Environmental Modelling & Software 105 (2018) 244–256

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Real-time forecasting and visualization toolkit for multi-seasonal time series

Jinduan Chen [a], Dominic L. Boccelli [b, *]

[a] Environmental Engineering Program, 737 Engineering Research Center, University of Cincinnati, Cincinnati, OH 45221-0012, USA
[b] Environmental Engineering Program, 742 Engineering Research Center, University of Cincinnati, Cincinnati, OH 45221-0012, USA

A R T I C L E I N F O

Article history:
Received 2 November 2017
Received in revised form
27 February 2018
Accepted 31 March 2018

Keywords:
Seasonality
Autocorrelation
Time series
Forecasting
Visualization

A B S T R A C T

Many environmental data sets are driven by multiple superimposed periods, yet most time series analysis software packages only support single-seasonality. The objective of this research was to develop a software toolkit utilizing multi-seasonal Autoregressive Integrated (msARI) models. A toolkit in MATLAB was developed for msARI-based identification, estimation, forecasting, and visualization. In the toolkit, an adaptive forecasting routine uses a continual event loop for real-time data acquisition and parameter re-estimation. A statistical quality control algorithm monitors model performance and re-estimates parameters when necessary. A set of visualization tools provide animated graphical representations of forecasts, prediction intervals and key performance metrics. The toolkit was applied to three case studies: electricity demands, water demands, and sewer flows. The analysis of the results demonstrated that the explicit modeling of multi-seasonality improved model predictions. Therefore, the msARI software presents a promising tool for modeling and predicting real-time data series.

1. Background and objectives

Many environmental processes are observed in discrete time intervals. These time series can be described and predicted by statistical models when physical models are unavailable or too complex to apply. Traditionally, the techniques of time series analysis have been applied to characterize the dynamics of air pollution (Robeson and Steyn, 1990), rainfall (Burlando et al., 1993), lake water levels (Irvine and Eberhardt, 1992), river flow (Kachroo and Liang, 1992), urban water consumption (Jowitt and Xu, 1992), salmon production (Hare and Francis, 1995), fishery landings (Koutroumanidis et al., 2006), and many others. With the recent advances in peripheral sensor and communication technologies, real-time data services have become more prevalent in environmental applications (e.g., Allen et al., 2011). However, new challenges have been presented to the software tools designed for real-time time series analysis. First, the tools have to adopt high data transfer rates, therefore fast algorithms for parameter estimation and forecasting are required. Second, the tools need to run consistently for a long time without human intervention, which necessitates self-adaptive model structures and/or parameter re-estimation methods (see, e.g., Chen et al., 1995). Third, dynamic visualization tools would benefit the perception of forecasting results and performance metrics. In addition to these real-time software considerations, the fundamental challenge associated with time series modeling remains the identification of the appropriate model structure.

While there are many approaches to time series modeling such as Box-Jenkins time series (Box and Jenkins, 1976; Wei, 2006), Artificial Neural Networks (ANN) (Hill et al., 1996) and Support Vector Machines (SVM) (Müller et al., 1997), our previous studies (Chen and Boccelli, 2014; Chen, 2015) showed that a variant of the classical Box-Jenkins time series named the multi-seasonal Auto-Regressive Integrated (msARI) model is promising for use within a real-time framework. For a selection of case studies, the empirical Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) diagrams showed much stronger characteristics of AutoRegressive (AR) processes than Moving-Average (MA) processes. The msARI models have also been shown to deliver competitive accuracy compared with more complex models (Adhikari and Agrawal, 2013). The relatively simple model structure and reduced number of parameters yield efficient parameter estimation and forecasting algorithms. Also benefiting from the model structure, parameter adaptation algorithms can be implemented in

* Corresponding author.
E-mail addresses: jinduan.uc@gmail.com (J. Chen), dominic.boccelli@uc.edu (D.L. Boccelli).

https://doi.org/10.1016/j.envsoft.2018.03.034
1364-8152/

Software and data availability

msARI toolkit
Name of software msARI toolkit
Developers Jinduan Chen and Dominic L. Boccelli
Contact address 737 Engineering Research Center, University of Cincinnati, Cincinnati, OH 45221–0012, USA
Telephone (513) 923-0706
E-mail jinduan.uc@gmail.com
Year first available 2014
Hardware required PC
Software required Matlab 7.12.0 (R2011a) or up
License GNU General Public License, Version 2
Git repository https://bitbucket.org/Jinduan/tsff
Programming language MATLAB
Program size 1 MB

Data set 1
Name of data Hourly Ontario and Market Demands, 2002–2013
Year first available 2014
Form of repository CSV file
Download webpage http://www.ieso.ca/Pages/Power-Data/Data-Directory.aspx
Size 2.56 MB

Data set 2
Name of data Hourly inflow/outflow data of production and storage facilities of the south-central water distribution network in Hillsborough County, FL, Apr 2012 to Dec 2012
Year first available 2013
Form of repository MATLAB .mat file
Availability https://data.mendeley.com/datasets/4yhprsgjrf/1
Size 52 kb

Data set 3
Name of data Hourly sewer flows monitored at Station S2 in Columbus, OH, Jun 1998 to Dec 2013
Year first available 2014
Form of repository MATLAB .mat file
Availability https://data.mendeley.com/datasets/4yhprsgjrf/1
Size 371 kb

an efficient manner. However, existing software packages do not have the facilities to specify and utilize the multi-seasonal structures, and there are few efforts focused on automatic parameter re-estimation.

The objective of the present research is to develop a real-time oriented software toolkit with a suite of utilities tailored for msARI models. At the core of the toolkit is an adaptive forecasting routine. By applying the tools on various data series the power and flexibility of the real-time analysis tools are demonstrated. In this article, the second section of the paper reviews the previous work on modeling theory, software development, and visualization techniques. The third section introduces the representation of the multi-seasonal structure and algorithms developed around the structure. The fourth section discusses the architecture of the toolkit, the design of the Application Programming Interface (API), and the real-time visualization tools. In the fifth section three case studies are examined. The final section summarizes the findings originated from the research and proposes future directions of the study.

2. Related work

2.1. Time series models

Both linear and non-linear models have been suggested to model environmental time series. The most popular methods are AutoRegressive Integrated Moving-Average (ARIMA) models for linear models and ANN (Kaastra and Boyd, 1996) and SVM (Sapankevych and Sankar, 2009) for non-linear models. More recently, Tiwari and Adamowski (2013) proposed a hybrid wavelet-bootstrap-neural network (WBNN) model for short term (1 day–2 months) urban water demand forecasting. The model is presented as an ensemble of several ANNs to improve forecasts and characterize uncertainties. Bennett et al. (2014) utilized autoregressive integrated moving average with exogenous variables (ARIMAX) and ANN techniques to predict energy use in low-voltage distribution networks and suggested a hybrid approach to improve model performance. Sehgal et al. (2014) used wavelet-bootstrap-multiple linear regression (WBMLR) to forecast daily river discharges and observed better performance than multiple linear regression (MLR) and ANN models. Overall, the relative predictive power of the linear ARIMA models compared with non-linear ANN and SVM models vary from case to case, and the performance for a particular type of model still relies heavily on the empirical selection of the model structure (Sfetsos, 2000; Ho et al., 2002; Wang et al., 2009; Adhikari and Agrawal, 2013). This study will focus on the extension and application of traditional time series models because of the compatibility with explicit periodic model structures and adaptive parameter re-estimation methods.

The linear auto-correlations in equally intervalled univariate time series were investigated systematically in the fundamental work of Box and Jenkins (1976). This work introduced a group of linear Gaussian models, also known as ARIMA models. Moreover, a variant of the Box-Jenkins time series model, the seasonal ARIMA (sARIMA) model, was proposed to characterize periodic behaviors existing in many environmental and socio-economic phenomena. The single-seasonal model was originally proposed to predict air traffic. Over the years, the formulation has been adopted in other applications such as representing monthly lake water levels (Irvine and Eberhardt, 1992), monthly river flow (Kachroo and Liang, 1992), and hourly road traffic flow (Williams et al., 1998). A multi-seasonal ARIMA model has two or more periods incorporated into the model formulation. For example, Caiado (2009) studied the daily urban water demand based on a double-seasonal ARIMA model that accounted for weekly and annual relationships in water demands.

2.2. Time series software

Three existing software tools for building ARIMA models are investigated in this research. The tools include two popular scientific/statistical computing platforms (R and MATLAB) and one special-purpose software package (Gretl).

The R packages *stats* and *forecast* (Hyndman et al., 2014) include many utility functions for conducting static ARIMA analysis. For example, functions acf() and pacf() compute and plot the sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) for model identification. Function arima() accepts a data series and estimates the parameters for a non-seasonal or