



# 10<sup>th</sup> INTERNATIONAL WORKSHOP ADVANCES IN CLEANER PRODUCTION

“TEN YEARS WORKING TOGETHER FOR A SUSTAINABLE FUTURE”

## Hybrid fuzzy c-means model for solar energy facilities clustering on American contaminated sites

FRANCO, D. G. B. <sup>a\*</sup>, STEINER, M. T. A. <sup>a</sup>

*a. Pontifícia Universidade Católica do Paraná, Curitiba*

*\*Corresponding author, david.barros@pucpr.br*

### Abstract

The present article used a hybrid fuzzy c-means model to cluster and define suitable locations, in terms of mapped area, distance to transmission lines and daily solar incidence, for solar energy capture facilities in the continental United States. The data used came from the National Solar Radiation Database (NSRDB), a collection of hourly measurements of solar radiation and meteorological data, and the RE-Powering America's Land project of the United States Environmental Protection Agency (EPA), whose purpose is to identify abandoned and contaminated areas that are ideal for renewable energy projects. Initially, data preprocessing was performed for substitution of missing data, normalization and principal component analysis (PCA). Then, the proposed hybrid clustering algorithm was applied. It is a fuzzy c-means model initialized by metaheuristics, namely genetic algorithm (GA), differential evolution (DE) and particle swarm optimization (PSO). The number of clusters was validated by three metrics: Calinski-Harabasz Index, Davies-Bouldin Index and Silhouette Coefficient. The three tests were unanimous, indicating two clusters as the ideal number, that is, a cluster for locations with potential for allocation of solar energy capture facilities and another for sites with no potential. As a result of the proposed hybrid approach, there was an increase in the training speed of the fuzzy c-means algorithm, which required a smaller number of iterations to reach the same objective function value. Visually, we can see the predominance of the allocation of the facilities in states with a higher average incidence of solar radiation, which is therefore the predominant factor in the convergence of the algorithm, which is in line with what was expected. Finally, the environmental-economic-social gains are considered with the revitalization of unproductive and contaminated land for the implantation of solar plants.

**Keywords:** *Clustering; Fuzzy c-means; Metaheuristics; Solar energy; Soil reuse.*