RRAWFLOW: Rainfall-Response Aquifer and Watershed Flow Model (v1.15)

A. J. Long[1]

[1]{ U.S. Geological Survey, 1608 Mountain View Rd, Rapid City, South Dakota, USA}

Abstract

The Rainfall-Response Aquifer and Watershed Flow Model (RRAWFLOW) is a lumped-parameter model that simulates streamflow, springflow, groundwater level, or solute transport for a measurement point in response to a system input of precipitation, recharge, or solute injection. I introduce the first version of RRAWFLOW available for download and public use and describe additional options. The open-source code is written in the R language and is available at *http://sd.water.usgs.gov/projects/RRAWFLOW/RRAWFLOW.html*, along with an example model of streamflow. RRAWFLOW includes a time-series process to estimate recharge from precipitation and simulates the response to recharge by convolution; i.e., the unit-hydrograph approach. Gamma functions are used for estimation of parametric impulse-response functions (IRFs); a combination of two gamma functions results in a double-peaked IRF. A spline fit to a set of control points is introduced as a new method for estimation of nonparametric IRFs. Options are included to simulate time-variant systems by several different methods. For many applications, lumped models simulate the system response with equal accuracy to that of distributed models, but moreover, the ease of model construction and calibration of lumped models makes them a good choice for many applications; e.g., estimating missing periods in a hydrologic record. RRAWFLOW provides professional hydrologists and students with an accessible and versatile tool for lumped-parameter modeling.

1   Introduction

1.1   Lumped versus distributed models

Hydrologic models, commonly referred to as a "lumped-parameter" or "lumped" models, generally have a small number of parameters, each representing a property of the entire

1

hydrologic system; conceptually, many physical processes are lumped into a few parameters. In contrast to lumped models, distributed models discretize the system into small compartments or cells, each of which has several parameters defined. All hydrologic models, however, are lumped to some degree. Models that frequently are considered physically based simulate numerous small-scale physics by lumping these processes into simplified mathematical forms (Beven, 1989). The use of the term "physically based" to describe any hydrologic model, therefore, should be discouraged (Beven and Young, 2013). Both distributed and lumped models, however, have components that can represent different hydrologic processes that can be interpreted in physically meaningful ways (Beven and Young, 2013). For example, the IRF estimated in many lumped models represents the physical response to an impulse into the system and provides mechanistic insights into that system, including the peak response time and magnitude and the hydrologic memory of the system (von Asmuth and Knotters, 2004; Beven and Young, 2013; Young, 2013). The IRF could be measured directly at the outflow point (e.g., a spring) if a short, intense recharge event follows a long, dry period. Most commonly however, the outflow, or system response, results from a series of superposed responses to repeating recharge events, and the lumped model is used to estimate the IRF iteratively and to simulate the system response.

In a comparison of lumped models to distributed models, Reed et al. (2004) concluded that lumped models had better overall performance than distributed models but also cited several other studies indicating that distributed or semi-distributed models may or may not provide improvement over lumped models. In another comparison, Smith et al. (2013) concluded that distributed models provided improvements over lumped models in 12–24% of the cases tested, depending on the criteria of evaluation. The mixed results of these comparisons indicate that lumped models are a good choice when the objectives do not require a distributed model.

A major advantage of a lumped model is its ease of construction and calibration because of the small number of parameters to estimate and because there is no need to assemble large datasets representing the physical properties of the system. Lumped models are useful for karst aquifers, where the geometry of the conduit network frequently is unknown. Lumped models provide an efficient means to simulate the response to possible future changes in the system input (e.g., precipitation). A lumped groundwater model might be more effective than a distributed model in regard to conditional validation and predictive modeling because of its

simplicity, as will be discussed. The primary advantage of distributed models is to simulate the response to possible changes within the system, such as urban development or increased groundwater pumping, for example. The choice to use a lumped or distributed model, therefore, depends on a study's objectives and available resources; a lumped model likely is the better choice if it meets the study's objectives.

1.2 RRAWFLOW overview

The Rainfall-Response Aquifer and Watershed Flow Model (RRAWFLOW) is a lumped model that is partially based on unit-hydrograph theory applied to streamflow (Nash, 1959). RRAWFLOW simulates a time-series record for a measurement point of streamflow, springflow, groundwater level, or solute transport in response to a system input of precipitation, recharge, or solute injection. A preliminary version of RRAWFLOW was developed by Long and Mahler (2013) and used to classify karst aquifers and characterize time-variant systems. This preliminary version also was used by Symtad et al. (2014) to simulate future scenarios of streamflow and groundwater level in a cave in Wind Cave National Park, United States and by the U.S. Geological Survey to simulate future scenarios of springflow and groundwater levels (*https://nccwsc.usgs.gov/display-project/4f8c652fe4b0546c0c397b4a/52d5615ae4b0f19e63da8647*). Although this preliminary version was applied primarily, but not exclusively, to karst, RRAWFLOW is suitable for aquifers and watersheds of any type, and non-karst systems generally are easier to model than karst systems. The overarching methods used in RRAWFLOW have been applied extensively to non-karst surface-water and groundwater systems (e.g., Nash, 1959; Blank et al., 1971; Delleur and Rao, 1971; Dooge, 1973; Neuman and de Marsily, 1976; Maloszewski and Zuber, 1982; Besbes and de Marsily, 1984; Beven, 1989; Jakeman and Hornberger, 1993; von Asmuth et al., 2002; Reed et al., 2004; von Asmuth and Knotters, 2004; Olsthoorn, 2008; Jurgens et al., 2012; Smith et al., 2013).

The purpose of this paper is to present a new version of RRAWFLOW with added functionality, to make the code publicly available, and to guide users in its operation. New functions in this version include (1) the gamma function for parametric impulse-response functions (IRFs), (2) a spline curve or straight-line segments fit through a set of control points for nonparametric IRFs, (3) a new option for time-variant systems that uses a continuously changing IRF scale, (5) two methods to determine wet and dry periods, and (5) any user-

defined IRF. To my knowledge, the spline-curve method previously has not been used for the IRF. The open-source code written in the R language (*http://www.r-project.org/index.html*) can be downloaded from *http://sd.water.usgs.gov/projects/RRAWFLOW/RRAWFLOW.html* along with a user's manual, example model, and a quick-start guide for the R novice.

Time-invariant and time-variant systems were described by Jenkins and Watts (1968). For example, Larocque et al. (1998) described high-flow periods exhibiting distinctly different response characteristics from low-flow periods. RRAWFLOW includes several options to simulate time-variant systems that generally are not available for distributed watershed models (e.g., PRMS; *http://wwwbrr.cr.usgs.gov/projects/SW_MoWS/PRMS.html*). If a distributed model is required for a specific study, RRAWFLOW might be a comlimentary exploratory tool to analyze the system's sensitivity to time-variant response characteristics. RRAWFLOW is useful for estimation of missing periods in a hydrologic record and as an educational tool for hands-on instruction of some of the basic principles in hydrology. Several example applications that demonstrate model options and calibration and validation procedures are included herein. Input, output, and calibration files are available from the RRAWFLOW website for one of these examples.

2   The model

The model's time-step interval is determined by the input data record, which must have equal time steps, and model output is generated for the same time step. Hydrological and meteorological data commonly are available for a daily time step, which is suitable for most simulations over a time frame of months to decades. Time steps shorter than one day can be used when high-resolution responses are of interest. Any time-step interval can be used because the equations are not time-unit specific. However, the time step should be equal to or less than the quickest identifiable response; longer time steps will result in a loss of information about response dynamics (Jakeman and Hornberger, 1993). RRAWFLOW also is independent of specific units for flow, water level, or solute concentration, and the user should maintain unit consistency. Air temperature is always in °C.

2.1   Precipitation recharge

Effective precipitation for a watershed is the amount of precipitation that results in streamflow exiting the watershed. This consists of infiltration to groundwater below the root

4

zone that reemerges as streamflow, springflow, shallow groundwater interflow, and overland runoff. Processes that apply to effective precipitation for watershed modeling also apply to infiltration recharge to groundwater that causes a response in springflow or groundwater level, except that overland runoff generally does not contribute to groundwater recharge. In RRAWFLOW, the term "recharge" is used for both watershed modeling and groundwater modeling. Methods used in RRAWFLOW to simulate precipitation recharge are described in Long and Mahler (2013), and the equations also are presented in Appendix A herein for convenience and reference in the RRAWFLOW User's Manual, which is available at *http://sd.water.usgs.gov/projects/RRAWFLOW/RRAWFLOW.html.*

2.2   Other recharge options

Recharge estimated outside of RRAWFLOW can be used as model input. For example, this applies to precipitation recharge estimated by a soil-water-balance model (e.g., Westenbroek et al., 2010) or sinking-stream recharge in karst aquifers that can be estimated by methods such as those described by Hortness and Driscoll (1998). This is system-input option 2 (Table 1).

2.3   Convolution

Convolution is a time-series operation (Jenkins and Watts, 1968; Smith, 2003) that is commonly used in non-distributed hydrologic models to simulate streamflow, springflow, or groundwater level in response to recharge (e.g., Nash, 1959; Dooge, 1973; Dreiss, 1989; Olsthoorn, 2008). The use of convolution in modeling also has been described as a linear-reservoir model and a transfer-function model (e.g., Nash, 1959; Young, 2013; von Asmuth et al., 2002). The discrete form of the convolution integral for uniform time steps used in RRAWFLOW is

$$y_i = \Delta t \sum_{j=0}^{i} \beta_j h_{i-j} u_j + \varphi_i + d_0 \qquad\qquad i,j = 0, 1, \dots, N \qquad\qquad (1)$$

where h_{i-j} is the IRF; u_j is the input, or forcing function; j and i are time-step indices corresponding to system input and output, respectively; N is the number of time steps in the output record; β_j is an optional time-varying IRF scaling coefficient; φ_i represents the errors resulting from measurement inaccuracy, sampling interval, or simplifying model assumptions; and d_0 is a hydraulic-head datum used in simulation of groundwater levels. d_0 is the level to

1   which hydraulic head would converge on if the local recharge was eliminated. Local recharge

2   is assumed to be the only forcing that results in hydraulic-head fluctuation or that causes

3   hydraulic head to rise above d_0. The errors φ_i are not explicitly simulated but are shown in Eq.

4   (1) for clarity. Further details and background are provided in Long and Mahler (2013).

5   The quantity $i - j$ represents the delay time from impulse to response, and the IRF represents a

6   distribution of these delay times. In RRAWFLOW, the input function u_j can be recharge or

7   input of a solute. The system response y_i can be streamflow exiting a watershed, springflow

8   from a groundwater system, groundwater level, or solute concentration at an outlet.

9   Physically, the IRF is the system response y_i per unit impulse of u_j and also can be described

10 as the response produced by a system when the input is a delta function (Smith, 2003).

11 Conceptually, convolution is the superposition of a series of IRFs that are initiated at the time

12 of each impulse of u_j and are scaled proportionally by the magnitude of the corresponding

13 impulse (Fig. 1).

14 **2.4   Solute transport**

15 RRAWFLOW can simulate transport of a solute, similarly to the approach of Maloszewski

16 and Zuber (1982). In this case, the user-provided system input u_j is the solute concentration,

17 and a constant recharge rate is assumed. The response in solute concentration at the outlet of a

18 system is simulated by the convolution integral (Eq. 1) with the IRFs described in the

19 following section. Convolution temporally disperses a system input of a solute, according the

20 IRF characteristics, at the system outlet. This is system-input option 3 (Table 1).

21 **2.5   Impulse-response function (IRF)**

22 The IRF characterizes the relation between system input and output by convolution (Eq. 1)

23 and has been described by other terms, including instantaneous unit hydrograph, transfer

24 function, and kernel (e.g., Nash, 1959; Dreiss, 1989; Berendrecht et al., 2003; Smith, 2003;

25 Jukić and Denić-Jukić, 2006). However, the term "transfer function" should only be applied

26 to the Fourier transform of the IRF (Smith, 2003). The IRF of a hydrologic system can be

27 approximated by a parametric function, where its shape is defined by one or more parameters,

28 or a nonparametric function that is not constrained by common curve types.

Deleted: To compare convolution with frequenc
analysis, the frequency spectrum is a function that
characterizes a single time series at a glance,
whereas the IRF is a function that characterizes th
relation between system input and output.

1   █████ Parametric IRFs

2 Parametric functions that have been used to approximate the IRF for hydrologic systems

3 include exponential, lognormal, and gamma functions (Nash, 1959; Besbes and de Marsily,

4 1984; Jakeman and Hornberger, 1993; von Asmuth et al., 2002; Berendrecht et al., 2003; von

5 Asmuth and Knotters, 2004; Long, 2009; Long and Mahler, 2013). The gamma function is

6 equivalent to the Pearson type III function: the three parameters of the Pearson type III

7 function can be combined into the two parameters of the gamma function (Haan, 2002).

8 Estimation of parametric IRFs generally consists of model-calibration techniques to optimize

9 the parameters with the aim of minimizing the difference between the observed and simulated

10 system response; i.e., fitting the model. The parametric functions previously described (other

11 than Pearson type III) have one or two of these fitting parameters. As the number of fitting

12 parameters increases, the risk of over-fitting the model also increases; i.e., fitting the errors φ

13 in Eq. (1).

14 For a parametric approximation of the IRF, RRAWFLOW uses the gamma function:

15 $$\gamma(t) = \frac{\lambda^{\eta} t^{\eta-1} e^{-\lambda t}}{\Gamma(\eta)} \qquad\qquad \lambda, \eta > 0 \qquad\qquad (2)$$

16 $$\Gamma(\eta) = \int_{t=0}^{\infty} t^{\eta-1} e^{-t} dt \qquad\qquad (3)$$

17 where λ and η are unitless shape parameters, and the mean and variance are η/λ and η/λ^2,

18 respectively. Eq. (3) is approximated in RRAWFLOW by the discrete form

19 $$\Gamma(\eta) = \Delta t \sum_{t=t_0}^{N} t^{\eta-1} e^{-t} \qquad\qquad (4)$$

20 where t is time centered on each discrete time step; t_0 and N are time centered on the initial

21 and final time steps, respectively; and Δt is the time step duration. The gamma function can

22 produce a variety of shapes, including exponential ($\eta = 1$), reverse-J ($\eta < 1$), and positively

23 skewed shapes with a peak at $t = (\eta - 1)/\lambda$ (Fig. 2; Haan, 2002). The gamma function can

24 produce nearly identical shapes to those of the lognormal function when $\eta > 1$, and therefore,

25 can produce nearly all possible shapes of the exponential and lognormal functions combined

26 when $\eta \geq 1$, plus the additional reverse-J shape when $\eta < 1$. The RRAWFLOW option to use

27 parametric IRFs is specified as IRF type 1 (Table 1).

1 The gamma function (Eq. 2), which has an area under its curve of unity, requires the

2 additional scaling coefficient ε for use as the IRF in many hydrologic applications:

3      $h(t) = \varepsilon \gamma(t)$ (5)

4 where ε [unitless] compensates for hydrologic systems that do not have a one-to-one relation

5 between system input and output (Olsthoorn, 2008). For example, if (1) the system input u_j is

6 in cubic meters per day of recharge, (2) the system response y_i is springflow with the same

7 units, and (3) 100% of this recharge emerges as springflow with nothing else contributing to

8 springflow, then ε would be set to unity. For most other hydrologic applications, ε would not

9 equal unity. Similarly, if 100% of a solute entering the system does not exit the system at the

10 observation point, then the area under the IRF should be less than unity ($\varepsilon < 1$). Maloszewski

11 and Zuber (1982) simulated solute transport with IRFs that were approximated by the

12 exponential or dispersion-model functions. The gamma function has a similar shape to that of

13 the dispersion-model function and could be used as an approximation of the dispersion-model

14 function, or the exact dispersion-model function can be provided to RRAWFLOW as a user-

15 defined IRF.

16 RRAWFLOW allows the use of as many as two superposed gamma functions, herein referred

17 to as double-gamma IRFs, to produce additional IRF shapes such as a double-peaked curve;

18 several examples are shown in Long and Mahler (2013), except with lognormal functions.

19 Approaches similar to this have been used to represent the components of quick flow and

20 slow flow in watershed modeling (Jakeman and Hornberger, 1993) and for conduit and

21 diffuse flow in karst systems (Pinault et al., 2001; Long, 2009; Long and Mahler, 2013). In

22 these examples, each parametric function represents one of two flow components. The use of

23 a double-gamma IRF also might be useful when a single function cannot produce the

24 necessary IRF approximation (e.g., an extra-long tail). The scaling coefficient ε can be set to

25 different values for the two gamma functions; e.g., to allow for a larger component of slow

26 flow than of quick flow.

27 ███ Nonparametric IRFs

28 The process of determining an unknown IRF from observed system input and output data is

29 known as deconvolution (e.g., Neuman and de Marsily, 1976). To define a nonparametric

30 IRF, an ordinate value is defined for each time step, and any shape desired is possible.

31 Deconvolution methods include Fourier harmonic time-series analysis (Blank et al., 1971;

Deleted: (Fig. 3),

1 Delleur and Rao, 1971), linear programming (Neuman and de Marsily, 1976), and time-
2 moment analysis (Dreiss, 1989). Estimations of nonparametric IRFs by model calibration
3 include those described by Pinault et al. (2001) and Jukić and Denić-Jukić (2006). A potential

4 problem with nonparametric IRFs is that hundreds or even thousands of IRF ordinates may be
5 needed to define the IRF, depending on the IRF length and time step. Optimization of each
6 individual ordinate would result in a mathematically underconstrained and over-fit model. An
7 extreme example of over-fitting is to determine the IRF by means of deconvolution in the
8 frequency domain (Smith, 2003) that results in a numerically perfect model fit but also an IRF
9 that commonly is highly oscillatory and cannot be explained physically (Blank et al., 1971;
10 and Delleur and Rao, 1971) because the errors φ (Eq. 1) are included in the fitting process.
11 Filtering the IRF in the frequency domain (i.e., transfer function; Smith, 2003) or smoothing
12 the IRF in the time domain (Long and Derickson, 1999) are options for IRF estimation by
13 Fourier analysis, which may require trial-and-error calibration. Further, an over-fitted model
14 results in a poor model fit when tested on a conditional validation period that was sequestered
15 from the fitting process. Pinault et al. (2001), Jukić and Denić-Jukić (2006), and Ladouche et
16 al. (2014) described different methods to constrain the nonparametric IRF and reduce the
17 number of fitting parameters.

18 The method proposed herein uses a small number of ordinates to define a smoothly shaped
19 nonparametric IRF: ordinates of the IRF are defined at spaced intervals (IRF control points),
20 and a spline curve is fit through these points (Fig. 3) (IRF type 2, Table 1). Another option is
21 to apply straight-line segments connecting the control points (IRF type 3, Table 1). Similar to
22 parametric IRFs, these two nonparametric options are convenient for the estimation of the IRF
23 through model calibration and conditional validation because of the ability to control the
24 number of fitting ordinates. If a model is suspected of having been over-fit, the number of
25 control points should be reduced; this consists of increasing the control-point intervals,
26 resulting in a smoother shape, or by reducing the tail length by setting posterior control points
27 to zero. Trial and error generally is required to determine the optimum number of control
28 points for a given application. The minimum number of control points is two: at least one to
29 define the non-zero part of the curve and one to define where the function becomes zero.

30 Another option allows a predefined IRF to be supplied to RRAWFLOW if the IRF is
31 determined by some other method (IRF type 4, Table 1). The scaling coefficient ε is not
32 necessary for nonparametric IRFs because the area is defined by the ordinate values.

9

1 **Linearity and time variance**

2 The terms "linear," "nonlinear," time invariant," and "time variant" may cause some
3 confusion. Estimated recharge (Appendix A) is a nonlinear process, where recharge as a
4 fraction of precipitation varies with antecedent soil-moisture conditions. Convolution (Eq. 1),
5 which simulates the system response to recharge, is a linear system (Jenkins and Watts, 1968;
6 Dooge, 1973). This linear system can be either time variant or time invariant, depending on
7 whether or not the IRF changes with time (Jenkins and Watts, 1968). Most commonly, a time-
8 invariant (i.e., static) IRF is assumed in hydrologic convolution models (e.g., von Asmuth et
9 al., 2002; Denić-Jukić and Jukić, 2003). In many hydrologic systems, however, the IRF
10 changes with changing climatic conditions, resulting in a change in response characteristics
11 (Larocque et al., 1998; Long and Mahler, 2013). Additional details and examples of time-
12 variant IRFs for hydrologic applications include Pinault et al. (2001), Jukić and Denić-Jukić
13 (2006), and Long and Mahler (2013).

14 **Time-variance (TV) options**

15 RRAWFLOW has three options for time variance in convolution (Table 1). In time-variance
16 (TV) option 1, the IRF is time invariant, or static. TV option 2 applies a time-variant IRF,
17 similarly to the method proposed by Long and Mahler (2013), which uses a minimal number
18 of fitting parameters but also represents the dominant transient characteristics of the system.
19 In this method, the system-input record is separated into climatically wet or dry periods. One
20 IRF represents all of the wet periods, and the other represents all of the dry periods. The IRF
21 scaling variable β (Eq. 1) is set to unity for TV options 1 and 2.

22 All of the parametric and nonparametric IRF-type options previously described can be used in
23 TV option 2 (Table 1). An advantage of this method is that both the size and shape of the IRF
24 can change, while the fitting parameters are kept to a minimum, because IRFs are not defined
25 continuously but rather for two different periods only. A potential disadvantage of this
26 method is that the IRF changes abruptly between wet and dry periods; however, this was not a
27 detrimental factor for several models in which this method was applied (Long and Mahler,
28 2013). Also, the superposition of many responses applied in convolution results in smooth
29 transitions in the simulated response between wet and dry periods. Jukić and Denić-Jukić

1 (2006) proposed a similar time-variant approach, where three different IRFs were applied to

2 one of three different hydrologic periods determined by an index of antecedent recharge.

3 Pinault et al. (2001) varied the IRF's vertical scale continuously with hydraulic head.

4 However, because hydraulic head also is used for model calibration, this approach cannot

5 undergo conditional validation or be used to simulate periods without observed system-

6 response data; e.g., future periods that might be simulated with climate projections. TV option

7 3 in RRAWFLOW (Table 1) is similar to the approach of Pinault et al. (2001), except that the

8 IRF scaling variable β (Eq. 1), varies according to the input for convolution u_j (e.g. recharge)

9 by

10 $$\beta_j = mx_j \qquad m \neq 0 \qquad\qquad (6)$$

11 where x_j is the moving average of u_j (Eq. 1) that is scaled to range from 0 to 1, and m

12 determines the range of β. A moving average of u_j is used so that the IRF transitions

13 smoothly. Generally, β is assumed to vary directly with x ($m > 0$). Advantages of this method

14 are that it requires fewer fitting parameters than TV option 2 and the IRF does not change

15 abruptly; the disadvantage is that only the vertical scale of the IRF changes, whereas the

16 shape is static. All of the parametric and nonparametric IRF-type options previously described

17 can be used in this option. TV option 3 has longer run times than TV options 1 or 2 because

18 of the additional computation required, mainly within the convolution loop.

19 For time-invariant systems, the cross-correlation function (CCF) has the same shape as the

20 IRF but only if the input to the convolution process is completely random (Jenkins and Watts,

21 1968). If the convolution input has a strong autocorrelation, typical of recharge in hydrologic

22 systems, then there is large error in using the CCF to estimate the IRF (Jenkins and Watts,

23 1968; Bailly-Comte et al., 2011) and therefore should be avoided.

24 ## 2.6   Determining wet and dry periods

25 RRAWFLOW includes two options to determine wet and dry periods on the basis of the

26 precipitation input record when using TV option 2. The first option assigns each calendar year

27 to a wet period if the annual mean precipitation is greater than the overall mean P_{mean} for the

28 entire input record, and other years are set to dry years (wet-switch option 1; Table 1). The

29 second option sets wet and dry periods according to the slopes of a cumulative precipitation

30 function in which upward or downward slopes indicate wet or dry periods, respectively (wet-

switch option 2; Table 1). This option (1) calculates a record of the cumulative departure from P_{mean}, (2) calculates the annual mean cummulative departure (CD_{mean}) from this record, (3) sets the time period from the mid-point of a calendar year to the mid-point of the following year to a wet period if CD_{mean} has a net increase during this interval, and (4) sets all other periods to dry. A third option allows the user to provide a record of wet and dry periods in the model input (wet-switch option 0; Table 1).

2.7 Model outputs

Model outputs consist of time series for simulated system response y_i, the dry-period and wet-period IRFs (if using TV option 2), the soil-moisture index s_i (Eq. A1, Appendix A), and the input to convolution u_j. Other outputs consist of a coefficient of efficiency E to measure the similarity between simulated and observed system response (residuals) and the hydrologic memory of the system. This system memory is the time that the response to an impulse effectively persists, which is defined by the length of the IRF. Because the gamma function is asymptotic and has infinite length, system memory is arbitrarily defined in RRAWFLOW as time t_m on the IRF time scale at which 95% of the curve area is in the range $0 - t_m$.

2.8 Evaluating model fit and over-fitting

The calibration period is the period of the data record used to calibrate the model. By default in RRAWFLOW, the conditional validation period is the part of the data record following the calibration period that is used to test the model calibration against system-response observation data not used in calibration (i.e., model prediction of streamflow or springflow). Assessing the conditional validation period is an indication of the expected model performance to predict a future period on the basis of climate simulations, for example; moreover, this assessment indicates if the model is being over-fit. This validation is considered conditional because the model cannot yet be tested against additional observational data that will be available in the future (Beven and Young, 2013). RRAWFLOW calculates a modified form of the Nash–Sutcliffe coefficient of efficiency (Nash and Sutcliffe, 1970; Legates and McCabe, 1999) to quantify model fit, as proposed by Long and Mahler (2013). This modification calculates the coefficient of efficiency E for a partial period, either calibration or conditional validation, in a manner that allows the two periods to be compared directly:

12

1

$$E = 1 - \frac{\left[\Sigma \left(y_{obs} - y_{sim} \right)^2 \right]_p}{\left[\Sigma \left(y_{obs} - y_{mean} \right)^2 \right]_T \left(\frac{l_p}{l_T} \right)}$$

1

(7)

2   where y_{obs} and y_{sim} are time series of the observed and simulated system responses,

3   respectively; y_{mean} is the mean value of y_{obs}; the subscripts p and T refer to the partial and total

4   periods, respectively; and l is the time length of the respective period. Conceptually, E is the

5   ratio of the magnitude of model residuals (numerator) to the overall variability in the

6   observation record (denominator) subtracted from unity and theoretically can vary from $-\infty$

7   (poorest fit) to unity (perfect fit).

8   In addition to quantifying model fit, E provides a useful way to evaluate possible over-fitting

9   of the model. Although model fit for the calibration period might improve as parameters are

10   added, if the validation period indicates that this added complexity is not helpful, the model

11   has been over-fit (von Asmuth et al., 2002). To test this condition, E is calculated for the

12   calibration and conditional validation periods separately (E_{cal} and E_{val}) by the modified Nash–

13   Sutcliffe coefficient of efficiency (Eq. 7), which makes E_{cal} and E_{val} directly comparable

14   (Long and Mahler, 2013). This method is particularly important for comparison of two

15   periods with different fluctuation amplitudes.

16   Legates and McCabe (1999) describe limitations of correlation-based measures to quantify

17   model fit, such as the coefficient of determination R^2, and the benefits of the Nash–Sutcliffe

18   coefficient of efficiency and the index of agreement. The sum of the squared and weighted

19   residuals (Doherty, 2005) is another useful metric used for this purpose. Hartmann, et al.

20   (2013) provides an example of using multiple metrics to evaluate model performance.

21   A value of E_{cal} that is much larger than E_{val} might indicate over-fitting, in which case a

22   simpler model (i.e., fewer fitting parameters) should be tested. For example, if a double-

23   gamma IRF is used, then a second model calibration with a single-gamma IRF could be tested

24   to determine if greater similarity in the E_{val} and E_{cal} values is achieved. For nonparametric

25   IRFs, a reduction in the number of IRF control points could be tested. A time-variant IRF

26   requires more parameters than a time-invariant IRF, and this also can be tested. Thoroughly

27   considering model complexity in light of E_{val} and E_{cal} provides context for conditional

28   validation. For example, Long and Mahler (2013) described decision criteria to evaluate

Deleted: An E value of zero indicates that the observed mean (y_{mean}) is an equally good predictor is the simulation (y_{sim}) (Legates and McCabe, 199... Because E compares the residuals to the overall response variability, E values for different simulat... sites can be directly compared; this would not be t... if the magnitude of the residuals alone is used in comparisons; e.g., as with the root mean square error.

Deleted: the coefficient of efficiencythe coeffici... of efficiency

Deleted: . For direct comparison of E_{cal} and E_{val}, the denominator must be consistent for both cases and therefore, the denominator is calculated on the basis of the total period and the numerator on the basis of the shorter period of interest (Long and Mahler, 2013). The denominator of Eq. (13) is sca... down to be consistent with the time period of the numerator1 where the subscript p refers to the par... period, either calibration or conditional validation... the subscript T refers the total period, and l is the time length of the respective period.

Deleted: For the total period, E is calculated by Eq. (13), which is equivalent to using Eq. (14) because the partial and total periods are equal. Therefore E, E_{cal}, and E_{val} all are directly comparable.

Deleted: Generally,

Deleted: will

Deleted: be larger (better fit)

Deleted: but a large difference indicates possibl...

model complexity and the number of fitting parameters on the basis of E_{cal} and E_{val}, with several examples of calibrated models.

The number of gamma functions or fitting parameters might correspond to different conceptual models of the system, and model complexity issues can be investigated by testing multiple conceptual models (e.g., Hartmann et al., 2013). Numerous other researchers have investigated issues related to model complexity and its effect on model-prediction uncertainty (e.g., Young et al., 1996; Jakeman and Hornberger, 1993; Arkesteijn and Pande, 2013). Prediction uncertainly crucially depends on model complexity (Arkesteijn and Pande, 2013). Although Vapnik–Chervonenkis generalization theory suggests that models with higher complexity tend to have higher prediction uncertainty, model complexity is not necessarily proportional to prediction uncertainty (Fienen et al., 2010). Doherty and others (2010) described a method for predictive uncertainty and sensitivity that tests the range of each parameter's potential values on the basis of expert knowledge and propagates this uncertainty to model predictions. Estimating this potential range of IRF parameter values, however, might be more difficult than, for example, estimating hydraulic conductivity or streambed roughness in a distributed model. Although a rigorous assessment of prediction uncertainty is beyond the scope of this article, effective tools are available for this purpose (Doherty, 2005; Fienen et al., 2010).

3   Example model applications

The model was applied to three hydrologic systems in the United States with responses of streamflow, springflow, and groundwater level. Several examples with different RRAWFLOW options and different levels of parameterization are described, including examples of model over-fitting. The hydrologic systems were selected to provide a wide range of examples that required different levels of model complexity. The first hydrologic system is streamflow from a watershed, for which a simple model was appropriate. Karst settings were selected for the second and third hydrologic systems to provided examples in which more complex models are needed. For precipitation and air-temperature inputs, gridded data (e.g. Daymet: http://daymet.ornl.gov/) can be used, or a single weather station can be assumed to represent the recharge area. All examples used a daily time step.

14

Model spin-up is the initial simulation period in which antecedent effects of the system are not fully incorporated into the simulation, and therefore, model output for this period is not valid. When the simulation is past the number of time steps equal to the system memory, then the system antecedent effects are fully incorporated into the model. Therefore, the model input record must start n time steps prior to the calibration period, where n is the system memory, as a number of time steps. Because the system memory is not known until the IRF is estimated, it is useful to start the simulation at the earliest date for which input data are available. Estimated system-input values can be used if observation data are not available for this antecedent-period requirement, and a constant value equal to the long-term mean can be used if a better estimate is not available; in this case, the antecedent effects will be smoothed. Another option is to select a period from the input data record and use this as input for the spin-up period.

The parameter optimization software PEST (Doherty, 2005) was used for parameter estimation in these examples. RRAWFLOW is a stand-alone model independent of PEST and, therefore, can be used with any optimization method, including trial and error. For optimization of nonparametric IRFs, the last control point was used to set the system memory by assigning a fixed (non-optimized) value of zero to that control point (Fig. 3). Posterior to this point, a series of control points fixed at zero was specified, resulting in a spline fit with a constant value of zero.

3.1   Streamflow in Boxelder Creek

Boxelder Creek is located in the Black Hills of South Dakota, USA, with a watershed area of 250 km^2 upstream from U.S. Geological Survey streamgage 06422500, with daily streamflow available from http://waterdata.usgs.gov/nwis. The watershed primarily is pine forest and contains metamorphic rocks of Precambrian age (Carter et al., 2001). Gridded daily precipitation and air temperature data from the Daymet dataset are available at 1-km grid spacing for 1980–2013. These data were obtained from the Geo Data Portal (http://cida.usgs.gov/gdp/) and spatially averaged for the watershed to produce a daily time series of precipitation and air temperature for 1980–2013, which was used as model input. The calibration period was 1980–1996, and validation was for 1997–2013 (Fig. 4). A 5-year model spin-up period was applied by inserting data for 1980–1984 into the period 1975–1979.

This estimated spin-up period affected the calibration period minimally, because the system memory was only about 3-months long.

The example models described used system-input option 1 (precipitation recharge, Table 1). Five example models are presented for Boxelder Creek, all of which used single-gamma IRFs (Table 2). All other trials with double-gamma IRFs resulted in minimization of one of the IRFs, indicating that single-gamma IRFs were appropriate for this system. All Boxelder Creek gamma functions optimized to $\eta < 1$, which is the reverse-J shape (Fig. 5). Example BC1 used TV option 1 (time-invariant IRF, Table 1), resulting in E_{cal} and E_{val} values of 0.62 and 0.46, respectively (Table 2). Examples BC2 and BC3 used TV option 2 (time-variant IRF) with wet-dry options 1 and 2, respectively. For BC3, E_{val} was higher (0.56) than for BC1 and BC2 (Table 2). Example BC4 and BC5 used TV option 3, in which a time-variant IRF that changes continually was used with a single-gamma IRF and moving-average (MA) windows of 1 and 10 years (Eq. 6), respectively. Of the five examples, BC3 had the highest E_{val} value; BC5 had the second highest E_{val} value and with fewer parameters (Table 2). Comparison of model fit for examples BC4 and BC5 indicates that the time-variant aspects of this system respond to general climatic changes over decadal periods more so than annual. Table 3 shows optimized IFR parameters for selected example models.

3.2 Springflow from Barton Springs

Barton Springs is a group of springs that flow from the Edwards aquifer, a carbonate aquifer in south-central Texas that is contained mostly within the Edwards Group (Lower Cretaceous geologic age). Model input data consisting of daily precipitation and air temperature and system-response observation data used for model evaluation are described in Long and Mahler (2013) along with details describing the hydrogeology, physiography, and climate. The example models described used system-input option 1 (precipitation recharge, Table 1). Seven example models are presented for Barton Springs (Table 2). Model fit varied more widely than for the Boxelder Creek models, possibly as a result of karst features in the Edwards aquifer that result in complex groundwater flow. For this reason, IRFs with added complexity are tested for the Barton Springs examples, and examples of over-fitting are demonstrated. Generally, E_{cal} is proportional to the number of optimized parameters for each example; however, high E_{cal} values often resulted in low values of E_{val}, which might indicate over fitting (Table 2).

16

Examples F1–F9 used wet-switch option 2. Examples F1, F5, F7 used gamma functions with an increasing number of parameters for the three examples, which were characterized as low E_{val} (F1), good E_{val} (F5), and an over-fit model (F7) that is indicated by a high E_{cal} and low E_{val} (Table 2). Example F5, with a moderate number of parameters, is considered the best choice of the three. Examples F8 and F9 used 7 and 16 optimized control points, respectively, in total (Figs. 6 and 7); F8 was considered a good choice, and F9 was over fit with too many control points (Table 2). Of the time-variant examples F5–F9, the two examples with the smallest number of optimized parameters (F5 and F8) had the largest E_{val} values (Table 2).

Examples F2 and F3 used TV option 3, with a single-gamma IRF and moving-average (MA) windows of 1 and 10 years (Eq. 6), respectively, but resulted in low E_{val} values. Similarly to Boxeler Creek, increasing the MA window from 1 to 10 years improved E_{cal} and E_{val} values, which indicates that the time-variant aspects of this system respond to general climatic changes over decadal periods more so than annual.

PEST was used to calculate 95% confidence intervals for the optimized parameters, as described in Doherty (2005), which are shown graphically for example F8 (Figs. 6). Example F9, with a total of 16 control points, had 32% wider parameter confidence intervals than did example F8, with 7 control points. Confidence intervals generally widen with an increasing number of parameters because of a decrease in individual parameter sensitivity.

3.3 Groundwater level in well LA88C

Well LA88C, located in western South Dakota, is open to the karstic Madison aquifer that is composed of limestone and dolostone of Mississippian geologic age. Model input data consisting of daily precipitation and air temperature and system-response observation data used for model evaluation are described in Long and Mahler (2013) along with details describing the hydrogeology, physiography, and climate. The example models described used system-input option 1 (precipitation recharge, Table 1).

Examples W1–W3 used TV option 2 with wet-switch option 2 for time variance (Table 1). All three examples for well LA-88C optimized to double-peaked IRFs, which were necessary for this system, as indicated by additional tests of single-peaked IRFs, probably as result of karst features (Figs. 8–10). In example W2, the last two control points for the wet period and the last control point for the dry period were optimized to zero, resulting in only 15 non-zero

17

control points. Of the three examples for well LA88C, W3 had the fewest optimized IRF parameters and resulted in the largest E_{val} value (Table 2), indicating that this is a good choice. IRFs for example W3 approach zero abruptly, resulting in negative values in the spline curve (Fig. 10); in these cases, RRAWFLOW sets all negative IRF ordinates to zero.

4   Discussion and conclusions

Watersheds simulated by Long and Mahler (2013) and Long (2009), which included karst and non-karst systems, had IRF shapes similar to the reverse-J shape for Boxelder Creek (Fig. 5), except that double-exponential IRFs were used to achieve this shape. A single reverse-J gamma function (Fig. 5) requires only three parameters, whereas the double-exponential IRF requires four parameters. IRFs for karst and non-karst watersheds commonly have quick-flow and slow-flow components (Jakeman and Hornberger, 1993; Long, 2009; Long and Mahler, 2013). The reverse-J IRF (Fig. 5) also exhibits quick-flow and slow-flow components in the form of a high peak and long tail, respectively, but fewer paramters for the gamma function is an advantage over the exponential function.

Although RRAWFLOW can be applied to any type of watershed or aquifer, karst aquifers might require more complex models. A non-karst system was compared with two karst systems, which indicated that the best model choices for the karst systems generally had a larger number of parameters than the best choices for the non-karst system (Table 2). Also, differences between wet- and dry-period IRFs were more pronounced for the karst systems than for Boxelder Creek (Figs. 5–9), possibly as result of heterogeneity. Example F5 is another karst example, with a reverse-J ($\eta < 1$) dry-period IRF and a delayed-peak ($\eta > 1$) wet-period IRF (Table 3). In karst aquifers, fluctuating groundwater levels might saturate or desaturate different conduit networks that result in different hydrologic responses between wet and dry periods.

Examples F5 and F8 are the two preferred models for Barton Springs and are nearly identical in terms of E_{val} and the number of optimized parameters (Table 2). Choosing between these two models, therefore, might be a matter of modeler preference. Use of the gamma function (example F5) has the advantage of being a common function. The control-points method (example F8) has the advantage that confidence intervals for the IRF can be easily shown in a graph (Fig. 6). Showing confidence intervals for a gamma function also could be done but

18

1 with additional steps involved, in which the gamma function would have to be calculated for
2 all combinations within the 95% parameter confidence intervals (i.e., Monte Carlo
3 simulations). Then this family of curves would be plotted, and the maximum upper and lower
4 curve extents would show the confidence intervals for the IRF. A disadvantage of the control-
5 points method is the need to select the temporal locations of control points and also to set the
6 system memory *a priori* by setting a zero-value control point at the end of the IRF. These
7 settings generally require trial and error.

8 Additional functionality can be added to RRAWFLOW by the user and could possibly be
9 included in future versions. For example, additional methods to estimate parametric or
10 nonparametric IRFs (e.g., the dispersion-model IRF) or the degree-day method for estimating
11 snowmelt (Rango and Martinec; 1995) could be added. If there were a need to include
12 precipitation recharge and sinking-stream recharge simultaneously in one system, this could
13 easily be added. An adjustment to the calculation of the soil-moisture index *s* could be
14 included to account for watershed changes such as tree coverage. Revisions, additions, and
15 corrections to the RRAWFLOW code can be sent to the author of this article for potential
16 incorporation into subsequent official versions. The code is not yet available in the
17 comprehensive R archive network (CRAN) but could be included in the future. Optimization
18 packages also are available in CRAN and could be built seamlessly into RRAWFLOW.

19 Comparison of the modified Nash–Sutcliffe coefficient of efficiency for the calibration and
20 conditional validation periods (E_{cal} and E_{val}) is useful for assessing over-fitting. Conceptual-
21 model options that maximize E_{val} can evaluated by multiple tests. Too many fitting
22 parameters, as well as too few, can result in low values of E_{val}. The ratio E_{val}/E_{cal} might be a
23 useful metric for comparison of different models and possibly in setting the lengths of the
24 calibration and validation periods. As in any model, this all should be considered in reference
25 to a physical understanding of the system; e.g., two distinct permeability domains might be
26 best simulated by two gamma functions.

27 The record length of the observed response should be considered in light of the system
28 memory: there is less confidence in the predictive strength of a model if the observed
29 response is shorter than the system memory than if it is longer, because, in the former case,
30 the effects of the IRF tail are not fully tested against observation. Ideally, the validation

Deleted: If E_{cal} is much larger than E_{val}, then the model may have been over-fit. Examples of model over-fitting (i.e., too many fitting parameters) include using a double-gamma IRF when a single-gamma IRF is suitable, applying time-variant IRF when time invariance is adequate, or using more control points for a spline fit than necessary. Generally, the objective should be to maximize E_{val} to achieve the best conditional validation possible, which frequently is at the expense of lowering E_{cal}. The selection of IRF type, number of control points and time-variance options to

period alone should be longer than the system memory, and if it is several times longer, then the full range of the IRF is tested several times over.

Konikow and Bredehoeft (1992) discuss the numerous uses of the term "validation," which has resulted in confusion, and also highlighted philosophical considerations associated with the term. They argue that conditional validation, as described herein [split-sample test in Konikow and Bredehoeft (1992)], is not useful for distributed groundwater models because of their limited predictive accuracy that results from non-unique solutions in calibration of complex models. Further, they argue that the split-sample-test period must be independent of any antecedent effects from the calibration period, which they say rarely can be achieved for a large-scale aquifer system. These arguments highlight an advantage of lumped models, because (1) a small number of parameters minimizes the problem of non-unique solutions, (2) selecting one model from multiple models via a validation process also reduces the problem of non-unique solutions, and (3) a lumped model provides an estimate of the system memory, which indicates the time span for antecedent effects following a calibration period. More than one half of the groundwater sites simulted by Long and Mahler (2013) had conditional validation periods that extended beyond the antecedent effects of the calibration period, and therefore, E_{val} could be calculated for this restricted period only if desired.

RRAWFLOW is useful for estimation of missing periods of a hydrologic record and is suitable for hydrograph-separation methods to estimate stream base flow, as described by Jakeman and Hornberger (1993) and Long (2009). For the simulated hydrograph, RRAWFLOW can be used to compute the base-flow component by executing the model without the quick-flow IRF. If using a single reverse-J gamma IRF instead of a double-exponential IRF that can be easily separated, the reverse-J function would need to be separted into its quick-flow (peak) and slow-flow (tail) components. To estimate the base-flow component of the observed hydrograph, a graphical separation program can be used, such as PART (Rutledge, 1998); however, because the different options and settings in PART (or similar programs) result in different base-flow estimates, the RRAWFLOW estimated base flow is helpful as guide to using PART (Long, 2009). For example, the PART settings can be adjusted so that the observed hydrograph separation has similar characteristics to those of the simulated hydrograph separation. The RRAWFLOW-simulated hydrgrograph separation also could be used as a benchmark model for comparison to more elaborate methods, such as geochemical hydrograph separation (e.g., Rimmer and Hartmann (2014).

20

5   Code availability

The RRAWFLOW program written in R language, RRAWFLOW manual, quick-start guide, and an example model are included in a download package available at *http://sd.water.usgs.gov/projects/RRAWFLOW/RRAWFLOW.html*. The model included is example BC3 for Boxelder Creek (Table 2). It is not necessary to know the R language to execute the model, but R must be installed on the user's computer. The example is set up to run on the Microsoft® Windows operating system but could be slightly modified to run on a Linux operating system. RRAWFLOW is in the public domain, and no license is needed.

Although the example model is set up for parameter optimization using the PEST software program (Doherty, 2005), RRAWFLOW can be used with any optimization routine, including trial and error. All RRAWFLOW input and output files are included along with PEST input, output, and executable files. The file 00_Quick_Start_Guide.pdf in the download package contains instructions for executing RRAWFLOW in the R environment and basic instructions for PEST execution for this example. The RRAWFLOW manual has detailed input and output instructions. The download package can be used as a template for a new modeling project by editing the input files accordingly. The R language program and PEST can be downloaded at no cost from *http://www.r-project.org/index.html* and *http://www.pesthomepage.org/*, respectively. Supplemental tables 1–4 contain model input values for all examples previously described.

6   Acknowledgements

21

Deleted: included in the online supplement to the article

Deleted: The

Deleted: 00_ReadMe.pdf in the supplement

Deleted: t

Deleted: supplement

Appendix A: Precipitation recharge

2   To simulate recharge from direct precipitation, a soil-moisture index s [unitless] is estimated

3   for each time step in RRAWFLOW. Quantitatively, s is the fraction of precipitation that

4   infiltrates and becomes recharge. To account for the antecedent effects of rainfall on soil

5   moisture, the past rainfall record is weighted by a backward-in-time exponential decay

6   function (Jakeman and Hornberger, 1993):

7    $$s_i = cr_i + \left(1 - \kappa_i^{-1}\right) s_{i-1}$$

8    $$= c\left[r_i + \left(1 - \kappa_i^{-1}\right) r_{i-1} + \left(1 - \kappa_i^{-1}\right)^2 r_{i-2} + ... \right]$$

(A1)

9    $i = 0, 1, ..., N \qquad 0 \leq s \leq 1$

10   where c [L^{-1}] is a scaling coeffient to constrain the value of s; κ [unitless] adjusts the effect of

11   antecedent rainfall and is related to evapotranspiration; r is total rainfall [L]; and i is the time

12   step. In RRAWFLOW, this method is option 1 for system input (Table 1). For watershed

13   modeling, the value of c can be set to satisfy the assumption that the total recharge volume

14   within a watershed is equal to the total outflow volume for the calibration period. This

15   assumption neglects the net change in total watershed storage during this period, which is

16   assumed to be small in comparison to the total inflow or outflow for the same period. Also,

17   this assumption does not apply if recharge to the watershed exits the watershed through deep

18   groundwater and bypasses the stream outlet. Recent rainfall has the largest effect on s in Eq.

19   (A1), whereas earlier rainfall has the least effect.

20   The effect of changing air temperatures on evapotranspiration is accounted for by (Jakeman

21   and Hornberger, 1993):

22   $$\kappa_i = \alpha \exp\left[(20 - T_i)f\right] \qquad f > 0,$$ (A2)

23   where α [unitless] is a scaling coefficient; T [°C] is mean air temperature at the land surface;

24   and f is a temperature modulation factor [°C^{-1}]. As air temperature T decreases, s in Eq. (A1)

25   increases with sufficient past rainfall. RRAWFLOW can be executed without air-temperature

26   data if unavailable (air-temperature option 2 in Table 1). Recharge for each time step u_i [L] is

27   calculated as the fraction s of precipitation by

Deleted: ≥

Deleted: ≥

Deleted: coefficient that limits s to the range 0–

Deleted: , typically in days

$$u_i = r_i s_i \qquad\qquad\qquad\qquad (A3)$$

Typically, *s* is largest during wet periods and rarely reaches a maximum value of 1.0 (Fig. 14).

Considerations for parameters *c* and *κ*

An additional function of parameters *c* and *κ* is to adjust for differences in the runoff effects between watershed and groundwater modeling. Also, for groundwater applications, *c* in Eq. (A1) cannot be determined empirically if the recharge area that affects a spring or well is not precisely defined. Therefore, for groundwater applications, *c* can be set to a value that results in a predefined maximum *s* value or estimated mean recharge rate, or *c* can be optimized through model calibration. In practice, the error in the estimation of *c* is compensated by an adjustment in the IRF area during model calibration; e.g., an overestimation of *c* by 10% would result in a 10% underestimation in IRF area.

Depending on the values of *c* and *κ*, the value of *s* can incorrectly have values <0 or >1; when this occurs, RRAWFLOW sets *s* to 0 or 1, respectively. This is most likely to occur early in the calibration process when parameter values might be far from optimum, and forcing the constraint $0 \le s \le 1$ assists in the efficiency of the calibration process. To ensure that the range of *s* is appropriate for the model area, this parameter should always be plotted after model calibration; i.e., *s* should be a physically plausible function that fluctuates in response to local precipitation and air temperature (Fig. 11). For example, in humid climates with high annual precipitation, *s* might frequently have a value >0.9, which is less likely in dry climates.

Snow precipitation

For cold climates where winter snowfall is common, a method proposed by Long and Maher (2013) is applied. To determine the form of precipitation for each time step, an air temperature threshold value T_s is set, below which precipitation is assumed to occur as snow (typically $T_s = 0°C$). To determine time steps when melting occurs, a melting threshold value T_m is set. If daily snow-depth data are available, T_m can be determined empirically as the mean air temperature for time steps when snow depth decreases to zero from a previous time step with a snow depth greater than zero. Long and Mahler (2013) determined that $T_m = 9°C$ for a

Deleted: ≥

Deleted: ≥

Deleted: $0 \ge s \ge 1$

study area in central North America. Sublimation is accounted for by a sublimation fraction S_f. Snow precipitation is summed for each series of snow-precipitation time steps occurring prior to each snowmelt time step by

$$p_m = (1 - S_f) \sum_{i=1}^{N} p_i \qquad\qquad T_i < T_s \qquad\qquad\qquad (A4)$$

where p_m is the accumulated snow precipitation that is assumed to melt when $T_i > T_m$, S_f is the sublimation fraction [unitless], p_i is the snow precipitation in height of water, and N is the number of snow-precipitation time steps occurring between melt time steps. Prior to calculating Eq. (A1), p_m is added to the rainfall record for the time step following a snowmelt time step because snowmelt is assumed to have the same effect as rainfall on the value of s.

References

Arkesteijn, L. and Pande, S.: On hydrological model complexity, its geometrical interpretations and prediction uncertainty, Water Resour. Res., 49, 7048-7063, 2013.

Bailly-Comte, V., Martin, J. B. and Screaton, E. J.: Time variant cross correlation to assess residence time of water and implication for hydraulics of a sink-rise karst system, Water Resour. Res., 47, 2011.

Berendrecht, W. L., Heemink, A. W., Van Geer, F. C. and Gehrels, J. C.: Decoupling of modeling and measuring interval in groundwater time series analysis based on response characteristics, Journal of Hydrology, 278, 1-16, 2003.

Besbes, M. and de Marsily, G.: From infiltration to recharge: use of a parametric transfer function, Journal of Hydrology, 74, 271-293, 1984.

Beven, K.: Changing ideas in hydrology — The case of physically-based models, Journal of Hydrology, 105, 157-172, 1989.

Beven, K. and Young, P.: A guide to good practice in modeling semantics for authors and referees, Water Resour. Res., 49, 5092-5098, 2013.

Blank, D., Delleur, J. W. and Giorgini, A.: Oscillatory kernel functions in linear hydrologic models, Water Resour. Res., 7, 1102-1117, 1971.

Carter, J.M., Driscoll, D.G., Hamade, G.R., Estimated recharge to the Madison and Minnelusa aquifers in the Black Hills area, South Dakota and Wyoming, water years 1931-98, US Geol. Surv., Reston Virginia, Water-Resources Investigations Report 00-4278, 66 pp., 2001.

Delleur, J. W. and Rao, R. A.: Linear systems analysis in hydrology: the transform approach, kernel oscillations and the effect of noise, in: Systems Approach to Hydrology, Water Resources Publications, Fort Collins, Colorado, 116-142, 1971.

Denić-Jukić, V. and Jukić, D.: Composite transfer functions for karst aquifers, Journal of Hydrology, 274, 80-94, 2003.

Doherty, J.: PEST—Model-Independent Parameter Estimation, User Manual (5th ed.), Watermark Numerical Computing, Brisbane, Australia, http://www.pesthomepage.org/, 2005.

Doherty, J.E., Hunt, R.J., and Tonkin, M.J.: Approaches to highly parameterized inversion: A guide to using PEST for model-parameter and predictive-uncertainty analysis, U.S.

25

Geological Survey Scientific Investigations Report 2010–5211, 71 p., http://pubs.usgs.gov/sir/2010/5211, 2010.

Dooge, J. C. I.: Linear theory of hydrologic systems, Washington, D.C., 1468 pp., 1973.

Dreiss, S. J.: Regional scale transport in a karst aquifer: 2. Linear systems and time moment analysis, Water Resour. Res., 25(1), 126-134, 1989.

Fienen, M. N., Doherty, J. E., Hunt, R. J. and Reeves, H. W.: Using prediction uncertainty analysis to design hydrologic monitoring networks: example applications from the Great Lakes water availability pilot project, US Geol. Surv., Reston Virginia, Scientific Investigations Report 2010–5159, 44 pp., 2010.

Haan, C. T.: Statistical Methods in Hydrology, Second ed., The Iowa State University Press, Ames, Iowa, 2002.

Hartmann, A., Wagener, T., Rimmer, A., Lange, J., Brielmann, H. and Weiler, M.: Testing the realism of model structures to identify karst system processes using water quality and quantity signatures, Water Resour. Res., 49, 1-14, doi:10.1002/wrcr.20229, 2013.

Hortness, J. E. and Driscoll, D. G.: Streamflow losses in the Black Hills of western South Dakota, US Geol. Surv., Reston Virginia, Water-Resources Investigations Report 98-4116 , 99 pp., 1998.

Jakeman, A. J. and Hornberger, G. M.: How much complexity is warranted in a rainfall-runoff model?, Water Resour. Res., 29, 2637-2649, 1993.

Jenkins, G. M. and Watts, D. G.: Spectral Analysis and its Applications, Holden-Day, San Francisco, California, 1968.

Jukić, D. and Denić-Jukić, V.: Nonlinear kernel functions for karst aquifers, J. Hydrol., 328, 360-374, 2006.

Jurgens, B. C., Bohlke, J. K. and Eberts, S. M.: TracerLPM (Version 1): An Excel® workbook for interpreting groundwater age distributions from environmental tracer data, U.S. Geological Survey, Reston, VA, Techniques and Methods Report 4-F3, 60 pp., 2012.

Konikow, L. F. and Bredehoeft, J. D.: Ground-water models cannot be validated, Adv. Water Resour., 15, 75-83, 1992.

Ladouche, B., Marechal, J. and Dorfliger, N.: Semi-distributed lumped model of a karst system under active management, Journal of Hydrology, 509, 215-230, 2014.

1   Larocque, M., Mangin, A., Razack, M. and Banton, O.: Contribution of correlation and
2   spectral analyses to the regional study of a large karst aquifer (Charente, France),
3   Journal of Hydrology, 205, 217-231, 1998.
4   Legates, D. R. and McCabe Jr., G. J.: Evaluating the use of 'goodness-of-fit' measures in
5   hydrologic and hydroclimatic model validation, Water Resour. Res., 35, 233-241,
6   1999.
7   Long, A. J.: Hydrograph separation for karst watersheds using a two-domain rainfall-
8   discharge model, Journal of Hydrology, 364, 249-256, 2009.
9   Long, A. J. and Derickson, R. G.: Linear systems analysis in a karst aquifer, J. Hydrol., 219,
10   206-217, 1999.
11   Long, A. J. and Putnam, L. D.: Linear model describing three components of flow in karst
12   aquifers using 18O data, Journal of Hydrology, 296, 254-270, 2004.
13   Long, A. J. and Mahler, B. J.: Prediction, time variance, and classification of hydraulic
14   response to recharge in two karst aquifers, Hydrol. Earth Syst. Sci., 17, 281-94, 2013.
15   Mahler, B. J., Valdes, D., Musgrove, M. and Massei, N.: Nutrient dynamics as indicators of
16   karst processes: Comparison of the Chalk aquifer (Normandy, France) and the
17   Edwards aquifer (Texas, U.S.A.), J. Contam. Hydrol., 98, 36-49, 2008.
18   Maloszewski, P. and Zuber, A.: Determining the turnover time of groundwater systems with
19   the aid of environmental tracers. 1. Models and their applicability, J. Hydrol. (Amst.),
20   57, 207-231, 1982.
21   Nash, J. E.: Systematic determination of unit hydrograph parameters, J. Geophys. Res., 54,
22   111-115, 1959.
23   Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I - A
24   discussion of principles, J. Hydrol., 10, 282-290, 1970.
25   Neuman, S. P. and de Marsily, G.: Identification of linear systems response by parametric
26   programming, Water Resour. Res., 12, 253-262, 1976.
27   Olsthoorn, T. N.: Do a bit more with convolution, Ground Water, 46, 13-22, 2008.
28   Pinault, J. L., Plagnes, V., Aquilina, L. and Bakalowicz, M.: Inverse modeling of the
29   hydrological and the hydrochemical behavior of hydrosystems; characterization of
30   karst system functioning, Water Resour. Res., 37, 2191-2204, 2001.
31   Rango, A. and Martinec, J.: Revisiting the degree-day method for snowmelt computations,
32   Water Resour. Bull, 31, 657-669, 1995.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F. and Seo, D. -J.: Overall distributed
model intercomparison project results, Journal of Hydrology, 298, 27-60, 2004.

Rimmer, A. and Hartmann, A.: Optimal hydrograph separation filter to evaluate transport
routines of hydrological models, Journal of Hydrology, 514, 249-257,
doi:10.1016/j.jhydrol.2014.04.033, 2014.

Rutledge, A. T.: Computer programs for describing the recession of ground-water discharge
and for estimating mean ground-water recharge and discharge from streamflow
records -update, US Geol. Surv., Reston Virginia, Water-Resources Investigations
Report 98-4148, 52 pp., 1998.

Smith, S. W.: Digital Signal Processing: A Practical Guide for Engineers and Scientists,
Newnes, Amsterdam, 2003.

Smith, M., Koren, V., Zhang, Z., Moreda, F., Cui, Z., Cosgrove, B., Mizukami, N.,
Kitzmiller, D., Ding, F., Reed, S., Anderson, E., Schaake, J., Zhang, Y., Andréassian,
V., Perrin, C., Coron, L., Valéry, A., Khakbaz, B., Sorooshian, S., Behrangi, A.,
Imam, B., Hsu, K. -., Todini, E., Coccia, G., Mazzetti, C., Ortiz Andres, E., Francés,
F., Orozco, I., Hartman, R., Henkel, A., Fickenscher, P. and Staggs, S.: The distributed
model intercomparison project - Phase 2: Experiment design and summary results of
the western basin experiments, Journal of Hydrology, 507, 300-329, 2013.

Symstad, A. J., Long, A. J., Stamm, J. F., King, D. A. and Bachelet, D. M.: Two Approaches
for Planning Natural Resource Management in a Changing Climate at Wind Cave
National Park, National Park Service Natural Resource Technical Report
NPS/WICA/NRTR—2014/918, 87 pp.,
http://irmafiles.nps.gov/reference/holding/510174, 2015.

von Asmuth, J. R., Bierkens, M. F. P. and Maas, K.: Transfer function-noise modeling in
continuous time using predefined impulse response functions, Water Resour. Res., 38,
231-2312, 2002.

von Asmuth, J. R. and Knotters, M.: Characterising groundwater dynamics based on a system
identification approach, Journal of Hydrology, 296, 118-134, 2004.

Westenbroek, S. M., Kelson, V. A., Dripps, W. R., Hunt, R. J. and Bradbury, K. R.: SWB - A
modified Thornthwaite-Mather soil-water-balance code for estimating groundwater
recharge, US Geol. Surv., Reston Virginia, Techniques and Methods 6-A31, 60 pp.,
2010.

1 Young, P., Parkinson, S. and Lees, M.: Simplicity out of complexity in environmental
2   modelling: Occam's razor revisited, Journal of Applied Statistics, 23, 165-210, 1996.
3 Young, P. C.: Hypothetico-inductive data-based mechanistic modeling of hydrological
4   systems, Water Resour. Res., 49, 915-935, 2013.
5

1   Table 1. RRAWFLOW options.

2   System input
3   1 = System input is precipitation that results in recharge to the system (Eqs. 1–4).
4   2 = System input is recharge estimated outside of RRAWFLOW (skip Eqs. 1–4).
5   3 = System input is solute concentration (skip Eqs. 1–4).

6   System output
7   1 = System output is groundwater level.
8   2 = System output is springflow or streamflow.
9   3 = System output is solute concentration (System input = 3).

10   IRF type
11   1 = Parametric IRF—gamma functions.
12   2 = Nonparametric IRF—spline fit to IRF control points.
13   3 = Nonparametric IRF—linear fit to IRF control points.
14   4 = Nonparametric IRF—user-defined IRF.

15   Time-variance (TV) option
16   1 = Time-invariant (static) IRF.
17   2 = Wet-period IRF and dry-period IRF are defined separately, and each are time invariant within these
18   respective periods.
19   3 = Variable IRF vertical scale, where β in Eq. (1) is variable.

20   Wet-switch option
21   0 = The wet and dry periods are provided by the user.
22   1 = Wet and dry periods are calculated by RRAWFLOW. Any calendar year when the mean precipitation is
23   above the mean for the entire precipitation record is a wet year, and other years are dry years.
24   2 = Wet and dry periods are calculated by RRAWFLOW according to the annual cumulative departure from
25   mean precipitation.

26   Air-temperature option
27   1 = Use air temperature adjustment (Eq. A2).
28   2 = Do not use air temperature adjustment.

29

30

1 **Tables**
2

Table 2. Summary of example models. The shaded rows indicate the best choices. [--, not applicable]

Site	Example	Figure	Time-variance (TV) option[1]	Wet-switch option[2]	Description of impulse-response function (IRF)	No. of optimized IRF parameters	Time variant	E_{cal}	E_{val}	Comments
Boxelder Creek	BC1	--	1	--	Single-gamma IRF (static)	3	no	0.62	0.46	Low E_{val}
	BC2	--	2	1	Single-gamma IRFs	6	yes	0.63	0.47	Low E_{val}
	BC3	5	2	2	Single-gamma IRFs	6	yes	0.61	0.56	Best E_{val}
	BC4	--	3	--	Single-gamma, variable-scale IRF; 1-year window	4	yes	0.63	0.46	Low E_{val}
	BC5	--	3	--	Single-gamma, variable-scale IRF; 10-year window	4	yes	0.68	0.53	Good E_{val} and small number of parameters
Barton Springs	F1	--	1	--	Single-gamma IRF (static)	3	no	0.84	0.63	Low E_{val}
	F5	--	2	2	Single-gamma IRFs	6	yes	0.88	0.72	Good E_{val}
	F7	--	2	2	Double-gamma IRFs	12	yes	0.90	0.63	Over fit
	F8	6	2	2	7 optimized control points	7	yes	0.88	0.72	Good E_{val}
	F9	7	2	2	16 optimized control points	16	yes	0.92	0.61	Over fit
	F2	--	3	--	Single-gamma, variable-scale IRF; 1-year window	4	yes	0.80	0.46	Low E_{val}
	F3	--	3	--	Single-gamma, variable-scale IRF; 10-year window	4	yes	0.84	0.61	Low E_{val}
Well LA88C	W1	8	2	2	Double-gamma IRFs	12	yes	0.92	0.73	--
	W2	9	2	2	18 optimized control points (double-peaked IRF)	18	yes	0.93	0.70	--
	W3	10	2	2	10 optimized control points (double-peaked IRF)	10	yes	0.88	0.75	Best E_{val} and smallest number of parameters

1   Table 3. Impulse-response function (IRF) parameters for selected example models.

Example	BC3	BC5	F5	F7	W1
Site	Boxelder Creek	Boxelder Creek	Barton Springs	Barton Springs	Well LA88C
Dry period					
λ_1	1.45E-02	2.08E-02	2.15E-04	5.26E-04	5.15E-03
η_1	5.40E-01	4.15E-01	6.26E-01	6.61E-01	1.15E+00
ε_1	8.32E+00	1.07E+01	2.01E+01	1.65E+01	1.08E+02
λ_2	0	0	0	3.82E-02	1.31E-02
η_2	0	0	0	5.10E+01	6.36E+00
ε_2	0	0	0	4.98E+00	1.42E+02
Wet period					
λ_3	6.34E-03	0	6.39E-03	7.77E-03	1.57E-03
η_3	3.12E-01	0	1.14E+00	1.22E+00	1.16E+00
ε_3	9.68E+00	0	1.09E+01	1.11E+01	2.50E+02
λ_4	0	0	0	9.53E-02	1.89E-02
η_4	0	0	0	4.95E+01	3.86E+01
ε_4	0	0	0	8.03E-01	5.28E+01

2

3

List of figures

1

2 Figure 1. Superposition of sequential impulse-response functions (IRFs). Each IRF is in response to an
3 impulse of the input function u_j and is scaled by the magnitude of that impulse.

4 Figure 2. Gamma functions for different values of its two shape parameters.

5 Figure 3. Impulse-response function defined by a spline curve fit through control points. The last
6 control point with a value of zero is not adjusted during model calibration.

7 Figure 4. Example B3: observed (gray) and simulated (black) flow for Boxelder Creek.

8 Figure 5. Example B3: impulse-response functions (IRFs) for Boxelder Creek using single-gamma
9 IRFs.

10 Figure 6. Example F8: nonparametric impulse-response functions (IRFs) for Barton Springs using a
11 total of 7 control points showing (a) optimized IRFs and (b) upper and lower 95% confidence limits
12 for the IRFs.

13 Figure 7. Example F9: nonparametric impulse-response functions for Barton Springs using a total of
14 16 control points.

15 Figure 8. Example W1: impulse-response functions (IRFs) for well LA88C using double-gamma IRFs
16 for both periods; i.e., wet and dry.

17 Figure 9. Example W2: nonparametric impulse-response functions for well LA88C using a total of 18
18 control points.

19 Figure 10. Example W3: nonparametric impulse-response functions for well LA88C using a total of
20 10 control points.

21 Figure 11. The soil-moisture parameter s (Eq. A1) for example BC2. The function s typically is largest
22 during wet periods (shaded gray).