

Responses to each of referees are provided below. A marked manuscript that shows where edits were made is at the end of this document.

Author responses to Anonymous Referee #1

Referee comment is shown, followed by Author's response.

In the submitted manuscript Dr Long presents the functioning and application of the RRAWFLOW model. The model consists of two main routines, one of them to calculate effective precipitation and recharge by a soil moisture index method. The other is meant to calculate groundwater dynamics by a variety of Impulse-response functions (IRFs). The model is highly adaptive to the user's tasks (1) by allowing for turning off the recharge routine (or replacing it), (2) by the wide range of IRFs that are capable of reflecting a wide range of hydrological behavior (including interplay of diffuse and concentrated subsurface flow), and (3) by being able to provide solute transport simulations simply by replacing the hydrological time series with hydrochemical time series and adapting the IRFs.

The universal applicability as well as the detailed elaborations of the model functioning and its application (including two examples) make me believe that this manuscript is an adequate contribution for GMD. However some corrections, mainly to its structure, are necessary to make it better understandable. The author tends to provide a lot of detail on special cases of the model application within the model description some of which may be better located within the discussion. Also some more elaborations about the use of alternative performance measures (other than Nash Sutcliffe) and alternative calibration routines (other than PEST) would be desirable. For more details please see the commented pdf of the manuscript.

Author's response: I thank the reviewer for these helpful comments. Some of the topics were moved the Discussion and Conclusions section, as suggested. Discussion of alternative performance measures were added. Because the referee submitted specific comments via a commented pdf of the manuscript, I provided responses with my comments added in a copy of this pdf, which was uploaded as a reply to the referee in the interactive discussions.

Author responses to Anonymous Referee #2

Referee comments are shown, followed by Author's responses.

The present manuscript (MS) introduces a conceptual rainfall-runoff model and describes basic functionalities. The model runs in open source language "R", the source code is provided and as such has a high potential to be used by the community either for research, for applied science or for education purposes. RRAWFLOW is flexible regarding input, is simple and parsimonious and ready to use with a sample data set. This is highly acknowledged and principally deserves publication in GMD. Here, I am fully in line with referee #1. BUT: A high level of redundancy exists to a published paper in HESS by the same author (Long and Mahler 2013). While the exact wording has been changed in the present

manuscript, most of the equations and also the data used to test RRAWFLOW have been published before (Long and Mahler 2013). Therefore, I propose that the model should be generalized and tested with new data to show its general value. Moreover, the manuscript should significantly be shortened. I detail my concerns down below.

---Author's response: The author thanks the referee this thoughtful review. Examples with new model data have been added, and the manuscript (MS) elsewhere has been shortened, except where other referees requested additions. The MS did not make it very clear that this is the first release of the RRAWFLOW code to the scientific and public communities. Long and Mahler (2013) presented it as a research model with no code or user's manual. The revised MS now clearly states in the abstract and introduction that the purpose of this paper is to present a new version of RRAWFLOW with added versatility, to make the code publicly available, and to guide users in its operation.

A) Redundancy with Long and Mahler 2013

In (Long and Mahler, HESS 2013) all main components of the RRAWFLOW have been described in great detail: The majority of the equations (1,2,3,5,6) and also the text paragraphs in between are included there. Wording is a little different but the meaning is more or less equivalent. As a difference, a more general parametric IRF (gamma function) is introduced in the present manuscript (MS) instead of concentrating on exponential or lognormal IRFs in (Long and Mahler 2013). Also time-variant IRFs are not a new feature here, also those have been discussed and tested in depth in the preceding paper. Most strikingly, also the data set used to test the model has been published before, Figure 6 in the present MS is equivalent to Figure 3 in (Long and Mahler 2013). I recommend to refer to the published material and strictly remove redundant data, paragraphs, equations and text passages.

---Author's response: Equations 1-5 have been moved to an appendix. This was necessary because specific variables in Eqs. 1-5 are referred to throughout the MS and are included in RRAWFLOW outputs, and the RRAWFLOW download contains a user's manual that cites all equations in this manuscript. To avoid confusion for the user, I feel it will be most helpful if all equations are contained in the body or appendix. Eq. 6 (now Eq. 1) is left in the main body because (1) I added one term to show how the equilibrium datum is applied in groundwater models, and (2) it pertains to Fig. 1, which is new in this MS to help explain convolution conceptually, which people seem to have trouble with. The related text has been shortened.

---Time-variance options, although some were presented previously, need to be discussed briefly for the purpose of explaining the different RRAWFLOW settings and options referred to in Table 1. The new continually changing IRF scale function needs to be presented also. An additional feature that I have included since submitting the MS is the option to have RRAWFLOW calculate the wet and dry periods on the basis of the precipitation record, which includes options for two different methods. Previously, the user had to supply the wet and dry periods. Therefore, a short section (section 2.6) was added to describe how these options work and the associated RRAWFLOW settings.

B) Applicability of RRAFLOW outside karst

For me it is not stated clear enough, if RRAFLOW is a model exclusively for karst systems or for other catchments/aquifers as well. This needs to be clearly clarified at the beginning of the MS. But here I see also a great chance for improvement: I propose to apply the model also to non-karst systems. If RRAFLOW can be used (and I am pretty sure that this is possible, because the structure is highly flexible), then this should be shown by new data. Differences in model performance, parameters, etc. should be highlighted in the discussion. This would add a large value to the manuscript and lower the problem of redundant information (see A above).

---Author's response: This is an excellent point. The introduction (section 1.2) now states much more clearly that RRAFLOW is suitable for hydrologic systems of any type, not just karst. Five example simulations for a non-karst watershed have been added, which are compared to the karst simulations, and the discussion highlights the differences and additional challenges when modeling karst. Then, in the Discussion and Conclusions, a comparison is made to another non-karst watershed simulated by Long and Mahler (2013). Because it is easier to simulate non-karst systems in RRAFLOW, the karst examples are useful to show how the different RRAFLOW functions can be used effectively to simulate more complex hydrologic systems.

C) Lengthy discussions on IRFs and tests with the same data set

Eleven of thirteen Figures show IRFs, theoretic or generated by modelling the Barton Springs or a well in the Madison Aquifer. This section needs to be condensed, particularly because the IRF-discussion is not new (again see A above). Again, I propose to apply RRAFLOW to other systems as well, if possible outside karst, if not possible to a system with a different karstification level. Finally, the IRFs received can be compared in order to show if system properties can be deducted from different IRF-shapes.

---Author's response: The figures related to Barton Springs and the well have been reduced to 5 figures. Four of these figures are used to demonstrate the new control-points IRF option. The related discussion also has been condensed, as suggested. A discussion of the different system properties related to the different IRFs was added to the Discussion and Conclusions section, as suggested.

D) Subdaily timestep

In l15-16 p5923 it is mentioned that sub-daily time steps can be used. This requires discussion and needs to be justified, because most model parameters and concepts are daily (soil moisture index, evapotranspiration, recharge, snowmelt etc.).

---Author's response: The equations are not specific to any particular time-step. Therefore, all references in the MS to daily time steps (except in examples) were deleted and are now referred to in general terms. A discussion was added at the beginning of the section 2 (The model) that explains that the time-step length is selected by the user and defined by the time step of the input record. A

discussion of choosing an appropriate time step also was added. I also revised the RRAWFLOW code and input-file format to make it easier to use time steps that are shorter or longer than one day.

Author responses to Anonymous Referee #3

Referee comments are shown, followed by Author's responses (----)

In this paper, a new code for convolution modeling of hydrological or mass transport response of hydrosystems is proposed. The code is written in R language and as such is made available for the wide increasing R-users community.

The paper is very well written and organized, the methods and mathematics used for the modeling approach are well described and the rationale supporting the overall approach is very accurately discussed. The model introduces new interesting concepts such as the spline smoothing approach applied to the IRF in order to reduce over-fitting (hence reducing computation time as well).

The paper is suitable and deserves publication in GMD, as it clearly is beneficial to the scientific hydrological community. However, the differences between the previous Long and Mahler (2013) paper should be clearly highlighted (technical description of the code, methodologies and approaches, applicability and versatility...).

---- Short answer: The author thanks the referee this helpful review. Redundancies with Long and Mahler (2013) have been either deleted, moved to an appendix, or summarized very briefly.

---- Long answer: The MS did not make it very clear that this is the first release of the RRAWFLOW code to the scientific and public communities. Long and Mahler (2013) presented it as a research model with no code or user's manual. The revised MS now clearly states in the abstract and introduction that the purpose of this paper is to present a new version of RRAWFLOW with added versatility, to make the code publicly available, and to guide users in its operation.

Equations 1-5 have been moved to an appendix. This was necessary because specific variables in Eqs. 1-5 are referred to throughout the MS and are included in RRAWFLOW outputs, and the RRAWFLOW download contains a user's manual that cites all equations in this manuscript. To avoid confusion for the user, I feel it will be most helpful if all equations are contained in the body or the appendix. Eq. 6 (now Eq. 1) is left in the main body because (1) I added one term to show how the equilibrium datum is applied in groundwater models, and (2) it pertains to Fig. 1, which is new in this MS to help explain convolution conceptually, which people seem to have trouble with. The related text has been shortened.

Time-variance options, although some were presented previously, need to be discussed briefly for the purpose of explaining the different RRAWFLOW settings and options referred to

in Table 1. The new continually changing IRF scale function needs to be presented also. An additional feature that I have included since submitting the MS is the option to have RRAWFLOW calculate the wet and dry periods on the basis of the precipitation record, which includes options for two different methods. Previously, the user had to supply the wet and dry periods. Therefore, a short section (section 2.6) was added to describe how these options work and the associated RRAWFLOW settings.

p.5928, l.15: Maybe mention that such terms are sometimes used improperly since for instance the transfer function actually refers to the Fourier transform of the IRF and that other terms are related to continuous-time instead of discrete-time functions.

---- Done

p.5928, l.20: I would remove these lines: I don't see the point of comparing convolution with spectral analysis. Convolution is one of many tools (like cross-correlation, autocorrelation, etc.) used in digital signal processing.

---- Done

p.5931, l.10: Since this is a methodological paper, it would be better to briefly describe the method used for determining the IRF by model calibration as employed by the same author in Long and Mahler (2013).

---- The reference to Long and Mahler (2013) was in error here. This paper did not present any nonparametric IRFs.

p.5931, l.13-18: Fourier transform-based deconvolution induces over-fitting because of the spectral signature of errors. Do the author think that filtering out high frequencies in the Fourier space previous to inverse transform be a good alternative to the spline interpolation approach in order to smooth the IRF by removing oscillations related to Fourier transform of the errors? Wouldn't such a Fourier filtering approach also lead to a good and consistent estimation of the IRF?

---- Good point. Since different filters would produce different IRFs, many different filters might need to be tested, which would be another form of model calibration. This would be an interesting approach. To briefly address this idea, I inserted the following text: "Filtering the IRF in the frequency domain (i.e., transfer function; Smith, 2003) or smoothing the IRF in the time domain (Long and Derickson, 1999) are options for IRF estimation by Fourier analysis, which may require trial-and-error calibration."

p.5931, l.24: Since optimization is based on a very limited number of points and spline (or linear) interpolating functions are used afterward to assess ordinates other than control points, what would be the influence of the interpolation scheme used (different

spline functions, spline vs linear interpolation) on the result of the modeling? Is there an appropriate minimum number of control points to consider?

---- Inserted text: "Trial and error generally is required to determine the optimum number of control points for a given application, and this number may be affected by the choice of a spline or linear interpolation. The minimum number of control points is two: at least one to define the non-zero part of the curve and one to define where the function becomes zero."

p.5935, l.5: Was 95% of the total area of the curve chosen arbitrarily or was it supported by some other criterion?

---- It is arbitrary and mainly useful for comparison of different sites. The word "arbitrarily" was inserted into the sentence for clarification.

p.5938, l.12-16: There are lots of packages devoted to optimization in CRAN repositories. The RRAWFLOW program would benefit using them in the future instead of external software like PEST.

---- Text inserted: "RRAWFLOW is a stand-alone model, independent of PEST, and therefore can be used with any optimization routine. For example, optimization packages are available in the comprehensive R archive network (CRAN) that could be built seamlessly into RRAWFLOW, possibly in future versions."

p.5942, l.9-13: Could a threshold value of the ratio E_{val}/E_{cal} be proposed to assess overfitting? I understand overfitting may result from the choice of, say, a double-gamma instead of a single-gamma IFR. But what if the double-gamma makes more physical meaning owing to experiencing or knowledge about the site tested?

---- I like this idea of the E_{val}/E_{cal} ratio. Indeed more research could be done to evaluate such a threshold. Text inserted: "The ratio E_{val}/E_{cal} might be a useful metric for comparison of different models and possibly in setting the lengths of the calibration and validation periods. As in any model, this all should be considered in reference to a physical understanding of the system; e.g., two gamma functions might be consistent with two distinct permeability domains."

p.5943, "code availability" section: Will the software be made available as a package through the comprehensive R archive network (CRAN)?

---- Not at this point. Text inserted: "The code is not yet available in CRAN but could be included in the future."

Marked manuscript follows:

1 RRAWFLOW: Rainfall-Response Aquifer and Watershed 2 Flow Model (v1.15)

3 A. J. Long¹

4 [1]{ U.S. Geological Survey, 1608 Mountain View Rd, Rapid City, South Dakota, USA }

5 Abstract

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

Deleted: RRAWFLOW

Deleted: included in the online supplement to the
article

Deleted: springflow

24 1 Introduction

Formatted: Heading 2

25 1.1 Lumped versus distributed models

26 Hydrologic models, commonly referred to as a “lumped-parameter” or “lumped” models,
27 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.
2 In contrast to lumped models, distributed models discretize the system into small
3 compartments or cells, each of which has several parameters defined. All hydrologic models,
4 however, are lumped to some degree. Models that frequently are considered physically based
5 simulate numerous small-scale physics by lumping these processes into simplified
6 mathematical forms (Beven, 1989). The use of the term “physically based” to describe any
7 hydrologic model, therefore, should be discouraged (Beven and Young, 2013). Both
8 distributed and lumped models, however, have components that can represent different
9 hydrologic processes that can be interpreted in physically meaningful ways (Beven and
10 Young, 2013). For example, the IRF estimated in many lumped models represents the
11 physical response to an impulse into the system and provides mechanistic insights into that
12 system, including the peak response time and magnitude and the hydrologic memory of the
13 system (von Asmuth and Knotters, 2004; Beven and Young, 2013; Young, 2013). The IRF
14 could be measured directly at the outflow point (e.g., a spring) if a short, intense recharge
15 event follows a long, dry period. Most commonly however, the outflow, or system response,
16 results from a series of superposed responses to repeating recharge events, and the lumped
17 model is used to estimate the IRF iteratively and to simulate the system response.

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

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

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

6 | **1.2 RRAWFLOW overview**

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

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

Deleted: I present a new version of RRAWFLOW with additional options that include

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

Deleted: and is included in the online supplement to this article.

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

Deleted: useful as an

Deleted: in the online supplement

17 2 The model

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

28 2.1 Precipitation recharge

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

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

Deleted: is not included in this case

10 2.2 Other recharge options

11 Recharge estimated outside of RRAWFLOW can be used as model input. For example, this
12 applies to precipitation recharge estimated by a soil-water-balance model (e.g., Westenbroek
13 et al., 2010) or sinking-stream recharge in karst aquifers that can be estimated by methods
14 such as those described by Hortness and Driscoll (1998). This is system-input option 2 (Table
15 1).

16 2.3 Convolution

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

$$24 \quad y_i = \Delta t \sum_{j=0}^i \beta_j h_{i-j} u_j + \varphi_i + d_0 \quad i, j = 0, 1, \dots, N \quad (1)$$

25 where h_{i-j} is the IRF; u_j is the input, or forcing function; j and i are time-step indices
26 corresponding to system input and output, respectively; N is the number of time steps in the
27 output record; β_j is an optional time-varying IRF scaling coefficient; φ_i represents the errors
28 resulting from measurement inaccuracy, sampling interval, or simplifying model assumptions;
29 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 frequency analysis, the frequency spectrum is a function that characterizes a single time series at a glance, whereas the IRF is a function that characterizes the 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:

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

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

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

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

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:

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

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.

Deleted: (Fig. 3),

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;

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.

Linearity and time variance

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

Time-variance (TV) options

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

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

Deleted: System linearity

Deleted: ; this is consistent with Jakeman and Homberger (1993), who described the response of streamflow to recharge as predominantly a linear process for a variety of watersheds of differing size

Deleted: For example, an increase in groundwater hydraulic gradient during a wet period generally results in an increase in groundwater velocity.

Deleted: This is applicable when different altitudes of the groundwater table result in different hydraulic gradients (e.g., toward a stream) or the saturation of different subterranean features with different hydraulic characteristics. Larocque et al. (1998), for example, observed that high-flow and low-flow periods could be identified by distinct hydraulic characteristics for a karst aquifer in France.

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

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

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-

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

7 **2.7 Model outputs**

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

16 **2.8 Evaluating model fit and over-fitting**

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

Deleted: . The

$$E = 1 - \frac{\left[\sum (y_{obs} - y_{sim})^2 \right]_p}{\left[\sum (y_{obs} - y_{mean})^2 \right]_T \left(\frac{l_p}{l_T} \right)} \quad (7)$$

where y_{obs} and y_{sim} are time series of the observed and simulated system responses, respectively; y_{mean} is the mean value of y_{obs} ; the subscripts p and T refer to the partial and total periods, respectively; and l is the time length of the respective period. Conceptually, E is the ratio of the magnitude of model residuals (numerator) to the overall variability in the observation record (denominator) subtracted from unity and theoretically can vary from $-\infty$ (poorest fit) to unity (perfect fit).

In addition to quantifying model fit, E provides a useful way to evaluate possible over-fitting of the model. Although model fit for the calibration period might improve as parameters are added, if the validation period indicates that this added complexity is not helpful, the model has been over-fit (von Asmuth et al., 2002). To test this condition, E is calculated for the calibration and conditional validation periods separately (E_{cal} and E_{val}) by the modified Nash–Sutcliffe coefficient of efficiency (Eq. 7), which makes E_{cal} and E_{val} directly comparable (Long and Mahler, 2013). This method is particularly important for comparison of two periods with different fluctuation amplitudes.

Legates and McCabe (1999) describe limitations of correlation-based measures to quantify model fit, such as the coefficient of determination R^2 , and the benefits of the Nash–Sutcliffe coefficient of efficiency and the index of agreement. The sum of the squared and weighted residuals (Doherty, 2005) is another useful metric used for this purpose. Hartmann, et al. (2013) provides an example of using multiple metrics to evaluate model performance.

A value of E_{cal} that is much larger than E_{val} might indicate over-fitting, in which case a simpler model (i.e., fewer fitting parameters) should be tested. For example, if a double-gamma IRF is used, then a second model calibration with a single-gamma IRF could be tested to determine if greater similarity in the E_{val} and E_{cal} values is achieved. For nonparametric IRFs, a reduction in the number of IRF control points could be tested. A time-variant IRF requires more parameters than a time-invariant IRF, and this also can be tested. Thoroughly considering model complexity in light of E_{val} and E_{cal} provides context for conditional 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 as the simulation (y_{sim}) (Legates and McCabe, 1999). Because E compares the residuals to the overall response variability, E values for different simulation sites can be directly compared; this would not be the case if the magnitude of the residuals alone is used in comparisons; e.g., as with the root mean square error.

Deleted: the coefficient of efficiency the coefficient 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 shorter period of interest (Long and Mahler, 2013). The denominator of Eq. (13) is scaled down to be consistent with the time period of the numerator where the subscript p refers to the partial period, either calibration or conditional validation; the subscript T refers to 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 possible

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

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

Deleted: related to the number of parameters

19 3 Example model applications

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

Deleted: two

Deleted: 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 of these sites. The precipitation records were separated into wet and dry periods, which were defined as years in which the annual precipitation was either above or below the long-term mean, respectively. The examples use a daily time step.

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

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

20 3.1 Streamflow in Boxelder Creek

21 Boxelder Creek is located in the Black Hills of South Dakota, USA, with a watershed area of
22 250 km² upstream from U.S. Geological Survey streamgage 06422500, with daily streamflow
23 available from <http://waterdata.usgs.gov/nwis>. The watershed primarily is pine forest and
24 contains metamorphic rocks of Precambrian age (Carter et al., 2001). Gridded daily
25 precipitation and air temperature data from the Daymet dataset are available at 1-km grid
26 spacing for 1980–2013. These data were obtained from the Geo Data Portal
27 (<http://cida.usgs.gov/gdp/>) and spatially averaged for the watershed to produce a daily time
28 series of precipitation and air temperature for 1980–2013, which was used as model input.
29 The calibration period was 1980–1996, and validation was for 1997–2013 (Fig. 4). A 5-year
30 model spin-up period was applied by inserting data for 1980–1984 into the period 1975–1979.

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

3 The example models described used system-input option 1 (precipitation recharge, Table 1).
4 Five example models are presented for Boxelder Creek, all of which used single-gamma IRFs
5 (Table 2). All other trials with double-gamma IRFs resulted in minimization of one of the
6 IRFs, indicating that single-gamma IRFs were appropriate for this system. All Boxelder Creek
7 gamma functions optimized to $\eta < 1$, which is the reverse-J shape (Fig. 5). Example BC1 used
8 TV option 1 (time-invariant IRF, Table 1), resulting in E_{cal} and E_{val} values of 0.62 and 0.46,
9 respectively (Table 2). Examples BC2 and BC3 used TV option 2 (time-variant IRF) with
10 wet-dry options 1 and 2, respectively. For BC3, E_{val} was higher (0.56) than for BC1 and BC2
11 (Table 2). Example BC4 and BC5 used TV option 3, in which a time-variant IRF that changes
12 continually was used with a single-gamma IRF and moving-average (MA) windows of 1 and
13 10 years (Eq. 6), respectively. Of the five examples, BC3 had the highest E_{val} value; BC5 had
14 the second highest E_{val} value and with fewer parameters (Table 2). Comparison of model fit
15 for examples BC4 and BC5 indicates that the time-variant aspects of this system respond to
16 general climatic changes over decadal periods more so than annual. Table 3 shows optimized
17 IFR parameters for selected example models.

18 3.2 Springflow from Barton Springs

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

Deleted: The collective flow from these springs was simulated in these examples. The springs respond to surface recharge to the Edwards aquifer that occurs from multiple sinking streams that cross onto the aquifer's recharge area and precipitation recharge (Mahler et al., 2008).

Deleted: Although recharge for Barton Springs also includes sinking-stream recharge, option 1 provided an approximation of total recharge.

Deleted: These examples also serve to describe model settings for simulation of streamflow, because the model settings for springflow and streamflow are identical.

Deleted: Therefore, the type of model applied and the number of parameters needed for the selected model are important considerations.

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

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

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

19 3.3 Groundwater level in well LA88C

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

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

Deleted: Examples F5–F9 used TV option 2, in which time variance is applied as different IRFs for wet and dry periods. Example F5 used single-gamma IRFs (Figs. 5 and 6), resulting in E_{cal} and E_{val} values of 0.88 and 0.72, respectively (Table 2), which represent an improvement in model fit compared to example N1. Example F7 used double-gamma IRF (Fig. 7), resulting in E_{cal} and E_{val} values of 0.90 and 0.63, respectively (Table 2); the added parameters for this example resulted in a higher E_{cal} value but lower E_{val} value than for example F5, indicating that example F5 is the better choice for hydrologic projections and that example F7 was over-fit. Example F8 used nonparametric IRFs with a spline curve fit to 4 and 3 optimized (non-zero) control points for the wet and dry periods, respectively (Fig. 8a; $E_{cal} = 0.88$, $E_{val} = 0.72$), which is the same model fit as for example F5. Example F9 is the same as example F8, except with 8 optimized control points for each of the two periods (Fig. 9), which resulted in a higher E_{cal} (0.92) but a much lower E_{val} (0.61) than for example F8, indicating an over-fit model.

Deleted: For these examples, E_{cal} and E_{val} values were equal to or smaller than all other time-variant examples (Table 2).

Deleted: ; this, and the fact that the wet and dry periods defined for examples F5–F9 generally are multi-year periods,

Deleted: and is contained within the regionally extensive Madison Limestone

Deleted: This formation is exposed at the land surface on all flanks of the Black Hills of South Dakota and Wyoming and dips radially outward in all directions below the land surface. Water level in the well responds to surface recharge to the Madison aquifer that occurs from direct precipitation and sinking streams that cross onto the aquifer's recharge area (Carter et al., 2001).

Deleted: As with Barton Springs, all E_{cal} values were larger than E_{val} (Table 2).

Deleted: Example W1 used double-gamma IRF (Figs. 10 and 11). Example W12 used nonparametric IRFs with spline curves fit to 10 and 8 optimized control points for the wet and dry periods, respectively (Fig. 12; $E_{cal} = 0.93$, $E_{val} = 0.70$).

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.

Deleted: Example W13 used nonparametric IRF with spline curves fit to 5 optimized control points for each of the two periods (Fig. 13; $E_{\text{cal}} = 0.88$, $E_{\text{val}} = 0.75$), resulting in a higher E_{val} value than for example W12, which indicates that the smaller number of control points is the better choice in this case.

Deleted: Example W13 also resulted in the smallest E_{cal} value of the three examples, which is not a disadvantage when the primary objective is to conditionally validate the model.

4 Discussion and conclusions

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.

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 parameters for the gamma function is an advantage over the exponential function.

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

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

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

Deleted: IRFs for streamflow commonly have a slow-flow component with a much smaller peak response and much longer tail than the quick-flow component, as in Fig. 3.

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

1 The record length of the observed response should be considered in light of the system
2 memory: there is less confidence in the predictive strength of a model if the observed
3 response is shorter than the system memory than if it is longer, because, in the former case,
4 the effects of the IRF tail are not fully tested against observation. Ideally, the validation
5 period alone should be longer than the system memory, and if it is several times longer, then
6 the full range of the IRF is tested several times over.

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

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

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

This study was supported by the Department of Interior South Central Climate Science Center. I thank Jeff Raffensperger and Janet Carter of the U.S. Geological Survey (USGS) for their constructive review comments in an early draft of this manuscript as well as the reviewers selected by Geoscientific Model Development. Although the R language has been used by the USGS, no warranty, expressed or implied, is made by the USGS or the U.S. Government as to the accuracy and functioning of the program and related program material, nor shall the fact of distribution constitute any such warranty, and no responsibility is assumed by the USGS in connection therewith.

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

To simulate recharge from direct precipitation, a soil-moisture index s [unitless] is estimated for each time step in RRAWFLOW. Quantitatively, s is the fraction of precipitation that infiltrates and becomes recharge. To account for the antecedent effects of rainfall on soil moisture, the past rainfall record is weighted by a backward-in-time exponential decay function (Jakeman and Hornberger, 1993):

$$\begin{aligned} s_i &= cr_i + (1 - \kappa_i^{-1})s_{i-1} \\ &= c \left[r_i + (1 - \kappa_i^{-1})r_{i-1} + (1 - \kappa_i^{-1})^2 r_{i-2} + \dots \right] \end{aligned} \quad (\text{A1})$$

$$i = 0, 1, \dots, N \quad 0 \leq s \leq 1$$

where c [L^{-1}] is a **scaling coefficient to constrain the value of s** ; κ [unitless] adjusts the effect of antecedent rainfall and is related to evapotranspiration; r is total rainfall [L]; and i is the time step. In RRAWFLOW, this method is option 1 for system input (Table 1). For watershed modeling, the value of c can be set to satisfy the assumption that the total recharge volume within a watershed is equal to the total outflow volume for the calibration period. This assumption neglects the net change in total watershed storage during this period, which is assumed to be small in comparison to the total inflow or outflow for the same period. Also, this assumption does not apply if recharge to the watershed exits the watershed through deep groundwater and bypasses the stream outlet. Recent rainfall has the largest effect on s in Eq. (A1), whereas earlier rainfall has the least effect.

The effect of changing air temperatures on evapotranspiration is accounted for by (Jakeman and Hornberger, 1993):

$$\kappa_i = \alpha \exp[(20 - T_i)f] \quad f > 0, \quad (\text{A2})$$

where α [unitless] is a scaling coefficient; T [$^{\circ}\text{C}$] is mean air temperature at the land surface; and f is a temperature modulation factor [$^{\circ}\text{C}^{-1}$]. As air temperature T decreases, s in Eq. (A1) increases with sufficient past rainfall. RRAWFLOW can be executed without air-temperature data if unavailable (air-temperature option 2 in Table 1). Recharge for each time step u_i [L] is calculated as the fraction s of precipitation by

Deleted: \geq

Deleted: \geq

Deleted: coefficient that limits s to the range 0-

Deleted: , typically in days

1 $u_i = r_i s_i$ (A3)

2 Typically, s is largest during wet periods and rarely reaches a maximum value of 1.0 (Fig.
3 14).

4 **Considerations for parameters c and κ**

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

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

21 **Snow precipitation**

22 For cold climates where winter snowfall is common, a method proposed by Long and Maher
23 (2013) is applied. To determine the form of precipitation for each time step, an air
24 temperature threshold value T_s is set, below which precipitation is assumed to occur as snow
25 (typically $T_s = 0^\circ\text{C}$). To determine time steps when melting occurs, a melting threshold value
26 T_m is set. If daily snow-depth data are available, T_m can be determined empirically as the mean
27 air temperature for time steps when snow depth decreases to zero from a previous time step
28 with a snow depth greater than zero. Long and Mahler (2013) determined that $T_m = 9^\circ\text{C}$ for a

Deleted: \geq

Deleted: \geq

Deleted: $0 \geq s \geq 1$

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

$$4 \quad p_m = (1 - S_f) \sum_{i=1}^N p_i \quad T_i < T_s \quad (A4)$$

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

10
11
12

Deleted: <#>Convolution¶

Convolution is a time-series operation (Jenkins and Watts, 1968; Smith, 2003) commonly used in non-distributed hydrological models to simulate streamflow, springflow, or groundwater level in response to recharge (e.g., Nash, 1959; Dooge, 1968; Dreiss, 1989; Olsthoorn, 2008). The use of convolution in hydrology has its origins in unit-hydrograph theory for watershed runoff modeling (Nash, 1959; Dooge, 1973). Convolution has been used to simulate solute transport of tracers (e.g., Maloszewski and Zuber, 1982; Long and Putnam, 2004; Jurgens et al., 2012), groundwater response to recharge from a sinking stream (Long and Derickson, 1999), flow through karst conduits (Comaton and Perrochet, 2002), and cave drip (Long and Mahler, 2013). Mathematically, convolution is the integration of an input function (e.g., recharge) with an IRF that describes the system response to unit impulse of the input function: ¶

$$y(t) = \int_{\tau=0}^{\infty} h(t-\tau)u(\tau)d\tau \quad (A5) \quad ¶$$

where $y(t)$ is the system response, or output; $h(t-\tau)$ is the IRF; $u(\tau)$ is the input, or forcing, function; τ and t are time variables corresponding to system input and output, respectively (Jenkins and Watts, 1968; Dooge, 1973; Olsthoorn, 2008). The quantity $t-\tau$ represents the delay time from impulse to response, and the IRF represents a distribution of these delay times. In RRAWFLOW, the input function $u(\tau)$ can be recharge or input of a solute. The system response $y(t)$ can be streamflow exiting watershed, springflow from a groundwater system, groundwater level, or solute concentration at an outlet. Physically, the IRF is the system response per unit impulse of $u(\tau)$ and also can be described as the response produced by a system when the input is a delta function (Smith, 2003). Conceptually, convolution is the superposition of a series of IRFs that are initiated at the time of each impulse of $u(\tau)$ and are scaled proportionally by the magnitude of the corresponding impulse (Fig. 1). ¶

For uniform time steps, the discrete form of Eq. (A5) is ¶

$$y_i = \sum_{j=0}^i h_{i-j}u_j + \phi_i \quad i, j = 0, 1, \dots, N \quad (A-6) \quad ¶$$

where h_{i-j} is the IRF; u_j is the input, or forcing, function; and 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; and ϕ represents the errors resulting from measurement inaccuracy, sampling interval, simplifying model assumptions. For simulation of groundwater levels, a datum h_0 at which hydraulic head equals zero must be established. Conceptually, h_0 is the level to which hydraulic head would converge if the local recharge was eliminated. Local recharge is assumed to be the only forcing that results in hydraulic-head fluctuation or that causes hydraulic head to rise above h_0 .

References

- 1
2 Arkesteijn, L. and Pande, S.: On hydrological model complexity, its geometrical
3 interpretations and prediction uncertainty, *Water Resour. Res.*, 49, 7048-7063, 2013.
- 4 Bailly-Comte, V., Martin, J. B. and Sreaton, E. J.: Time variant cross correlation to assess
5 residence time of water and implication for hydraulics of a sink-rise karst system,
6 *Water Resour. Res.*, 47, 2011.
- 7 Berendrecht, W. L., Heemink, A. W., Van Geer, F. C. and Gehrels, J. C.: Decoupling of
8 modeling and measuring interval in groundwater time series analysis based on
9 response characteristics, *Journal of Hydrology*, 278, 1-16, 2003.
- 10 Besbes, M. and de Marsily, G.: From infiltration to recharge: use of a parametric transfer
11 function, *Journal of Hydrology*, 74, 271-293, 1984.
- 12 Beven, K.: Changing ideas in hydrology — The case of physically-based models, *Journal of*
13 *Hydrology*, 105, 157-172, 1989.
- 14 Beven, K. and Young, P.: A guide to good practice in modeling semantics for authors and
15 referees, *Water Resour. Res.*, 49, 5092-5098, 2013.
- 16 Blank, D., Delleur, J. W. and Giorgini, A.: Oscillatory kernel functions in linear hydrologic
17 models, *Water Resour. Res.*, 7, 1102-1117, 1971.
- 18 Carter, J.M., Driscoll, D.G., Hamade, G.R., Estimated recharge to the Madison and Minnelusa
19 aquifers in the Black Hills area, South Dakota and Wyoming, water years 1931-98, US
20 Geol. Surv., Reston Virginia, Water-Resources Investigations Report 00-4278, 66 pp.,
21 2001.
- 22 Delleur, J. W. and Rao, R. A.: Linear systems analysis in hydrology: the transform approach,
23 kernel oscillations and the effect of noise, in: *Systems Approach to Hydrology*, Water
24 Resources Publications, Fort Collins, Colorado, 116-142, 1971.
- 25 Denić-Jukić, V. and Jukić, D.: Composite transfer functions for karst aquifers, *Journal of*
26 *Hydrology*, 274, 80-94, 2003.
- 27 Doherty, J.: PEST—Model-Independent Parameter Estimation, User Manual (5th ed.),
28 Watermark Numerical Computing, Brisbane, Australia,
29 <http://www.pesthomepage.org/>, 2005.
- 30 [Doherty, J.E., Hunt, R.J., and Tonkin, M.J.: Approaches to highly parameterized inversion: A](#)
31 [guide to using PEST for model-parameter and predictive-uncertainty analysis, U.S.](#)

1 [Geological Survey Scientific Investigations Report 2010–5211, 71 p.,](#)
2 [Hartmann, A., Wagener, T., Rimmer, A., Lange, J., Brielmann, H. and Weiler, M.: Testing](http://pubs.usgs.gov/sir/2010/5211, 2010.</p><p>3 Dooge, J. C. I.: Linear theory of hydrologic systems, Washington, D.C., 1468 pp., 1973.</p><p>4 Dreiss, S. J.: Regional scale transport in a karst aquifer: 2. Linear systems and time moment
5 analysis, <i>Water Resour. Res.</i>, 25(1), 126-134, 1989.</p><p>6 Fienen, M. N., Doherty, J. E., Hunt, R. J. and Reeves, H. W.: Using prediction uncertainty
7 analysis to design hydrologic monitoring networks: example applications from the
8 Great Lakes water availability pilot project, US Geol. Surv., Reston Virginia,
9 Scientific Investigations Report 2010–5159, 44 pp., 2010.</p><p>10 Haan, C. T.: Statistical Methods in Hydrology, Second ed., The Iowa State University Press,
11 Ames, Iowa, 2002.</p><p>12 <a href=)
13 [the realism of model structures to identify karst system processes using water quality](#)
14 [and quantity signatures, *Water Resour. Res.*, 49, 1-14, doi:10.1002/wrcr.20229, 2013.](#)

15 Hortness, J. E. and Driscoll, D. G.: Streamflow losses in the Black Hills of western South
16 Dakota, US Geol. Surv., Reston Virginia, Water-Resources Investigations Report 98-
17 4116 , 99 pp., 1998.

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

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

22 Jukić, D. and Denić-Jukić, V.: Nonlinear kernel functions for karst aquifers, *J. Hydrol.*, 328,
23 360-374, 2006.

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

28 [Konikow, L. F. and Bredehoeft, J. D.: Ground-water models cannot be validated, *Adv. Water*](#)
29 [Resour., 15, 75-83, 1992.](#)

30 Ladouche, B., Marechal, J. and Dorfliger, N.: Semi-distributed lumped model of a karst
31 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.

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

3 [Rimmer, A. and Hartmann, A.: Optimal hydrograph separation filter to evaluate transport](#)
4 [routines of hydrological models, *Journal of Hydrology*, 514, 249-257,](#)
5 [doi:10.1016/j.jhydrol.2014.04.033, 2014.](#)

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

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

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

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

24 von Asmuth, J. R., Bierkens, M. F. P. and Maas, K.: Transfer function-noise modeling in
25 continuous time using predefined impulse response functions, *Water Resour. Res.*, 38,
26 231-2312, 2002.

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

29 Westenbroek, S. M., Kelson, V. A., Dripps, W. R., Hunt, R. J. and Bradbury, K. R.: SWB - A
30 modified Thornthwaite-Mather soil-water-balance code for estimating groundwater
31 recharge, US Geol. Surv., Reston Virginia, Techniques and Methods 6-A31, 60 pp.,
32 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
2

Tables

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

3
 4
 5

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).