Advances in Water Resources 82 (2015) 106–123

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Quantitative assessment of groundwater controls across major US river basins using a multi-model regression algorithm

Laura E. Condon [a,b,d,*], Amanda S. Hering [c], Reed M. Maxwell [a,b,d]

[a] Department of Geology and Geological Engineering, Colorado School of Mines, United States
[b] Integrated GroundWater Modeling Center, United States
[c] Department of Applied Mathematics and Statistics, Colorado School of Mines, United States
[d] Climate Change Water and Society (CCWAS), Integrative Graduate Education and Research Traineeship (IGERT), United States

ARTICLE INFO

Article history:
Received 24 December 2014
Received in revised form 14 April 2015
Accepted 15 April 2015
Available online 23 April 2015

Keywords:
Groundwater
Integrated modeling
Multi-model regression

ABSTRACT

Spatial patterns in the physical controls of groundwater depth and flux are assessed quantitatively using results from a first of its kind, integrated groundwater surface water simulation over the majority of the contiguous US. We apply a novel, k-regression algorithm to the simulated system to simultaneously identify spatial subsets of grid cells with similar relationships between explanatory variables and groundwater metrics while quantifying behavior using multiple linear regression. The combination of this statistical approach with the results of a large-scale, high-resolution groundwater simulation allows us to evaluate the ability to represent complex groundwater behavior with simple linear models across an unprecedented range of climates and physical settings. In almost all of the eight major basins considered, we identify at least some areas where the coefficient of determination for the linear regression model is larger than 0.7, and in many cases this is achieved for more than 50% of the total basin area. In general, we show that water table depth is most strongly related to location within a basin and slope, while conductivity and recharge are more important predictors for groundwater flux metrics. Results also illustrate spatial variability in these relationships; further demonstrating the historic difficulty in developing spatially contiguous classifications of groundwater behavior. This work highlights the potential to combine new statistical techniques with integrated hydrologic models to help improve our understanding of complex heterogeneous systems.

1. Introduction

The study of water table configuration and groundwater movement is an enduring research topic. Early work by Hubbert [1], King [2] and Toth [3] established a conceptual formulation of groundwater as a subdued replica of topography. Within this model, groundwater recharges at high elevations and converges to low-lying discharge points, creating a pattern of deep water tables along topographic divides that become progressively shallower around surface water bodies. Toth and others further demonstrated that systems of recharge and discharge can be nested across multiple spatial scales and that topographic slope is an important control on groundwater depth and surface water exchanges [3–5]. Taking advantage of these relationships, Beven and Kirkby [6] developed the widely used TOPMODEL, which simulates groundwater configuration using the topographic index calculated from drainage area and slope.

Notwithstanding the undisputed connection between groundwater and topography, many studies have shown that geology and climate are similarly important groundwater controls [7–12]. At the most basic level, geologic properties, like hydraulic conductivity, moderate the ease with which water moves through the subsurface and can control preferential flow paths. Similarly, climate dictates the potential recharge rate (i.e. precipitation minus evaporation, PME) and in turn, the total flux through the groundwater and surface water systems. Of course, within real world domains, these connections are not so simple. Topography, elevation and climate are highly heterogeneous variables with demonstrated scaling behavior. Nonlinear interactions between these three drivers result in complex groundwater systems which are both difficult to observe and difficult to characterize.

While there are numerous studies of local groundwater behavior within heterogeneous domains, McDonnell and Woods [13] note that so far this approach has not produced results that are

* Corresponding author at: Department of Geology and Geological Engineering, Colorado School of Mines, United States.
E-mail address: lcondon@mymail.mines.edu (L.E. Condon).

http://dx.doi.org/10.1016/j.advwatres.2015.04.008
0309-1708/

easily transferable to larger scales or other local areas due to the high degree of complexity within hydrologic systems and 'tremendous variability in space, time and process'. They argue that hydrologists still lack guidance on dominant processes and mechanisms within a basin and that this gap could be addressed with a unified classification system. Hydrologic response units for surface water systems are not a new concept (e.g. [14]); however, developing classification systems for the integrated groundwater surface water system has proven more challenging. Winter [15] introduced the concept of fundamental hydrologic landscape units (FHLU) that define contiguous spatial units based on climate, topographic and geologic variables that take into consideration both groundwater and surface water systems. Wolock et al. [12] applied this concept to the contiguous US and defined 20 hydrologic regions from the variables identified by Winter [15] using Principal Component Analysis (PCA) combined with a minimum variance criterion and the nearest neighbor chain algorithm. While the FHLU regions were shown to be good predictors of some hydrologic behavior, groundwater was not explicitly considered.

Indeed, as a result of data scarcity and inconsistencies, very few studies have explicitly sought to characterize groundwater behavior at continental scales. Gleeson et al. [16,17] used Water Table Ratios [9] combined with gridded conductivity, potential recharge and elevation fields to assess the degree to which the groundwater table is controlled by topography across the contiguous US; although here too the water table was not explicitly simulated. Schaller and Fan [5] evaluated basin scale groundwater surface water exchanges over the same extent using observational data. While their approach does not rely on inferred groundwater behavior, they are still limited by data availability. Though these studies are quite different in focus and approach, both reported spatial differences in the dominant processes, and ultimately the controls, of groundwater behavior.

Here we evaluate groundwater behavior by combining an explicit simulation of the groundwater and surface water systems across the majority of the contiguous US with a novel statistical algorithm. Numerical modeling is an established tool in hydrology; however, integrated models that solve the surface and subsurface simultaneously are a relatively modern development [18–20] that have primarily been applied to regional and local studies (see Table 1 in Maxwell et al. [21]). Computational advances along with new data sources have only recently facilitated high-resolution simulation at the continental scale [22]. Although the integrated approach is still computationally intensive; the ability to generate *simulated truths* over such a large area at a high level of physical complexity and spatial resolution allows us to explicitly evaluate groundwater behavior across a broad range of settings and to potentially develop robust relationships that could be used to improve more simplified models.

Although there are limitations to working from simulated results, this approach provides the unique advantage of high resolution, consistent, gridded model outputs. This facilitates direct statistical analysis of spatial heterogeneity in groundwater controls. In statistics, there are a number of ways to empirically model systems where the causal relationships between response and explanatory variables are not the same for all observations. Finite mixture and Markov switching models are established tools in the fields of economics, marketing and machine learning [23]. Unlike cluster analysis, which groups observations solely with the explanatory variables (i.e. with no consideration of the link to dependent variables), mixed regression approaches take into account both the response and the explanatory variables. In hydrology, this type of statistical analysis has traditionally been limited to studies of subsurface characterization and contaminant transport (e.g. [24,25]). However, similar techniques have also been applied to remote sensing data of soil moisture variability [26] and vegetation types [27]. Still, to our knowledge, statistical analysis of regional scale controls on groundwater behavior has not previously been conducted.

In this study, we leverage advances in both integrated modeling and statistical methods to quantitatively identify spatial patterns in the relative importance of physical groundwater controls and evaluate the potential to characterize complex groundwater systems with simple linear models. Analysis is based on a first of its kind, high resolution (1 km), fully-integrated groundwater surface water model of the majority of the contiguous US (CONUS) spanning 6.3 M km^2 [22]. We apply a novel *k*-regression algorithm, which is related to finite mixture regression approaches, to hydrologic model outputs to rigorously characterize (1) the relative importance of topography, climate and recharge as groundwater predictors and (2) spatial variability in these relationships. We take advantage of the explicit groundwater simulation combined with spatially gridded model inputs to evaluate behavior across a wide range of climates, physical settings and scales. This approach

Table 1
Summary of the explanatory variables including, their abbreviation, a short description, spatial aggregation level and variable type.

#	Abbreviation	Description	Spatial aggregation level	Variable type
1	Elev	Elevation (m)	Local (L)	Location
2	Slope	Topographic slope (–)		Slope
3	lnK	Natural log hydraulic conductivity (m/h)		Conductivity
4	PME	Recharge: precipitation – evaporation (m/h)		Recharge
5	TI	Topographic index (–)	TI	Location
6	S_Elev%	Subbasin elevation (% of regional elevation range)	Subbasin (S)	Location
7	S_Relief	Subbasin topographic relief (m)		Slope
8	S_Relief%	Subbasin relief fraction (% of regional elevation range)		Slope
9	S_Slope	Average subbasin topographic slope (–)		Slope
10	S_lnK	Subbasin median natural log hydraulic conductivity (m/h)		Conductivity
11	S_varK	Subbasin variance of hydraulic conductivity (m/h)		Conductivity
12	S_PME	Subbasin recharge: precipitation–evaporation (m/y)		Recharge
13	Order	Stream order (–)	Upstream area (U)	Location
14	U_Elev%	Upstream elevation (% of regional elevation range)		Location
15	U_Area	Upstream drainage area (km^2)		Location
16	U_Area%	Upstream drainage area fraction (% of regional drainage area)		Location
17	U_Relief	Upstream topographic relief (m)		Slope
18	U_Relief%	Upstream relief fraction (% of regional elevation range)		Slope
19	U_Slope	Average upstream topographic slope (–)		Slope
20	U_lnK	Upstream median natural log hydraulic conductivity (m/h)		Conductivity
21	U_varK	Upstream variance of hydraulic conductivity (m/h)		Conductivity
22	U_PME	Upstream recharge: precipitation – evaporation (m/yr)		Recharge