

2025-06-05

## Partitioning Around Medoids on Product Spaces: A Clustering Approach for Cylindrical Data

Yahia Hammami

*Université de Carthage, yahia.hammami@ept.ucar.tn*

Houyem Demni

*University of Cassino and Southern Lazio, houyem.demni@unicas.it*

Amor Messaoud

*Université de Carthage, amor.messaoud@ept.ucar.tn*

*See next page for additional authors*

Follow this and additional works at: <https://arrow.tudublin.ie/saml>



Part of the [Statistics and Probability Commons](#)

### Recommended Citation

Hammami, Yahia; Demni, Houyem; Messaoud, Amor; and Porzio, Giovanni C., "Partitioning Around Medoids on Product Spaces: A Clustering Approach for Cylindrical Data" (2025). *SAML-25 Workshop on Statistical and Machine Learning*. 16.

<https://arrow.tudublin.ie/saml/16>

This Article is brought to you by the EUT+ Academic Press a free to read and publish press of the European University of Technology.



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 4.0 International License](#).

---

## Authors

Yahia Hammami, Houyem Demni, Amor Messaoud, and Giovanni C. Porzio

# Partitioning Around Medoids on Product Spaces: A Clustering Approach for Cylindrical Data

Yahia Hammami

Université de Carthage, Ecole Polytechnique de  
Tunisie

LR99ES22 Laboratoire d'Economie et de Gestion  
Industrielle (LEGI-EPT)

La Marsa, Tunisia

yahia.hammami@ept.ucar.tn

Amor Messaoud

Université de Carthage, Ecole Polytechnique de  
Tunisie

LR99ES22 Laboratoire d'Economie et de Gestion  
Industrielle (LEGI-EPT)

La Marsa, Tunisia

amor.messaoud@ept.ucar.tn

Houyem Demni

University of Cassino and Southern Lazio,  
European University of Technology (EUt+)

Cassino, Italy

houyem.demni@unicas.it

Giovanni C. Porzio

University of Cassino and Southern Lazio,  
European University of Technology (EUt+)

Cassino, Italy

porzio@unicas.it

## Abstract

Clustering is a common unsupervised task in data analysis and machine learning. It deals with finding clusters of objects that are characterized by the highest similarity within the same cluster and the highest dissimilarity between different clusters.

One of the most used algorithms in clustering is the popular Partitioning Around Medoids (PAM), also known as k-medoids [4, 5]. The algorithm imposes the center of clusters to be some of the data points, and it looks for a minimal value of the sum of the dissimilarity to all the objects. One of the recognized properties of such a method is its robustness toward outliers due to the minimization of the total dissimilarity to other points.

This clustering technique has been designed for linear data. However, observations that cannot be depicted in an Euclidean space but instead in more complex manifolds are often encountered in applications. If so, suitable variations are needed.

Particularly, this work extends the PAM algorithm to data consisting of both linear and circular variables, that is, to data lying on the surface of a cylinder. Clustering this kind of data is a quite challenging task due to the complexity of the product space. Cylindrical data are described by a circular and a linear component measured on different scales. The most important difference between the two components is that on a circular scale the data points are periodic, that is,  $0^\circ$  and  $360^\circ$  represent the same value. On a linear scale instead, the values  $0^\circ$  and  $360^\circ$  are located in different places. This information must be properly accounted for by any similarity or distance measure: standard clustering methods cannot be applied inherently to these mixed data types.

Cylindrical data appear in many fields and have a wide range of applications. For instance, the HSV (Hue, Saturation, Value) color space features in color image processing has a

periodic hue component [2]. In environmental sciences, meteorological data often combine wind direction with wind speed, temperature,  $\text{SO}_2$  concentration, or other air quality indicators [3]. In fire ecology, the fire perimeter orientation can be considered as a two-dimensional or a three-dimensional observation, yielding circular or spherical data. In combination with the fire size, this leads to cylindrical data [1].

The proposed cylindrical PAM method adopts a similarity measure derived from a probabilistic model. The angular component is assumed to be drawn from a von Mises distribution, while the linear components are drawn from a Gaussian distribution. The performance of the proposed cylindrical PAM is evaluated through some numerical and real examples. A comparison is also given to demonstrate its effectiveness in handling mixed data types.

## Keywords

Circular and linear variables, k-medoids, similarity.

## Acknowledgments

The work of Houyem Demni and Giovanni C. Porzio was funded by the project ECS 0000024 “Ecosistema dell’innovazione - Rome Technopole” financed by EU in NextGenerationEU plan through MUR Decree n. 1051 23.06.2022 PNRR Missione 4 Componente 2 Investimento 1.5 - CUP H33C2200042000.

## References

- [1] A. M. Barros, J. M. Pereira, and U. J. Lund. 2012. Identifying geographical patterns of wildfire orientation: A watershed-based analysis. *Forest Ecology and Management* 264 (2012), 98–107.
- [2] A. Fujita, D. Y. Takahashi, and A. G. Patriota. 2017. A clustering method for data in cylindrical coordinates. *Mathematical Problems in Engineering* 2017 (2017), 3696850. doi:10.1155/2017/3696850
- [3] E. García-Portugués, R. M. Crujeiras, and W. González-Manteiga. 2013. Exploring wind direction and  $\text{SO}_2$  concentration by circular-linear density estimation. *Stochastic Environmental Research and Risk Assessment* 27 (2013), 1055–1067.

- [4] L. Kaufman. 1990. Partitioning around medoids (program PAM). In *Finding Groups in Data*. Vol. 344. Wiley, 68–125.
- [5] L. Kaufman and P. J. Rousseeuw. 1987. Clustering by means of medoids. In *Statistical Data Analysis Based on the  $L_1$ -Norm and Related Methods*, Yadolah Dodge (Ed.). North-Holland, 405–416.