

## Evaluation of Methods for Data Imputation in Time Series of Meteorological Variables of Northern Chile Using Data Analysis Techniques\*

<sup>1</sup>Sergio CERDA, <sup>2</sup>Francisco GARCÍA, <sup>3</sup>David CONTRERAS and <sup>4</sup>Alonso INOSTROSA-PSIJAS

<sup>1,2,3</sup>Universidad Arturo Prat, Iquique, Chile

<sup>4</sup>Universidad de Valparaíso, Escuela de Ingeniería Informática, Valparaíso, Chile

Correspondence should be addressed to: Sergio CERDA; [sergio.cerda.lozano@unap.cl](mailto:sergio.cerda.lozano@unap.cl)

\* Presented at the 40<sup>th</sup> IBIMA International Conference, 23-24 November 2022, Seville, Spain

Copyright © 2022. Sergio CERDA, Francisco GARCÍA, David CONTRERAS and Alonso INOSTROSA-PSIJAS

### Abstract

The study of climate change through extreme climate indicators to mitigate its effects has become significantly relevant in recent years. These indicators require temperature and precipitation data from weather stations for their calculation, their completeness being essential. Unfortunately, in the case of Northern Chile, the time series of meteorological stations present problems of empty data. Thus, in this article, different methods for data imputation are evaluated to identify those that can improve the series' quality regarding its completeness and representativeness in the behavior of meteorological variables. The CRISP-DM methodology adapted to structure the research phases is used for the above. Residual error analysis and correlation are carried out to evaluate the methods, highlighting that the best methods are CLP, IDC, and RN depending on the variable. Finally, concluding that there is more than one method with minimum residual error and positive correlation that can be used for the imputation of data in the meteorological variables of Northern Chile.

**Keywords:** Series Temporales, variables meteorológicas, métodos de imputación.