Journal of Hydrology 555 (2017) 938–955

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

The combined use of dynamic factor analysis and wavelet analysis to evaluate latent factors controlling complex groundwater level fluctuations in a riverside alluvial aquifer

Yun-Yeong Oh [a], Seong-Taek Yun [a,*], Soonyoung Yu [a], Se-Yeong Hamm [b]

[a] Korea-CO_2 Storage Environmental Management (K-COSEM) Research Center, Department of Earth and Environmental Sciences, Korea University, Seoul 02843, Republic of Korea
[b] Division of Earth Environmental System, Pusan National University, Busan 46241, Republic of Korea

ARTICLE INFO

Article history:
Received 6 August 2017
Received in revised form 22 October 2017
Accepted 27 October 2017

This manuscript was handled by P. Kitanidis, Editor-in-Chief, with the assistance of Simon A. Mathias, Associate Editor

Keywords:
Groundwater level fluctuation
Wavelet based clustering
Dynamic factor analysis
Wavelet analysis
Multiresolution cross-correlation
Effective dynamic efficiency

ABSTRACT

To identify and quantitatively evaluate complex latent factors controlling groundwater level (GWL) fluctuations in a riverside alluvial aquifer influenced by barrage construction, we developed the combined use of dynamic factor analysis (DFA) and wavelet analysis (WA). Time series data of GWL, river water level and precipitation were collected for 3 years (July 2012 to June 2015) from an alluvial aquifer underneath an agricultural area of the Nakdong river basin, South Korea. Based on the wavelet coefficients of the final approximation, the GWL data was clustered into three groups (WCG1 to WCG3). Two dynamic factors (DFs) were then extracted using DFA for each group; thus, six major factors were extracted. Next, the time–frequency variability of the extracted DFs was examined using multiresolution cross-correlation analysis (MRCCA) with the following steps: 1) major driving forces and their scales in GWL fluctuations were identified by comparing maximum correlation coefficients (r_{max}) between DFs and the GWL time series and 2) the results were supplemented using the wavelet transformed coherence (WTC) analysis between DFs and the hydrological time series. Finally, relative contributions of six major DFs to the GWL fluctuations could be quantitatively assessed by calculating the effective dynamic efficiency (D_{ef}). The characteristics and relevant process of the identified six DFs are: 1) WCG1DF$_{4,1}$ as an indicative of seasonal agricultural pumping (scales = 64–128 days; r_{max} = 0.68–0.89; $D_{ef} \leq 23.1\%$); 2) WCG1DF$_{4,4}$ representing the cycle of regional groundwater recharge (scales = 64–128 days; r_{max} = 0.98–1.00; $D_{ef} \leq 11.1\%$); 3) WCG2DF$_{4,1}$ indicating the complex interaction between the episodes of precipitation and direct runoff (scales = 2–8 days; r_{max} = 0.82–0.91; $D_{ef} \leq 35.3\%$) and seasonal GW-RW interaction (scales = 64–128 days; r_{max} = 0.76–0.91; $D_{ef} \leq 14.2\%$); 4) WCG2DF$_{4,4}$ reflecting the complex effects of seasonal pervasive pumping and the local recharge cycle (scales = 64–128 days; r_{max} = 0.86–0.94; $D_{ef} \leq 16.4\%$); 5) WCG3DF$_{4,2}$ as the result of temporal pumping (scales = 2–8 days; r_{max} = 0.98–0.99; $D_{ef} \leq 7.7\%$); and 6) WCG3DF$_{4,4}$ indicating the local recharge cycle (scales = 64–128 days; r_{max} = 0.76–0.91; $D_{ef} \leq 34.2$ %). This study shows that major driving forces controlling GWL time series data in a complex hydrological setting can be identified and quantitatively evaluated by the combined use of DFA and WA and applying MRCCA and WTC.

1. Introduction

Hydraulic head in groundwater is one of the most important metrics in hydrogeology and is essentially measured as groundwater level (GWL) (Post and von Asmuth, 2013; Todd and Mays, 2005). There are multiple uses of GWL data, for example, to establish GW flow patterns (Prudhomme et al., 2013), to determine the response of an aquifer to stresses such as pumping or recharge (Bredehoeft, 2002; Chae et al., 2010; von Asmuth et al., 2008), to characterize the interaction between groundwater and surface water (Huntington and Niswonger, 2012; Menció et al., 2014; Rosenberry and LaBaugh, 2008), to determine aquifer properties by examining time variant GWL changes (Ha et al., 2007; Hall and Moench, 1972; Lee et al., 2017; Moench and Barlow, 2000; Oh et al., 2016a; Singh, 2004), and to calibrate groundwater flow models (Foglia et al., 2007; Hansen et al., 2013).

It is now possible to collect GWL data automatically and frequently in observation wells with digital equipment. The frequent and continuous GWL measurements are used to interpret various

* Corresponding author.
E-mail address: styun@korea.ac.kr (S.-T. Yun).

https://doi.org/10.1016/j.jhydrol.2017.10.070
0022-1694/© 2017 Elsevier B.V. All rights reserved.

Nomenclature

a	contraction wavelet coefficient
b	translation coefficient
i	order of observation station
j	order of latent factor
J	maximum wavelet decomposition level
k	length of time series
M	optimal number of dynamic factors
N	number of sampled time series
n	element of dynamic factor in M
p	wavelet resolution level
$\mathbf{Q}$	error covariance matrices of $\varepsilon_i(t)$
q	time domain position
r_{xy}	cross-correlation coefficient
r_{max}	maximum r_{xy}
$\mathbf{R}$	error covariance matrices of $\eta_j(t)$
S	original signal (time series)
s	number of latent factors
t	time [day]
x	input time series
y	output time series
$\bar{x}$	average of x
$\bar{y}$	average of y

Greek symbols

α	DF loading
$\varepsilon_i(t)$	*white noises of $Y_i(t)$*
$\eta_j(t)$	white noises of $F_j(t)$
$\theta_{J,q}$	approximation Wcf of J and q
ϕ	scale function
$\lambda_{p,q}$	detail Wcf of p and q
$\mu_i(t)$	level parameter
σ_x	standard deviations of x
σ_y	standard deviations of y
τ	lag time [day]
ψ	mother wavelet function

Abbreviations

AJ	Approximation component of level J
CCA	cross-correlation analysis
CCF	cross-correlation function
CHB	Changnyeong Haman River Barrage gauge station
CHRB	Changnyeong Haman River Barrage
C_{xy}	cross-correlogram
CWT	continuous wavelet transform
Db5	Daubechies-5 mother wavelet
D_{ef}	dynamic efficiency, [%]
Dp	Detail component of level p
DF(s)	dynamic factor(s)
DF loading	dynamic factor loading
DFA	dynamic factor analysis
DN	Deongnam RWL gauge station
DWT	discrete wavelet transform
$\mathrm{DF}_{\mathrm{M,n}}$	n^{th} DF in M of DF model
DG	wavelet clustered group for DFs
E_{DF}	ratio of sum of E_w at major scales
E_w	wavelet energy
f_l	length of a filter function
$f(t)$	target time series
$F_j(t)$	j^{th} common latent factor
GWL	groundwater level
GY	Georyonggang RWL gauge station
HAM-	GWL observation wells
JD	Jindong RWL gauge station
MFTS	multifactor time series
MRA	multiresolution analysis
MRCCA	multiresolution CCA
R_GY	Georyonggang rainfall gauge station
R_JD	Jindong rainfall gauge station
R_YS	Yeongsan rainfall gauge station
RWL	river water level
WA	wavelet analysis
Wcf(s)	wavelet coefficient(s)
WCG	wavelet clustered group for GWLs
$Y_i(t)$	i^{th} observed time series

spatiotemporal characteristics, such as the hydrometeorological cycle, infiltration and recharge, groundwater – surface water interactions, groundwater use, aquifer geometry, and hydraulic anisotropy and heterogeneity. GWL data as a multi-factor time series (MFTS) embeds a multitude of processes in the hydrologic cycle (Post and von Asmuth, 2013). It is usually assumed that there are common driving forces behind observed MFTS data whereby individual observations can be explained by a few latent factors (Anderson, 1963; Márkus et al., 1999). Several statistical tools have been suggested to determine latent factors.

Factor analysis (FA) is a conventional statistical tool to determine latent factors. However, the application of FA to MFTS data often produces unreliable or misleading results (e.g., spurious regression), particularly when delayed interdependence occurs among observed variables (Jolliffe, 2002). Moreover, the majority of hydrological time series, including precipitation, river water level (RWL) and GWL, have autoregression and a long memory effect (Larocque et al., 1998; Schuurmans et al., 2007), mostly due to continuous and cyclic physical processes with a lagged response (Whitcher et al., 2002). Thus, there has been a need for developing new techniques taking into account the dynamic structure of observations, such as non-stationarity, auto-regression and heteroscedasticity (Fathian et al., 2016; Ritter and Muñoz-Carpena, 2006; Westra et al., 2014).

Dynamic factor analysis (DFA) has been applied as an alternative method for MFTS data, revealing its dynamic structure (Harvey, 1990; Muñoz-Carpena et al., 2005). DFA has been used to describe the variation among variables using a few underlying latent variables, denoted as dynamic factors (DFs), reflecting their dynamic characteristics (Berendrecht and van Geer, 2016; Zuur et al., 2003). The major advantages of DFA are: 1) the reduction of the dimensionality of large datasets, improving the efficiency of the analysis as FA and 2) the applicability to interdependent and non-stationary time series data (Kuo and Lin, 2010; Shojaei et al., 2016). In hydrogeology, DFA has been used to recognize the trends of GWL, including recharge and extraction (Márkus et al., 1999). For such cases, DFA was combined with a transfer function noise model to include explanatory variables such as precipitation and drainage (Berendrecht et al., 2004) or couped with a simple regression model to identify trends in GWL and surface water levels (Muñoz-Carpena et al., 2005). For example, Kaplan et al. (2010) discriminated the factors explaining GWL fluctuations in coastal floodplain wetlands, including regional groundwater circulation, surface water elevation, and net local recharge. Kovács