SCIENTIFIC BRIEFING

HYDROLOGICAL PROCESSES
Hydrol. Process. (2011)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hyp.8185

A synthetic hydrologic-response dataset

Benjamin B. Mirus,[1]*
Keith Loague,[1]
Nicoleta C. Cristea,[2]
Stephen J. Burges,[2]
Stephanie K. Kampf[3]

[1] *Department of Geological and Environmental Sciences Stanford University Stanford, CA 94305-2115, USA*
[2] *Department of Civil and Environmental Engineering University of Washington, Seattle, WA 98195-2700, USA*
[3] *Natural Resources Ecology Laboratory Colorado State University Fort Collins, CO 80523-1499, USA*

*Correspondence to:
Benjamin B. Mirus, U.S. Geological Survey, 345 Middlefield Rd, MS 420, Menlo Park, CA 94025-3561, USA.
E-mail: bmirus@stanford.edu

Received 28 January 2011
Accepted 25 May 2011

Abstract

Synthetic data have long been employed in hydrology for model development and testing. The objective of this study was to generate a synthetic dataset of hydrologic response with higher spatial and temporal resolution than could presently be obtained in the field, spanning a longer period than the typical duration of monitoring campaigns in experimental catchments. The synthetic dataset was generated for a rangeland catchment with the Integrated Hydrology Model (InHM), and is presented for future use by the community. The InHM boundary-value problem is based upon the previously reported hypothetical reality of Tarrawarra-like hydrologic response. Whereas the emphasis in developing the hypothetical reality was on parameterising InHM to reproduce observations from the Tarrawarra catchment, the emphasis in generating the synthetic dataset is on developing an internally valid hydrologic-response dataset that extends well beyond the period of observations at Tarrawarra. The synthetic dataset spans 11 years of continuous forcing and response data (e.g. integrated response, distributed fluxes, state variable dynamics). The dataset should be useful for a wide range of problems including evaluation of simple rainfall runoff modelling techniques, design of measurement networks, development of data-assimilation algorithms, and studies on information theory. The dataset is available at: ftp://pangea.stanford.edu/pub/loague/. Copyright © 2011 John Wiley & Sons, Ltd.

Introduction

There are well known problems associated with the observation and measurement of hydrologic response. Any set of hydrologic observations will face issues with measurement errors, gaps in records, resolution limitations, and the small support volume of most instruments. Additionally, detailed hydrologic datasets such as those described by Western and Grayson (1998), Slaughter *et al.* (2001), Ebel *et al.* (2007a), and Heppner and Loague (2008) require extensive, long-term field campaigns that are often prohibitively expensive. As a result, catchment-scale hydrologic-response datasets with high spatial and temporal resolution are rare, and continuous long-term monitoring of distributed fluxes and state variable dynamics in experimental catchments is uncommon. Synthetic data have been used in previous studies to provide an additional source of hydrologic process information (Table I). Synthetic datasets have proven useful for hypothesis testing in situations where the necessary observations are scare or unavailable. For example, Loague and Abrams (2001) employed synthetic data to illustrate the dynamics of the Horton and Dunne overland flow mechanisms for a hypothetical reality based on the R-5 catchment. Synthetic datasets provide an immediate solution to certain shortcomings related to the current practical restrictions on the resolution and duration of field observations in experimental catchments.

Sophisticated numerical models of hydrologic response based on the known physics of surface and subsurface flow require careful field measurements to produce meaningful simulation results (Ebel and Loague, 2006). Our experiences with comprehensive physics-based simulation have highlighted the importance of designing model boundary-value problems using a foundation of continuous observational records (e.g. VanderKwaak and Loague, 2001; Loague and VanderKwaak, 2004; Loague *et al.*, 2005;

Table I. Characteristics of selected synthetic datasets in hydrology

Reference	Model	Application
Freeze (1972a,b)	Deterministic	Concept development
Wood (1976)	Deterministic	Parameter uncertainty
Freeze (1980)	Stochastic	Concept development
Smith and Hebbert (1983)	Deterministic	Concept development
Milly and Eagleson (1988)	Deterministic	Concept development
Gan and Burges (1990)	Deterministic	Model testing
Troch *et al.* (1993)	Deterministic	Model testing
Loague and Abrams (2001)	Stochastic	Concept development

Ebel and Loague, 2006,2008; Heppner *et al.*, 2007; Ebel *et al.*, 2007b,2008,2010; Mirus *et al.*, 2007,2009,2011; Heppner and Loague, 2008). In cases when sufficient observational details are available, a sophisticated physics-based model provides a foundation for generating realistic, internally valid synthetic data for a range of possible climatic forcing conditions. Unfortunately, the extensive datasets of both integrated and distributed hydrologic response required to support the parameterisation and rigorous evaluation of catchment-scale boundary-value problems are rare.

The scarcity of distributed hydrologic-response data limits the use of physics-based models for operational hydrology. As a result, simpler rainfall-runoff models offer the quantitative support for designing policy in the decision management arena when few data are available. However, the dearth of distributed datasets at the catchment scale has also limited the number and breadth of unbiased tests that compare and evaluate simplified rainfall-runoff modelling approaches. Further testing is needed to determine which field measurements provide the most useful information for improving model performance and assessing the utility of different underlying modelling techniques for given applications (e.g. Loague and Freeze, 1985). Internally valid synthetic data generated using comprehensive physics-based models can provide information needed for testing simpler models outside their calibrated ranges, provided they span a sufficient duration to cover a wide range of conditions. Simulated hydrologic response can also be employed to examine issues related to data worth and design of more effective measurement networks (Ebel and Loague, 2008), develop algorithms for data assimilation (Dunne and Entekhabi, 2005; Zhou *et al.*, 2008), explore problems related to information theory (Mogheir *et al.*, 2004), and test techniques for upscaling and downscaling soil moisture (Crow *et al.*, 2000; Vereecken *et al.*, 2007; Kaheil *et al.*, 2008).

The objective of this work is to establish and distribute an internally valid, long-duration hydrologic-response dataset with greater spatial and temporal resolution than current field measurement capabilities allow. The synthetic dataset, heretofore referred to as the Synthetic Rangeland 1 (SR-1) dataset, was generated using the physics-based InHM (VanderKwaak, 1999). Parameterisation of InHM herein employs the Tarrawarra-like hypothetical reality (Mirus *et al.*, 2009) as a foundation. The SR-1 dataset extends the six-month hypothetical reality to more than a decade of comprehensive, continuous hydrologic response with high spatial and temporal resolution.

Approach

InHM was developed to simulate fully coupled 3D variably saturated subsurface flow and 2D surface flow. InHM was selected to generate this SR-1 dataset because it represents the known hydrologic-response mechanisms with no a priori specification. InHM has been applied at the catchment scale to simulate observed hydrologic response for a range of environmental conditions from steep forested catchments (Ebel *et al.*, 2007b; Mirus *et al.*, 2007) to gently sloping rangeland catchments (Heppner *et al.*, 2007; Mirus *et al.*, 2009). The equations and numerical methods employed by InHM are described by VanderKwaak (1999).

The gently sloping 10·5 hectare rangeland catchment, known as Tarrawarra, is located in southeastern Australia (Western and Grayson, 1998). The Tarrawarra catchment was selected as the foundation for developing the hypothetical reality and SR-1 dataset because of the extensive distributed measurements of hydrologic state variables and relatively long record of continuous forcing data. For the hypothetical reality, InHM was parameterized to capture the observed record of surface runoff, soil-water content, and subsurface pressure head dynamics available for Tarrawarra during the Austral winter of a relatively wet year (1996). The SR-1 dataset is distinguished from the Tarrawarra-like hypothetical reality by including the seasonal wetting and drying of the catchment for a range of wet and dry years where no field observations are available. The observations at Tarrawarra and the best-fit InHM parameterisation are provided by Western and Grayson (1998), and Mirus *et al.* (2009), respectively.

The SR-1 dataset relies upon the physics represented by InHM, the observed hydrologic response at Tarrawarra, and the continuous climatic records in the region to both fill in missing data and extend the hydrologic-response data further. Collected during the period between September 1995 and November 1997, the record of observations at Tarrawarra is sporadic. The SR-1 dataset spans 1 January 1996 through 31 December 2006 and comprises spatially and temporally dense information. For example, between 1995 and 1997, 13 surveys of soil moisture patterns are reported for Tarrawarra, each based on roughly 600 soil-water content measurements taken over the course of one day. In contrast, the SR-1 dataset includes snapshots of surface water depths at

1,335 locations and pressure head/volumetric water content at 73,425 locations every half hour for individual storms. The emphasis in generating the SR-1 dataset is on providing realistic catchment-scale hydrologic-response data with higher spatial and temporal resolution than could be obtained with current field measurement systems.

Continuous precipitation and meteorological measurements from the 11-year period were used to drive the continuous InHM simulation. Precipitation intensities at 30-min intervals were taken from locally observed records (Australian Bureau of Meterology, 2007). Potential evapotranspiration was calculated using the Penman-Monteith equation (Shuttleworth WJ. 1993) with measurements of relative humidity, wind speed, and net radiation taken from climate records at Tarrawarra for 1996 through 1999 (Andrew Western, personal communication, 2007) and a nearby weather station for 2000 through 2006 (Australian Bureau of Meteorology, 2007). Negative evapotranspiration estimates were set equal to zero. Gaps in the meteorological records (between three and nine hours) were interpolated linearly between the measurements immediately before and after the missing periods.

Synthetic Dataset

The SR-1 dataset comprises the forcing data (precipitation and potential evapotranspiration) and the InHM simulated output, which includes both the integrated response (discharge hydrograph) and distributed response (internal state variables and fluxes). The discharge hydrograph is at the catchment outlet (Figure 1). The distributed response is in the form of (i) snapshots of internal state variables and fluxes throughout the entire domain at selected output times, and (ii) continuous time series of state variables and fluxes at 11 selected locations throughout the domain (Figure 1). Vertical profiles at each of the 11 locations include the surface node and subsurface nodes at approximately 0·02, 0·1, 0·5, and 1 m depths. The state variables given in the SR-1 dataset are water depths on the surface and pressure heads and volumetric water contents in the subsurface. The fluxes given are surface water and groundwater velocity vectors, surface–subsurface exchanges, and boundary fluxes. The dataset includes periods with shorter output time intervals to highlight the hydrologic response during the 11 storms summarized in Table II. These 11 storms represent a range from shorter to longer duration and lower to higher peak rainfall intensities. Output times during the selected storms are at 30-min intervals. Between the selected storms the output times for the continuous simulations were based on the adaptive time-step employed by InHM. Table III provides a summary of the hydrologic forcing and response data. While the InHM simulation was continuous for the entire 11-year period, the output time series have been broken up into periods of 3-month duration as well as by individual storms to limit the file size and facilitate manipulation of the data. The files are provided in tab-delimited text format (.dat) and are openly available via the ftp site: ftp://pangea.stanford.edu/pub/loague/.

Discussion

The work presented here differs from previous synthetic datasets in that it uses a comprehensive physics-based numerical model to produce a continuous record of hydrologic response at the catchment scale (Table I). The hydrologic realism of the SR-1 dataset relies on the physics upon which InHM rests, the hydrologic-response observations from Tarrawarra, the resulting best-fit, time-invariant parameterisation of InHM, and the observed climate records in the region. The SR-1 dataset employs effective hydraulic properties for each hydrogeologic unit to represent the average impact of heterogeneity, hysteresis, preferential flow, and evapotranspiration on catchment-scale hydrologic response. As a result, the SR-1 data should be used with the understanding that it is influenced by these assumptions regarding the flow physics and hydraulic properties.

We are not aware of any hydrologic-response dataset, synthetic or real, with comparable resolution, duration,

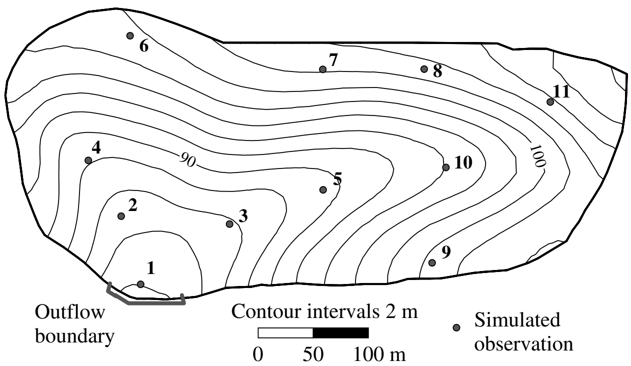

Figure 1. The boundary-value problem for the SR-1 dataset, showing the topography, locations of the observation node profiles, and the down-gradient subsurface/surface outflow boundary for the otherwise closed catchment

Table II. Characteristics of the selected storms

Start date (mm/dd/yyyy)	Total depth (mm)	Storm duration (h)	Maximum intensity (mm h^{-1})	Time to maximum intensity (h)
04/11/1996	135	444	12	179
06/23/1996	49	48	13	34
09/03/1996	30	108	6	57
06/25/1998	49	99	5	92
07/29/1998	29	135	4	21
02/28/1999	48	62	27	49
08/27/1999	34	103	5	12
07/26/2003	107	239	6	39
11/13/2004	53	64	9	33
09/12/2005	60	185	8	140
09/29/2005	22	52	21	7

Table III. Summary of the SR-1 dataset

Characteristic	Units	Type of data [a]	Temporal	Measurements resolution	Comments per output time
Precipitation	$m\ s^{-1}$	U, FD	0·5 hour	1	Observed, Tarrawarra
Evapotranspiration [b]	$m\ s^{-1}$	U, FD	0·5 hour	1	Estimated, Penman-Monteith
Volumetric-water content	$m^3\ m^{-3}$	D, SV	variable	73,425	InHM output
Pressure head	m	D, SV	variable	73,425	InHM output
Surface water depth	m	D, SV	variable	1,335	InHM output
Subsurface velocities	$m\ s^{-1}$	D, FV	variable	73,425	InHM output
Surface water velocities	$m\ s^{-1}$	D, FV	variable	1,335	InHM output
Surface/subsurface exchange	—	D, FV	variable	1,335	InHM output
Boundary fluxes[c]					
Subsurface, base [d]	$L\ s^{-1}$	D, FV	variable	1,335	InHM output
Subsurface, down gradient [d]	$L\ s^{-1}$	D, FV	variable	1,026	InHM output
Surface outflow [e]	$L\ s^{-1}$	I, FV	variable	1	InHM output

[a] U—uniform; I—integrated; D—distributed; SV—state variable; FD—Forcing data; FV—Flux vector.
[b] Saturation-limited, depth-distributed potential evapotranspiration over top 0·40 m.
[c] Closed catchment, with fluxes out the base and down-gradient boundary.
[d] Radiation flux boundary condition, calculates fluxes according to upstream hydraulic head gradient.
[e] Critical depth boundary condition, discharge value is summed over 19 boundary nodes.

and breadth to the one described here. Employed as a surrogate for real information, the SR-1 dataset can be sampled from to calibrate simpler models, optimize measurement networks, and develop data-intensive algorithms for upscaling, downscaling, and data assimilation. The SR-1 dataset provides a unique opportunity for testing simpler models against spatially and temporally dense information. The SR-1 dataset includes sufficient information to test empirical models such as regressions and unit hydrographs, conceptual models based on topographic indices or bucket assumptions, and physical approximations such as kinematic wave approaches for routing surface or subsurface flow.

Acknowledgements

This work was supported, in part, by National Science Foundation Grant No. EAR-0537410. Andrew Western provided additional meteorological data; Greg Williams compiled the data used in calculating potential evapotranspiration. Thanks, also, to the School of Earth Sciences at Stanford University for facilitating the distribution of the SR-1 dataset. Comments from two anonymous reviewers helped improve the manuscript.

References

Australian Bureau of Meterology. 2007. http://www.bom.gov.au/index.shtml, verified 5/17/2011.

Crow WT, Wood EF, Dubayah R. 2000. Potential for downscaling soil moisture maps derived from spaceborne imaging radar data. *Journal of Geophysical Research* **105**: 2203–2212.

Dunne S, Entekhabi D. 2005. An ensemble-based reanalysis approach to land data assimilation. *Water Resources Research* **41**: W02013, DOI:10·1029/2004WR003449.

Ebel BA, Loague K. 2006. Physics-based hydrologic response simulation: Seeing through the fog of equifinality. *Hydrological Processes* **20**: 2887–2900, DOI:10·1002/hyp.6388.

Ebel BA, Loague K. 2008. Rapid simulated hydrologic response within the variably saturated near surface. *Hydrological Processes* **22**: 464–471, DOI:10·1002/hyp.6926.

Ebel BA, Loague K, Borja RI. 2010. The impacts of hysteresis on variably-saturated hydrologic response and slope failure. *Environmental Earth Sciences* **61**: 1215–1225.

Ebel BA, Loague K, VanderKwaak JE, Dietrich WE, Montgomery DR, Torres R, Anderson SP. 2007a. Near-surface hydrologic response for a steep, unchanneled catchment near Coos Bay, Oregon: 1. Sprinkling experiments. *American Journal of Science* **307**: 678–708, DOI:10·2475/04·2007·02.

Ebel BA, Loague K, VanderKwaak JE, Dietrich WE, Montgomery DR, Torres R, Anderson SP. 2007b. Near-surface hydrologic response for a steep, unchanneled catchment near Coos Bay, Oregon: 2. Physics-based simulations. *American Journal of Science* **307**: 709–748, DOI:10·2475/04·2007·03.

Ebel BA, Loague K, Montgomery DR, Dietrich WE. 2008. Physics-based continuous simulation of long-term near-surface hydrologic response for the Coos Bay experimental catchment. *Water Resources Research* **44**: W07417, DOI:10·1029/2007WR006442.

Freeze RA. 1972a. Role of subsurface flow in generating surface runoff: 1. Base flow contributions to channel flow. *Water Resources Research* **8**: 609–623.

Freeze RA. 1972b. Role of subsurface flow in generating surface runoff: 2. Upstream source areas. *Water Resources Research* **8**: 1272–1283.

Freeze RA. 1980. A stochastic-conceptual analysis of rainfall-runoff processes on a hillslope. *Water Resources Research* **16**: 391–408.

Gan TY, Burges SJ. 1990. An assessment of a conceptual rainfall-runoff model's ability to represent the dynamics of small hypothetical catchments: 2. Hydrologic responses for normal and extreme rainfall. *Water Resources Research* **26**: 1605–1619.

Heppner CS, Loague K, VanderKwaak JE. 2007. Long-term InHM simulations of hydrologic response and sediment transport for the R-5 catchment. *Earth Surface Processes and Landforms* **32**: 1273–1292.

Heppner CS, Loague K. 2008. A dam problem: Simulated upstream impacts for a Searsville-like watershed. *Ecohydrology* **1**: 408–424.

Kaheil YH, Gill MK, McKee M, Bastidas LA, Rosero E. 2008. Downscaling and Assimilation of Surface Soil Moisture Using Ground Truth Measurements. *IEEE Transactions on Geoscience and Remote Sensing* **46**: 1375–1384.

Loague KM, Freeze RA. 1985. A comparison of rainfall-runoff modeling techniques on small upland catchments. *Water Resources Research* **21**: 229–248.

Loague K, Abrams RH. 2001. Stochastic-conceptual analysis of near-surface hydrologic response. *Hydrological Processes* **15**: 2715–2728.

Loague K, Heppner CS, Abrams RH, VanderKwaak JE, Carr AE, Ebel BA. 2005. Further testing of the Integrated Hydrology Model (InHM): Event-based simulations for a small rangeland catchment located near Chickasha, Oklahoma. *Hydrological Processes* **19**: 1373–1398.

Loague K, VanderKwaak JE. 2004. Physics-based hydrologic response simulation: Platinum bridge, 1958 Edsel, or useful tool. *Hydrological Processes* **18**: 2949–2956, DOI:10·1002/hyp.5737.

Milly PCD, Eagleson PS. 1988. Effect of storm scale on surface runoff volume. *Water Resources Research* **24**: 620–624.

Mirus BB, Ebel BA, Heppner CS, Loague K. 2011. Assessing the detail needed to capture rainfall-runoff dynamics with physics-based hydrologic-response simulation. *Water Resources Research* **47**: DOI:1029/2010WR009906.

Mirus BB, Ebel BA, Loague K, Wemple BC. 2007. Simulated effect of a Forest Road on Near-Surface Hydrologic Response: Redux. *Earth Surface Processes and Landforms* **32**: 126–142, DOI:10·1002/esp.1387.

Mirus BB, Loague K, VanderKwaak JE, Kampf SK, Burges SJ. 2009. A Hypothetical Reality of Tarrawarra-Like Hydrologic Response. *Hydrological Processes* **23**: 1093–1103.

Mogheir Y, de Lima JLMP, Singh VP. 2004. Characterizing the spatial variability of groundwater quality using the entropy theory: I. Synthetic data. *Hydrological Processes* **18**: 2165–2179. DOI:10·1002/hyp.1465.

Shuttleworth WJ. 1993. Evaporation. In *Handbook of Hydrology*. Maidment DR (ed.). McGraw-Hill, Inc: New York.

Slaughter CW, Marks D, Flerchinger GN, Van Vactor SS, Burgess M. 2001. Thirty-five years of research data collection at the Reynolds Creek Experimental Watershed, Idaho, United States. *Water Resources Research* **37**: 2819–2823.

Smith RE, Hebbert RHB. 1983. Mathematical simulation of interdependent surface and subsurface hydrologic processes. *Water Resources Research* **19**: 987–1001.

Troch PA, Mancini M, Paniconi C, Wood EF. 1993. Evaluation of a Distributed Catchment Scale Water Balance Model. *Water Resources Research* **29**: 1805–1817.

VanderKwaak JE. 1999. *Numerical simulation of flow and chemical transport in integrated surface-subsurface hydrologic systems*. PhD Dissertation, University of Waterloo, Waterloo.

VanderKwaak JE, Loague K. 2001. Hydrologic-response simulations for the R-5 catchment with a comprehensive physics-based model. *Water Resources Research* **37**: 999–1013.

Vereecken H, Kasteel R, Vanderborght J, Harter T. 2007. Upscaling hydraulic properties and soil water flow processes in heterogeneous soils: A review *Vadose Zone Journal* **6**: 1–28, DOI:10·2136/vzj2006·0055.

Western AW, Grayson RB. 1998. The Tarrawarra dataset: Soil moisture patterns, soil characteristics and hydrological flux measurements. *Water Resources Research* **34**: 2765–2768.

Wood EF. 1976. An analysis of the effects of parameter uncertainty in deterministic hydrologic models. *Water Resources Research* **12**: 925–932.

Zhou Y, McLaughlin D, Entekhabi D, Ng GC. 2008. An ensemble multiscale filter for large nonlinear data assimilation problems. *Monthly Weather Review* **136**: 678–698.