Cyclic clustering approach to impute missing values for cyclostationary hydrological time series

Mohammad Reza Mahmoudi[1]

Abstract

In different scientific fields, some parts of the collected dataset may include incomplete data with at least one missing value. To tackle this issue, statisticians and machine learning experts have introduced different missing value imputation (MVI) approaches. Hydrological data are mostly time series with stationary (or cyclic) trends that its observations have high autocorrelation (or cyclic auto-correlation). In addition, these time series frequently exhibit random and non-constant oscillations from their trends. Therefore, the selected MVI approach for hydrological time series should consider these serious aspects. One aspect of hydrological time series data that has not been closely examined in the MVI studies up to now is the existence of cyclic trend, named cyclostationarity. In this paper, a novel MVI approach for cyclostationary hydrological time series, called cyclic clustering, will be discussed. The performance of the proposed method is studied by employing numerous simulated datasets. The results indicate that the cyclic clustering approach is robust to impute missing values for both stationary and cyclostationary hydrological time series. The applicability of the introduced approach is also investigated in a real-world problem.

Keywords Missing data · Missing value · Imputation · Hydrological time series · Cyclostationary time series · Clustering

1 Introduction

In different scientific fields, to make inference about an objective problem, a particular dataset is prepared. However, in real-world problems, some parts of the prepared dataset may include incomplete data with at least one missing value. The issue of missing values has been widely considered by statisticians and others (Schafer and Graham 2002). The missing values problem is a usual issue in geographical, hydrological, climatological and meteorological databases (Elshorbagy et al. 2002; Gill et al. 2007a; Kim and Ahn 2009; Yozgatligil et al. 2013; Aissia et al. 2017; Dikbas 2017; Gao et al. 2018; Chen et al. 2019; Dembélé et al. 2019; Militino et al. 2019; Hristopulos and Baxevani 2020; Arriagada et al. 2021). The datasets may be missed because of different reasons such as error

✉ Mohammad Reza Mahmoudi
mahmoudi.m.r@fasau.ac.ir

1 Department of Statistics, Faculty of Science, Fasa University, Fasa, Fars, Iran