Signal Processing 91 (2011) 553–561

Contents lists available at ScienceDirect

Signal Processing

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

FARIMA with stable innovations model of Great Salt Lake elevation time series

Hu Sheng [a,b], YangQuan Chen [b,*]

[a] Department of Electronic Engineering, Dalian University of Technology, Dalian 116024, China
[b] Electrical and Computer Engineering Department, Utah State University, Logan, UT 84322, USA

ARTICLE INFO

Article history:
Received 11 November 2009
Received in revised form
22 January 2010
Accepted 27 January 2010
Available online 16 February 2010

Keywords:
Hurst parameter
Long-range dependent
Stable distribution
FARIMA
FIGARCH
FARIMA with stable innovations

ABSTRACT

Great Salt Lake (GSL) is the largest salt lake in the western hemisphere, the fourth-largest terminal lake in the world. The elevation of GSL has critical effect on the people who live nearby and their properties. It is crucial to build an exact model of GSL elevation time series in order to predict the GSL elevation precisely. Although some models, such as ARIMA or FARIMA (fractional auto-regressive integrated moving average), GARCH (generalized auto-regressive conditional heteroskedasticity) and FIGARCH (fractional integral generalized auto-regressive conditional heteroskedasticity) have been proposed to characterize the variation of GSL elevation, which have been unsatisfactory. Therefore, it became a key point to build a more appropriate model of GSL elevation time series. In this paper a new model based on FARIMA with stable innovations is applied to analyze the data and predict the future elevation levels. From the analysis we can see that the new model can characterize GSL elevation time series more accurately. The new model will be beneficial to predict GSL elevation more precisely.

1. Introduction

Great Salt Lake (GSL), located in the northern part of the US State of Utah, is the largest salt lake in the western hemisphere, the fourth-largest terminal lake in the world. In an average year the lake covers an area of around 1700 square miles (4400 km^2), but the lake's size fluctuates substantially due to its elevation (shallowness). GSL is located on a shallow playa, so small changes in the water-surface elevation result in large changes in the surface area of the lake. For instance, in 1963 it reached its lowest recorded level at 950 square miles (2460 km^2), but in 1987 the surface area was at the historic high of 3300 square miles (8547 km^2) [1]. The variations of the GSL elevation have an enormous impact on the people who live nearby. The rise in 1987 had caused 285 million US dollars worth of damage to lakeside industries, roads, railroads, wildfowl management areas, recreational facilities and farming that had been established on the exposed lake bed [2].

GSL is divided into a north and a south part by a rock-fill causeway. Because of the importance of the GSL elevation, the United States Geological Survey (USGS) has been collecting water surface elevation data from the south part of GSL since 1875 and continuously since October 1902. The north part of the lake has been monitored since 1960 [3]. The USGS operates gages that collect water surface elevation data on the south part of the lake at the Boat Harbor gage, and on the north part of the lake at the Saline gage [4]. We found that the distribution of the data from north part is evidently heavy tailed, so the north part water surface elevation data of the lake was analyzed in the paper. Several studies have been performed to build the precise model of the GSL elevation time series and a variety of techniques have

* Corresponding author.
E-mail addresses: yangQuan.chen@usu.edu, yqchen@ece.usu.edu (Y. Chen).

0165-1684/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2010.01.023

been used to estimate historical GSL elevation time series, including geological and archeological methods [5–7]. Despite these preliminary efforts, all the conventional methods and models were found to be insufficient to character the lake levels and predict its future. One reason for such inadequacy might be the existence of long-range dependence in the GSL elevation time series [8]. Another reason might be the non-convergence of the second order moment of the GSL elevation time series. So the fractional order signal processing (FOSP) techniques are probably the better techniques to model and predict it [9]. Fractional order signal processing is, in recent years, becoming a very active research area. FOSP provides many powerful techniques to analyze fractional systems which have both short and long-term memories or time series with heavy tailed distribution [10]. FOSP is based on the knowledge of fractional order calculus (FOC). FOC is a generalization of the conventional differential and integral operators [11]. It is the basis of the fractional systems described by fractional order differential equations. The simplest fractional order dynamic systems include the fractional order integrators and fractional order differentiators. The fractional autoregressive integrated moving average (FARIMA) with stable innovations model is a typical fractional order system. It behooves us to develop models that combine both features: infinite variance and long-range dependence [12]. FARIMA with stable innovations model is based on linear fractional stable noise (LFSN) which is stationary, self-similar and heavy tailed processes. "Model accepts physical interpretation, since it explains how the observed data appear as a superposition of independent effects" [13]. The traditional models, such as autoregressive (AR), moving average (MA), autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) processes, can only capture the short-range dependence [14]. FARIMA and FIGARCH models give a good fit for long-range dependent(LRD) time series, but it cannot character the time series with heavy tailed character preciously. Therefore, the FARIMA with stable innovations model is sufficient to describe the GSL elevation time series.

The paper is organized as follows. In Section 2 we define the term self-similarity and the stable distribution. In Section 3, FARIMA, FIGARCH and FARIMA with stable innovations models are presented. Section 4 introduces the LRD and the stable parameter estimators. In Section 5, a FARIMA with stable innovations model of GSL elevation time series was established to analyze the data and predict the future elevation level of GSL. The results of the FARIMA with stable innovations model of GSL elevation time series are presented in Section 6.

2. Preliminaries

2.1. LRD and hurst parameter

LRD processes appear in many contexts. Which is characterized by the Hurst parameter ($0<H<1$). In LRD processes, there is a strong coupling between values at different time [15]. This indicates that the decay of the autocorrelation function is hyperbolic and decays slower than exponential decay, and that the area under the autocorrelation function curve is infinite. LRD analysis are widely applied in may time series, such as financial data, communications networks data and biological data. A stationary process is said to have long-range correlations if its covariance function $C(n)$ (assume that the process has finite second order statistics) decays slowly as $n\rightarrow\infty$. For $0<\alpha<1$,

$$\lim_{n\rightarrow\infty}\frac{C(n)}{n^{-a}}=c, \tag{1}$$

where c is finite, positive constant. That is to say, for large n, $C(n)$ looks like c/n^{α} [16]. The parameter α is related to the Hurst parameter via the equation $\alpha=2-2H$ [17]. Hurst parameter is a simple direct parameter which characterizes the degree of long-range dependence. The Hurst exponent, which was proposed for analysis of long-term storage capacity of reservoirs more than a half century ago [18], is used today to measure the intensity of LRD. Indeed, a lot of time series are often described in the model with long-range dependence characters. $0<H<0.5$ indicates that the time series is a negatively correlated process, or anti-persistent process, and $0.5<H<1$ indicates it is a positively correlated process [17]. It should be stressed that here we consider only processes with finite second moments. If the process has infinite second order statistics then we cannot describe the LRD with covariance function. But the processes with infinite second order statistics really can exhibit LRD character. And there are well-defined self-similar processes with stationary increments and infinite second moments [17]. Typical examples are stable self-similar processes and fractional ARIMA processes with infinite variance innovations [19–21].

Fractional Gaussian noise (FGN) and the class of FARIMA processes are most popular models to simulate self-similar processes. FGN, as an example of exactly self-similar process, is used to model phenomena in many disciplines, e.g. in computer networks signal processing, economics and queuing systems. Time series with short-range and long-range dependent also can be well described by FARIMA processes. The autocorrelation of FARIMA process decays hyperbolically as the time increases, which causes significant difference with relation to traditional, short-range dependent stochastic processes such as Markov, Poisson or ARMA processes.

2.2. Stable distribution

Non-Gaussian signals and noises more frequently tend to produce large-amplitude aberrancy from the average value than Gaussian ones. Non-Gaussian signals and noises are more likely to exhibit sharp spikes or occasional bursts of outlying observations than one would expect from normally distributed signals. Underwater acoustic signals, low-frequency atmospheric noise, and many types of man-made noise have all been found to belong to this class. The stable distributions provide a useful theoretical tool for this type of signals and noise [22].