

## The Spatial-temporal Exponential distance-based Fuzzy clustering of COVID-19 data based on B-splines

Pierpaolo D'Urso<sup>1</sup>; Livia De Giovanni<sup>2</sup>; Vincenzina Vitale<sup>1\*</sup>

<sup>1</sup> Department of Social end Economic Sciences, Sapienza University of Rome, P.za Aldo Moro,5 00185 Rome, Italy. E-mail: pierpaolo.durso@uniroma1.it (P. D'Urso), vincenzina.vitale@uniroma1.it (V. Vitale)

<sup>2</sup> Department of Political Sciences, LUISS university, Viale Romania, 32 00197 Rome, Italy. E-mail: Idegiovanni@luiss.it (L. De Giovanni),

## Abstract:

Italy was the first European country seriously hit by pandemic. Its Northern regions of Lombardy and Veneto early experienced the negative consequences of COVID-19 pandemic on their regional health systems. After a first total lockdown, in March 2020, Italy faced other two pandemic waves, that involved all regions from the North to the South. The highest price has been paid in terms of deaths, especially among the older population. In this work, the interest has been that of clustering the 20 Italian regions according to the daily time-series of the cumulative cases over population (per 10,000 inhabitants), of the cumulative cases over monitored cases and of the cumulative deaths over population (per 10 000 inhabitants). Data, spanned from 2020-02-24 to 2021-02-08, have been those officially provided by the Italian Civil Protection Department.

The proposed fuzzy clustering model embedded both space and time information, the former by adding a spatial penalty term to the objective function (Coppi et al, 2010), the latter by transforming each time series onto (finite dimensional) vectors of coefficients associated to the functional basis of the Cubic B-splines, thus reducing computational burden.

In addition, the model is able to deal with outliers since a suitable exponential transformation of the Euclidean distance (Wu & Yang, 2002) has been introduced in the Fuzzy c-Medoids Algorithm.

The clustering results showed model capability to intercept both global and local spatial outliers (like Aosta Valley and Sicily) identifying the same partition into three increasing risk profiles (the so called yellow, orange and red areas) defined by the Experts Committee and based on 21 selected indicators. The spatial penalty term, based on a contiguity matrix, has been able to take into account the effect due to the positive autocorrelation among units.

## Keywords:

Robust algorithm; Partitioning Around Medoids; Time-series; Italian data; pandemic

## References

Coppi, R., D'Urso, P., Giordani, P., 2010. A fuzzy clustering model for multivariate spatial time series. J. Classification 27 (1), 54–88.

Wu, K.-L., Yang, M.-S., 2002. Alternative c-means clustering algorithms. Pattern Recognit. 35 (10), 2267–2278.