Journal of Hydrology 536 (2016) 50–60

Contents lists available at ScienceDirect

Journal of Hydrology

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

A dynamic factor modeling framework for analyzing multiple groundwater head series simultaneously

W.L. Berendrecht [a,*], F.C. van Geer [b]

[a] Berendrecht Consultancy, Stakenbergerhout 107, 3845 JE Harderwijk, The Netherlands
[b] TNO Geological Survey of the Netherlands, P.O. Box 80015, 3508 TA Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 30 September 2015
Received in revised form 5 February 2016
Accepted 15 February 2016
Available online 24 February 2016
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Wolfgang Nowak, Associate Editor

Keywords:
Dynamic factor model
Time series modeling
Multivariate analysis
Groundwater head
Gap filling
Outlier detection

SUMMARY

In this paper we present an approach in which we combine a dynamic factor model (DFM) and predefined response functions to analyze a set of groundwater head series simultaneously. Each groundwater head series is decomposed into: (a) one or more deterministic components as a response to known driving forces, (b) one or more common dynamic factors, representing spatial patterns not related to any of the input series and (c) one specific dynamic factor for each groundwater head series, describing unique variation for that series. The approach reduces the degrees of freedom for each response function, enables the application to irregular observed data, and exploits the correlation between residual series of a set of groundwater head series. The common dynamic factors may be interpreted as spatial patterns due to e.g. limitations in the model specification or concept, spatially correlated errors in input variables, or driving forces which have not been included in the model. In the latter case the model can be applied in the context of an alarming system, e.g. to monitor regional trends. The specific dynamic factor depicts the variation of a particular groundwater head series that cannot be related to any other time series of the set nor to any input series. Therefore the specific dynamic factor is suitable for analyzing local variations and detecting incidental measurement errors, for example in a quality control procedure. The DFM framework is illustrated with a set of 8 groundwater head series and applied for filling gaps in time series, reconstructing high-frequency data, and detecting outliers.

1. Introduction

As many other hydrological variables, groundwater head is being monitored since many decades, resulting in an enormous number of time series. For example over 35,000 groundwater head time series are available in the national repository DINO of the Geological Survey of the Netherlands (https://www.dinoloket.nl). During the last 30 years, many groundwater head time series have been analyzed using time series models. Objectives of these analyses were to assess temporal trends and to quantify the response of groundwater heads to driving forces like precipitation, evapotranspiration and groundwater abstraction (Kim et al., 2005; Gehrels et al., 1994).

In the early years, the most applied time series models for groundwater head were transfer function noise models as described by Box and Jenkins (1994). These models are based on a temporal discretization with a constant observation frequency. In the past decades, more and more automatic data loggers are used for the observation of groundwater heads, enabling high frequent sampling (e.g. 1 day). In many cases, a Box–Jenkins model for high-frequency groundwater head series includes a large number of degrees of freedom, because the modeling time step is small with respect to the response time. This high dimensionality complicates estimation of the model parameters. Therefore, Von Asmuth et al. (2002) suggested a more robust and parsimonious approach by using predefined response functions, which reduces the degrees of freedom tremendously. Such predefined functions can be applied in continuous time as demonstrated by Von Asmuth et al. (2002), or in discrete time using a Kalman filter framework to accommodate changes in observation frequency as well as irregularly spaced observations (Bierkens et al., 1999; Berendrecht et al., 2003).

Most applications of time series models to groundwater head series have been single output models, decomposing one single groundwater head series into components attributed to known input series, and a residual noise component (e.g. Ahn, 2000; Tankersley et al., 1993). Often groundwater head series obtained from the same hydrological system show similar patterns.

* Corresponding author.
E-mail addresses: info@berendrechtconsultancy.nl (W.L. Berendrecht), frans.vangeer@tno.nl (F.C. van Geer).

http://dx.doi.org/10.1016/j.jhydrol.2016.02.028
0022-1694/

Obviously, the components of different groundwater head series driven by the same input series (e.g. precipitation) show similarities in their response. However, also the residual noise components of different groundwater head time series may be correlated. In multiple output time series models a set of output time series is being analyzed simultaneously, taking into account mutual correlation. Multiple output time series models are well known in time series analysis (Lewis and Reinsel, 1985; Shea, 1987; Lütkepohl, 2007). An application of the model introduced by Shea to groundwater head series is given by Van Geer and Zuur (1997). Camacho et al. (1987) introduced the 'contemporaneous' autoregressive moving-average (ARMA) model, in which an output time series is modeled as a linear combination of historical observations of the series itself as well as time series in the surrounding area. Deutsch and Pfeifer (1981) and Stoffer (1986) describe the 'space–time' ARMA model, where physical knowledge is used to define at forehand the spatial coherence in order to reduce the number of degrees of freedom.

Another class of multivariate time series models is the dynamic factor model originally proposed by Geweke (1977). Dynamic factor modeling (DFM) is a multivariate time-series analysis technique used to describe the variation among many variables in terms of a few underlying but unobserved variables called common dynamic factors. One of the attractive properties of DFM is that it reduces the dimensionality of large systems of multivariate time series, which makes them very efficient compared to the methods described above. In addition, it allows to identify underlying common patterns or latent effects in time series.

Dynamic factor models have been used in econometric and psychological related fields (Molenaar, 1985; Molenaar et al., 1992; Harvey, 1989) and environmental sciences (Zuur et al., 2003; Zuur and Pierce, 2004). Márkus et al. (1999) applied dynamic factor analysis in hydrology to identify common patterns of groundwater level in a karstic area of Hungary. Although they identified two common trends as recharge and extraction, no explanatory variables were included in the model. Berendrecht (2004) combined DFM with a transfer function noise (TFN) model to include explanatory variables in the analysis of multiple groundwater time series. Ritter and Muñoz-Carpena (2006) combined DFM with a simple regression model to identify common trends in groundwater and surface water levels.

In this paper we continue on the research of Berendrecht (2004) and combine the DFM framework with predefined response functions (Von Asmuth et al., 2002) for modeling explanatory variables. The method presented here is efficient and robust in the sense that it reduces the degrees of freedom for each response function and that it allows for application to irregularly observed data. The paper demonstrates how DFM exploits the residual correlation of time series to reveal common dynamic factors and specific dynamic factors. Whereas most applications of dynamic factor modeling focus on identifying common trends (e.g. Zuur et al., 2003; Ritter and Muñoz-Carpena, 2006), this paper shows that DFM is a powerful method for gap filling, time series extension and reconstruction of high-frequency data. Moreover we show that specific dynamic factors can be used to reveal outliers.

The major contributions of this paper are therefore:

- Defining and explaining a dynamic modeling framework for analyzing irregularly observed hydrological time series simultaneously. This includes methods for selecting the number of common dynamic factors, factor rotation to allow for meaningful interpretation of factors, and efficient filtering, smoothing, and parameter estimation procedures.
- Combining dynamic factor modeling with predefined response functions for explanatory variables, enabling a parsimonious and robust modeling of multiple time series.
- Decomposition of series into common and specific factors, which provides information on coherence and common dynamic patterns in the observed hydrological system. This paper demonstrates how these factors can be used for filling gaps in time series, reconstructing high-frequency data within the context of evaluating monitoring networks, and detecting location-specific outliers.

Section 2 introduces the dynamic factor modeling framework for analyzing irregularly observed time series. It presents the discrete-time state space formulation of the DFM including explanatory variables by means of predefined response functions. The section comprehensively describes how the dynamic factor model decomposes residual series into common and specific factors using maximum likelihood estimation combined with an efficient multivariate Kalman filtering technique. It addresses the issue of selecting the number of common factors and describes the method of factor rotation as a means to allow for a plausible interpretation of factors. In Section 3 we illustrate the method by applying the model to a set of 8 groundwater head series, mainly driven by precipitation and evapotranspiration. We use the model to fill gaps in time series and to reconstruct high-frequency data. Results show that for time instants with missing data, dynamic factor modeling produces smaller prediction errors than single output models, as the DFM includes information from observations taken from surrounding locations. Section 3 also shows that analysis of specific dynamic factors may help to detect outliers. As these outliers are location-specific and therefore not observed in surrounding time series, these outliers may be indicative for incidental measurement errors. Section 4 gives a discussion of the DFM framework and results presented in this paper.

2. Theory and background

2.1. The univariate TFN model

For linear groundwater systems the following multi-input discrete transfer function noise (TFN) model can be used to model groundwater head fluctuations (Tankersley et al., 1993; Knotters and Van Walsum, 1997; Von Asmuth et al., 2008):

$$y_t = d_t + n_t + \mu + \varepsilon_t \tag{1}$$

$$d_t = \sum_{j=1}^{p} \Theta_j(\mathrm{B})u_{j,t} = \sum_{j=1}^{p}\left[\sum_{i=0}^{\infty}\Theta_{j,i}u_{j,t-i}\right] \tag{2}$$

$$n_t = \Phi(\mathrm{B})\eta_t = \sum_{i=0}^{\infty}\Phi_i\eta_{t-i} \tag{3}$$

with y_t the observed groundwater head at time t [L], d_t the deterministic component at time t relating one or more input series to the observed time series using a transfer function model [L], n_t the stochastic or residual component at time t [L], μ a constant local drainage level relative to some reference level [L], ε_t measurement noise at time t [L] which is assumed to be a zero mean white noise process with variance σ^2_ε, $u_{j,t}$ the jth input variable at time step t [L], $\Theta_j(\mathrm{B})$ a transfer function for input variable j [–], B a backward shift operator [–] defined as $\mathrm{B}^i u_{j,t} = u_{j,t-i}$, $\Theta_{j,i}$ is the ith coefficient of the transfer function for input variable j [–], $\Phi(\mathrm{B})$ a noise transfer function [–], Φ_i is the ith coefficient of the noise transfer function [–], and η_t a zero mean white noise process [L] with variance σ^2_η.

Box and Jenkins (1994) define the transfer function $\Theta(\mathrm{B})$ in Eq. (2) as a fraction, where the numerator is a moving average (MA) function $\omega(\mathrm{B})$, and the denominator an autoregressive (AR) function $\delta(\mathrm{B})$, so that