Short Communication

iMedPub Journals
http://www.imedpub.com

Global Environment,
Health and Safety

2022

Vol 6. No. S1

The Known Sub-Sequence Algorithm (KSSA) for making optimal decisions about imputation methods for multiple missing data in time series for environmental research

Ivan Felipe Benavides

Instituto de Estudios del Pacífico, Universidad Nacional de Colombia, sede Tumaco

ibenavidesm@unal.edu.co

Citation: Benavides, The Known Sub-Sequence Algorithm (KSSA) for making optimal decisions about imputation methods for multiple missing data in time series for environmental research. Glob Environ Health Saf.2022, 6:S1:3

Abstract

Missing data in time series is a frequent problem for the environmental sciences. This is a serious limitation for statistical analysis and therefore, imputation (the process of filling missing data) is a keystone task. Several imputation methods have been proposed and implemented in programming software, however, their efficiency is data-dependent. There is no universal imputation method best for all-time series, but instead, each method suits the structure of particular groups of time series. ¿Which imputation method is best to fill a time series? the main problem is that the target time series (of interest for imputation) already contains missing data, so validation of methods cannot be performed directly on it. Instead it needs a full time series (no missing data) to simulate missing data, perform imputations and compare actual to impute. However, the best imputation method for the full time series is not necessarily the best for the target time series. The Known Sub- Sequence Algorithm (KSSA) is a novel approach to solve this problem by validating imputation methods directly on target time series. It uses the information contained within sub-sequences between missing data gaps to produce an optimal decision about a best imputation method for any particular target time series, no matter the structure it has. This is done by means of a process of iterative bootstrapping that randomly samples sub-sequences of the target time series in order to learn from them to find a best method form a set of candidates. This is a promising machine learning algorithm that will help environmental scientists and decision makers working with time series. KSSA will soon be implemented as the 'kssa' R-package in CRAN and is currently available in GitHub.

Received: March 08, 2022; **Accepted:** March 11, 2022; **Published:** March 15, 2022

Biography

Ivan Felipe Benavides is Biologist from Universidad de Narino in Colombia, PhD in Ecology from Universidad Austral de Chile, current postdoc researcher at Universidad Nacional de Colombia, and consultant for SoftwareShop and Datambiente in Latin America for environmental issues. He is an expert in environmental data science, which includes biostatistics, modeling, experimental design, machine learning and algorithm development to solve environmental problems.

References

1. Benavides et al (2022) Assessing methods for multiple imputation of systematic missing data in marine fisheries time series with a new validation algorithm. Aquaculture and Fisheries.[CrossRef] [GoogleScholar] [Indexed at].
2. Benavides et al (2022). Forecasting marine fishery landings for six species in the Colombian Pacific Ocean using SARIMA models Applied Sciences. In prep. [CrossRef] [GoogleScholar] [Indexed at].
3. Salazar et al 2021. Generalized additive models with delayed effects and spatial autocorrelation patterns to improve the spatiotemporal prediction of the skipjack (Katsunowus pelamis) distribution in the Colombian Pacific Ocean. RSMA.[CrossRef] [GoogleScholar] [Indexed at].
4. Molina-Cuaichar et al (2021) Evaluation of physical and chemical soil properties under different management types in the south-western Colombian Andes. Forest Systems.[CrossRef] [GoogleScholar] [Indexed at].
5. Benavides et al (2018). The variation of infiltration rates and physical- chemical soil properties across a land cover and land use gradient in a Paramo of southwestern Colombia. JSWC.[CrossRef] [GoogleScholar] [Indexed at].