>No</td>
<td>2.4</td>
</tr>
<tr>
<td>Bxr</td>
<td>20.0</td>
<td>21.0</td>
<td>0.69</td>
<td>0.53</td>
<td>Exp</td>
<td>Logn</td>
<td>Exp</td>
<td>Logn</td>
<td>Yes</td>
</tr>
<tr>
<td>Dow</td>
<td>3.2</td>
<td>10.7</td>
<td>0.82</td>
<td>0.71</td>
<td>Exp</td>
<td>–</td>
<td>Exp</td>
<td>–</td>
<td>No</td>
</tr>
<tr>
<td>Bud</td>
<td>4.1</td>
<td>10.2</td>
<td>0.90</td>
<td>0.76</td>
<td>Exp</td>
<td>–</td>
<td>Exp</td>
<td>–</td>
<td>No</td>
</tr>
<tr>
<td>BARsp</td>
<td>11.8</td>
<td>21.0</td>
<td>0.87</td>
<td>0.74</td>
<td>Exp</td>
<td>–</td>
<td>Exp</td>
<td>–</td>
<td>No</td>
</tr>
<tr>
<td>COMsp</td>
<td>70.0</td>
<td>13.0</td>
<td>0.87</td>
<td>0.62</td>
<td>Exp</td>
<td>–</td>
<td>Exp</td>
<td>–</td>
<td>Yes</td>
</tr>
<tr>
<td>Madison aquifer sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LA88C</td>
<td>14.4</td>
<td>7.8</td>
<td>0.90</td>
<td>0.78</td>
<td>Exp</td>
<td>Logn</td>
<td>Exp</td>
<td>Logn</td>
<td>No</td>
</tr>
<tr>
<td>Tilf</td>
<td>8.5</td>
<td>12.0</td>
<td>0.99</td>
<td>0.85</td>
<td>Exp</td>
<td>Logn</td>
<td>Exp</td>
<td>Logn</td>
<td>No</td>
</tr>
<tr>
<td>RG</td>
<td>2.0</td>
<td>22.6</td>
<td>1.00</td>
<td>0.92</td>
<td>Exp</td>
<td>–</td>
<td>Exp</td>
<td>–</td>
<td>Yes</td>
</tr>
<tr>
<td>RFsp</td>
<td>29.2</td>
<td>0.0</td>
<td>0.83</td>
<td>–</td>
<td>Exp</td>
<td>Logn</td>
<td>Exp</td>
<td>Logn</td>
<td>No</td>
</tr>
<tr>
<td>FALr</td>
<td>61.9</td>
<td>10.9</td>
<td>0.55</td>
<td>0.67</td>
<td>Exp</td>
<td>Logn</td>
<td>Exp</td>
<td>Logn</td>
<td>No</td>
</tr>
<tr>
<td>WCL</td>
<td>9.3</td>
<td>16.0</td>
<td>0.99</td>
<td>0.91</td>
<td>Logn</td>
<td>–</td>
<td>Logn</td>
<td>–</td>
<td>No</td>
</tr>
<tr>
<td>LScr</td>
<td>23.2</td>
<td>0.0</td>
<td>0.76</td>
<td>–</td>
<td>Exp</td>
<td>Exp</td>
<td>Exp</td>
<td>Exp</td>
<td>No</td>
</tr>
<tr>
<td>SPFcr</td>
<td>10.2</td>
<td>11.8</td>
<td>0.58</td>
<td>0.70</td>
<td>Exp</td>
<td>Exp</td>
<td>Exp</td>
<td>Exp</td>
<td>Yes</td>
</tr>
<tr>
<td>Cave drip</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTD</td>
<td>0.9</td>
<td>1.0</td>
<td>0.10</td>
<td>0.80</td>
<td>Exp</td>
<td>Logn</td>
<td>–</td>
<td>–</td>
<td>–$^d$</td>
</tr>
<tr>
<td>RmDr</td>
<td>1.1</td>
<td>0.0</td>
<td>0.97</td>
<td>–</td>
<td>Logn</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–$^d$</td>
</tr>
<tr>
<td>Fractured-rock watershed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEVcr</td>
<td>6.2</td>
<td>14.8</td>
<td>0.50</td>
<td>0.71</td>
<td>Exp</td>
<td>Exp</td>
<td>Exp</td>
<td>–</td>
<td>No</td>
</tr>
</tbody>
</table>

IRF, impulse-response function; w1, wet period primary; w2, wet period secondary; d1, dry period primary; d2, wet period secondary; Exp, exponential curve; Logn, lognormal curve; –, not applicable.

$^a$ Dimensionless; computed on a daily time step.

$^b$ Ratio of system memory to full system-response period, dimensionless.

$^c$ Also includes a backward validation period of 12.5 yr.

$^d$ Period was too short to determine stationarity. Simulation had one wet period and no dry period.
Fig. 1. Examples of compound impulse-response functions (IRFs) consisting of the superposition of an exponential and a lognormal curve (a and b) and two exponential curves (c and d).
Fig. 2. Study areas showing the Madison and Edwards aquifers.
Fig. 3. Modeling results for Barton Springs showing observed and simulated springflow.
Fig. 4. Principle component biplot showing simulated sites and metrics for (a) the first two principle components and (b) the first and third principle component. Plotting positions were scaled to range from -1 to 1 to show sites and metrics on a similar scale. Site and metric labels are shown in Tables 2 and 3.
Fig. 5. Impulse-response functions for selected sites in clusters $A_s$, $B_s$, and $C_s$ plotted for scale-independent comparison. Black line is wet period; gray line is dry period. The vertical axes extend from zero to the maximum peak height, and the horizontal axes extend from zero to the system memory.