

Evaluating the Influence of Missing Data From the Crop Vegetation Index Time Series on Copernicus HR-VPP Phenological Products

Alexey Valero-Jorge ¹, M^a. Auxiliadora Casterad ¹ and José Tomás Alcalá ²

¹ Departamento de Sistemas Agrícolas Forestales y Medio Ambiente (Unidad asociada a EEAD-CSIC Suelos y Riegos). Centro de Investigación y Tecnología Agroalimentaria de Aragón.

² Departamento de Métodos Estadísticos, Instituto Universitario de Investigación en Matemáticas y Aplicaciones (IUMA). Universidad de Zaragoza.

Abstract: Phenological parameters extracted from the time series (TS) of spectral indices are crucial for the characterization of crops. The accuracy of these parameters can be affected by missing data in the TS. The Copernicus Land Monitoring Service (CLMS) provides these parameters and information on their quality in the period of occurrence. The aim of this work was to assess, for extensive herbaceous crops, the impact of missing vegetation index time data on the phenological parameters provided by the CLMS at the time of occurrence. A methodology was proposed using the package developed in this research, TSGenerator, for the download, processing, and analysis of the data applied to evaluate the parameters at the start and end of the season and the day of maximum development (SOS, EOS, and MAX) between the years 2018 and 2023. It used 252 images of the BIOPAR-VI module products, 6 images of phenology parameters, and 2025 barley and maize plots from Monegros and Zaidín irrigated lands in Spain. In barley, SOS and MAX were the most affected, with an average of 42.9% and 40.9 % of missing data according to the Copernicus HR-VPP product evaluation parameter (± 21 days). However, for maize, the most affected parameters were SOS and EOS, with 36.6% and 41.0 % of missing data. The correlation between the QFLAG-VPP quality parameter provided by Copernicus and the one proposed in this study, and the average percentage of non-missing data in the two most affected parameters, was $r = 0.89$ for barley and $r = 0.74$ for maize. The vegetation index series for barley presented about a 50 % probability of missing data in SOS and MAX according to a generalized additive fit model. A similar percentage was given for maize at SOS and EOS time points. This work represents an advance in the knowledge of the effect of missing data at the specific times of SOS, EOS, and MAX.

Keywords: missing data; time series; crops; Sentinel-2; phenological metrics