Deployment of Renewable Energy Sources: Empirical Evidence in Identifying Clusters with Dynamic Time Warping

[1]Ann Maharaj, [2]Livia De Giovanni, [3]Pierpaolo D'Urso, [4]Mita Bhattacharya

[1,4]Monash University, Australia, [2]LUISS, Rome, Italy, [3]Sapienza, University of Rome, Italy

Workshop on "Statistical Computing and Robust Inference for High Dimensional Data", Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, December 11-14 2023

Overview

Background

- Deployment of renewable energy sources has caused a seismic shift in the world energy arena.
- Individual and coordinated effort across countries and regions is shaping the world in future, business models, and are supported globally to achieve net zero goals by 2050.
- In the current environment, renewable energy is widely adopted across many regions and countries to rebalance energy portfolios and reduce emissions from greenhouse gas emissions for sustainable development.
- Several authors including *Gugler et al.(2016), Khan et al. (2020) and Bhattacharya et al. (2022)* have researched various aspects of renewable energy in various parts of the world.

Motivation

- Renewable energy clusters share common advantages and difficulties in terms of transferring the key characteristics in deploying renewable energy resources across countries and regions.
- Clustering together of countries could capture similarities that has the potential to lead to common decisions that could benefit the countries in the same cluster.
- Energy clusters are becoming more and more common within a country informing groups with local entities acting in a common interest:
  - To improve energy security and efficiency for the region.
  - In meeting the Sustainable Development Goal 7 (SDG7) (affordable and clean energy) put forward by the United Nations https://sdgs.un.org/goals.
- Our aim is to identify clusters of countries at different levels of deployment of renewable energy sources.

Data

- We use data from the World Development Indicators (WDI) series maintained by the World Bank https://databank.worldbank.org/source/world-development-indicators.
- Renewable energy sources includes hydro, solid biofuels, wind, solar, liquid biofuels, biogas, geothermal, marine and waste as indicated in the World Bank Metadata Glossary https://databank.worldbank.org/metadataglossary/all/series
- We consider renewable energy as a percentage of primary energy supply for all available OECD and non-OECD countries with complete records from 1995 to 2018 (25 years), as well as with complete classifications according to four income groups and seven regions as indicated on the WDI website.

Methods

- We propose a **Fuzzy Clustering** model based on the **Partitioning Around Medoids (PAM)** with **Dynamic Time Warping (DTW)** to identify clusters of countries where within a particular cluster, the levels of deployment of renewable energy sources are similar while across clusters, they are different.
- **Fuzzy Clustering**: It appears more attractive than the traditional clustering methods in real applications characterised by a not clear separation among clusters the fuzzy memberships to the clusters show a possible second-best scenario uncovered by traditional clustering methods
- **Partitioning Around Medoids (PAM)**: In adopting this approach, the medoid time series are observed data and not "virtual" prototypes as in the fuzzy c-means method, making easier interpretation of the identified clusters
- **Dynamic Time Warping (DTW)**: the DTW distance or multivariate time series stretches or compresses the patterns of two objects locally in order to make their shape as similar as possible.

Dynamic Time Warping (DTW)

- Univariate time series are usually algebraically represented as a two-way data array or time data array, of

$$\mathrm{X} \equiv \{x_{it} : i = 1, \ldots, I;\; t = 1, \ldots, T\}$$

  where i indicates the generic unit (object) and t the generic time; x_{it} represents the variable observed in the i-th unit at time t.

- Let $\mathrm{x}_i \equiv \{x_{i1}, \ldots, x_{it}, \ldots, x_{iT}\}$ and $\mathrm{x}_{i'} \equiv \{x_{i'1}, \ldots, x_{i't'}, \ldots, x_{i'T'}\}$ be two time series, where T and T' need not be identical. In the DTW framework, they are defined as the "query" (or test) and the "reference" object respectively.

- The total distance between x_i and $\mathrm{x}_{i'}$ is then computed through the so-called "warping curve", or "warping path", which ensures that each data point in x_i is compared to the "closest" data point in $\mathrm{x}_{i'}$.

Dynamic Time Warping (DTW)

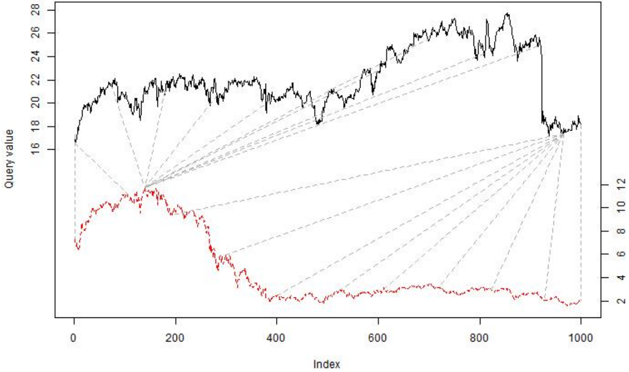


Figure: Dynamic Time Warping distance

Fuzzy Clustering with Dynamic Time Warping

Let $\mathrm{X} = (\mathrm{x}_1, \ldots, \mathrm{x}_I)$ be a set of I univariate time series, and $\widetilde{\mathrm{X}} = (\widetilde{\mathrm{x}}_1, \ldots, \widetilde{\mathrm{x}}_C)$ a subset of X with cardinality C. For classifying univariate time series we consider the following Dynamic Time Warping Fuzzy C-Medoids (DTW-FCMd) clustering model:

$$\begin{cases} \min\limits_{u_{ic}} : \sum\limits_{c=1}^{C} u_{ic}^m D^2(\mathrm{x}_i, \widetilde{\mathrm{x}}_c) = \sum\limits_{c=1}^{C} u_{ic}^m \sum_{l=1}^{L} (x_{i,\hat{\varphi}_l} - \widetilde{x}_{c,\hat{\psi}_l})^2 m_{l,\hat{\Phi}} \\ \text{s.t.: } \sum\limits_{c=1}^{C} u_{ic} = 1, \quad u_{ic} \geq 0, \end{cases} \tag{1}$$

where

- $D^2(\mathrm{x}_i, \widetilde{\mathrm{x}}_c)$ is the squared DTW distance computed between the i-th time series and the c-th medoid;
- u_{ic} is the membership degree of the i-th time series to the c-the cluster;
- $m > 1$ is the fuzziness parameter—the greater the value of m the more fuzzy is the obtained partition.

Then, each time series is allocated into the cluster corresponding to its closest medoid time series, based on their pairwise DTW distance.

Fuzzy Clustering with Dynamic Time Warping

The local optimal solution of (1) is:

$$u_{ic} = \frac{1}{\sum_{c'=1}^{C} \left[\frac{D^2(\mathrm{x}_i, \widetilde{\mathrm{x}}_c)}{D^2(\mathrm{x}_i, \widetilde{\mathrm{x}}_{c'})} \right]^{\frac{1}{m-1}}}. \tag{2}$$

- Since the solutions (2) of (1) are recursive, it is not guaranteed that the global minimum is reached; more than one random start is suggested to obtain a stable solution.
- The fuzziness parameter m determines the shapes of the clusters and produces a fine tuning between the membership degrees close to 0 or 1 and those with intermediate values.
  - As m increases, the fuzzier the membership degrees are. Conversely, if m is close to 1, the resulting partition is hard.
  - In the fuzzy clustering literature, several heuristic procedures to select m have been proposed but there is a lack of sounding theoretical basis to justify the selection of the fuzziness parameter. For this reason, as it is suggested in literature, we set $m = 1.5$

Fuzzy Internal Validity Indices

- The optimal number of clusters C can be determined by considering internal validity indices for fuzzy clustering.
- We consider the following internal validity indices:
  - **Fuzzy Silhoutte (FS) Index**: Takes on the maximum value for the optimal number of clusters. (*Campello and Hruschka, 2006*)
  - **Xie-Beni (XB) Index**: Takes on the minimum value for the optimal number of clusters. (*Xie and Beni, 1991*)
  - **Partition Coefficient (PC) Index**: Takes on the maximum value for the optimal number of clusters. (*Bezdek, 1981*)
  - **Partition Entropy (PE)**: Takes on the minimum value for the optimal number of clusters. (*Bezdek, 1981*)
  - **Modified Partition Coefficient (MPC) Index**: Takes on the maximum value for the optimal number of clusters. (*Davè, 1996*)

Fuzzy Membership Degrees and Hard Clusters

A suitable cut-off point of the highest membership values is used to determine it a case can be considered to have fuzzy membership across clusters (*Maharaj and D'Urso, 2011)*. In particular,

- In a 2-cluster solution: A case is considered to have fuzzy membership in two clusters if membership degrees are less than 0.7 in one cluster and greater that 0.3 in the other cluster.
- In a 3-cluster solution: A case is considered to have fuzzy membership in two clusters if membership degrees are less than 0.6 in one cluster and greater that 0.4 in the other cluster. A case is considered to have fuzzy membership in all three clusters if membership degrees are less that 0.4 but greater than 0.3 in all three clusters.
- In a 4-cluster solution: A case is considered to have fuzzy membership in two clusters if membership degrees, are less than 0.5 in any one cluster.

Application

Table: Distribution of OECD and Non-OECD countries

Non-OECD	OECD	Total
92 (71%)	38 (29%)	130

Table: Income Levels and Regions

Income Levels	
High	Lower-middle
Upper-middle	Low
Regions	
ECA: Europe and Central Asia	
EAP: East Asia and the Pacific	
LAC: Latin and Central America	
MENA: Middle East and North Africa	
NAM: North America	
SSA: Sub-Saharan Africa	
SA: South Asia	

Application: Categorical Summaries

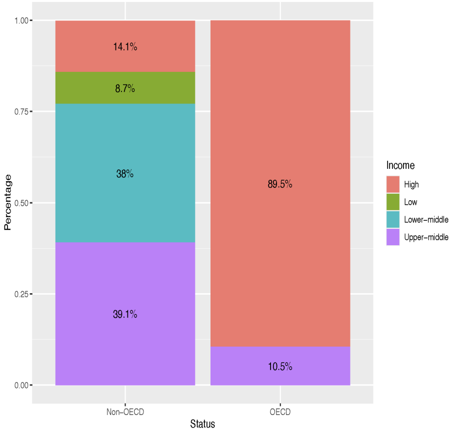


Figure: Countries by OECD membership and income levels

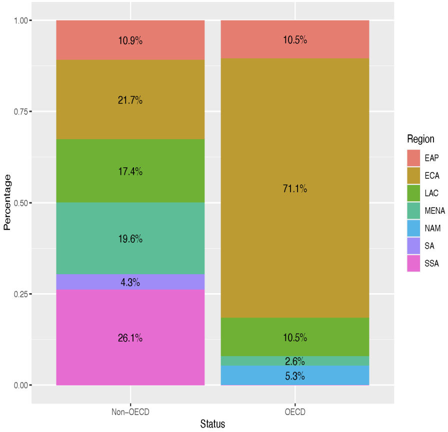


Figure: Countries by OECD membership and regions

Application:Categorical Summaries

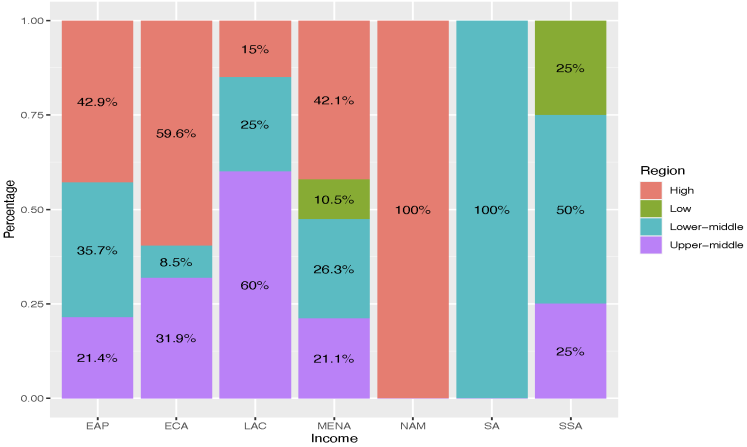


Figure: Countries by Income and Regions

Application: Cluster Solutions

Table: Cluster Validity Indices

m=1.5	FSIL	XB	PC	MPC	PE
2 clusters	0.4	3.3	0.6	0.2	0.6
3 clusters	0.8	0.1	0.9	0.8	0.2
4 clusters	0.7	0.6	0.8	0.8	0.3
5 clusters	0.7	0.4	0.8	0.7	0.4

- Based on all criteria, the 3-cluster solution is optimal, with largest values for FSIL and PC, equal largest value (for 4-cluster) for MPC, and smallest values for XB and PE.
- We also examine the 4-cluster solution given that except for the MPC and XB criteria, the values of the other criteria are second best.

Application: 3-Cluster Solution

Table: 3-Cluster solution

Cluster	Countries	Countries %
C1	26	20%
C2	36	28%
C3	68	52%
Total	130	100%

The medoids of the three clusters which determine the centres are Mynmar (EAP), Georgia (ECA) and United Kingdom (ECA), respectively.

- Cluster 1 have similar renewable energy usage as a percentage of primary energy supply to that of Mynmar.
- Cluster 2 have similar renewable energy usage as a percentage of primary energy supply to that of Georgia.
- Cluster 3 have similar renewable energy usage as a percentage of primary energy supply to that of United Kingdom.

Application: 3-Cluster Solution

Mean renewable energy as a percentage of primary energy supply from 1995 to 2018 for countries in each of the clusters.

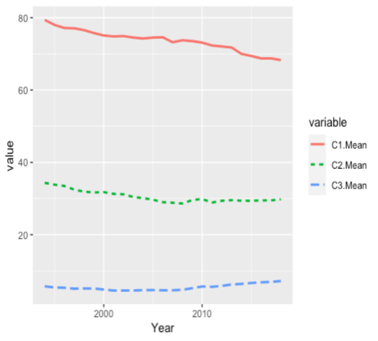


Figure: 3-Cluster Prototypes

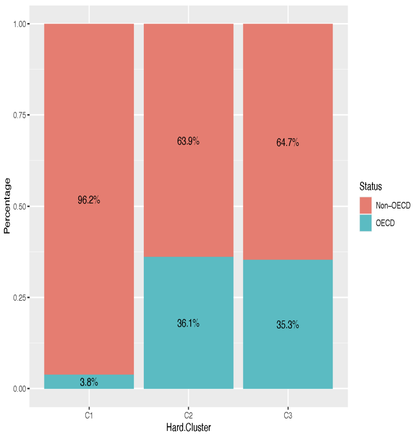


Figure: 3-Cluster solution across OECD and Non-OECD countries

Application: 3-Cluster Solution

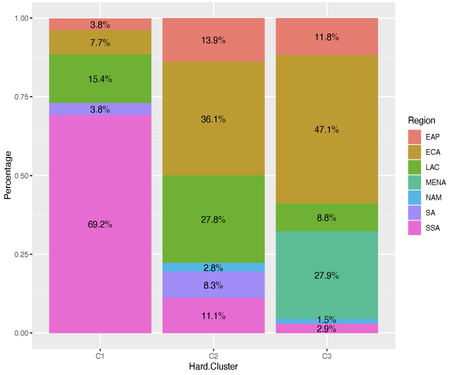


Figure: 3-Cluster solution across regions

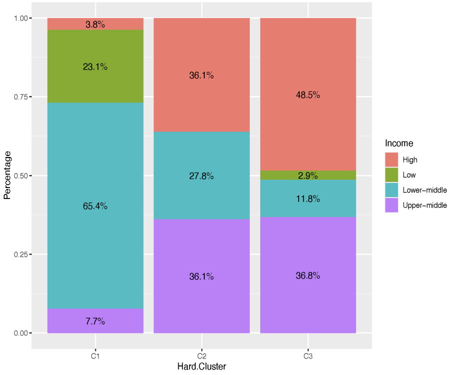


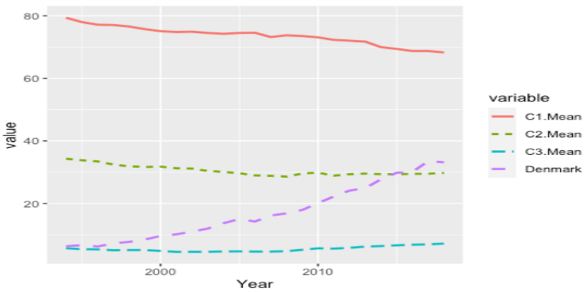
Figure: 3-Cluster solution across income groups

Application: Fuzzy 3-Cluster Solution

Table: Fuzzy 3-Cluster solution

Cluster	Number	Countries with Fuzzy Membership				Membership Degrees		
						C1	C2	C3
C1	24							
C2	37							
C2C3	1	Denmark	ECA	High	OECD	0.01	0.52	0.48
C3	68							
Total	130							

Fuzzy 3-Cluster Prototypes



Application: 4-Cluster Solution

Table: 4-Cluster solution

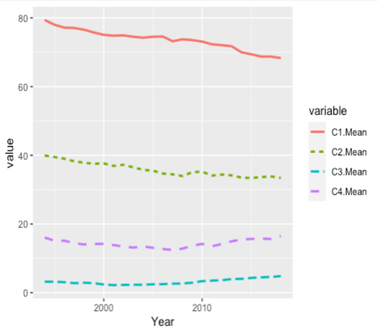
cluster	countries	countries %
C1	26	20%
C2	24	18%
C3	48	37%
C4	32	25%
Total	130	100%

- The medoids of Clusters 1 ,2 3, and 4 are Mynmar (EAP), Indonesia (EAP), Jordon (MENA) and Ecuador (LAC), respectively.
- Cluster 1 in the 4-cluster solutions remains the same as it was in the 3-cluster solution.
- 12 countries from Cluster 2 in the 3-cluster solution have moved to Cluster 4 in the 4-cluster solution.
- 20 countries from Cluster 3 in the 3-cluster solution have moved to Cluster 4 in the 4-cluster solution.

Application: 4-Cluster Solution

Mean renewable energy as a percentage of primary energy supply from 1995 to 2018 for countries in each of the clusters.

Figure: 4-Cluster prototypes



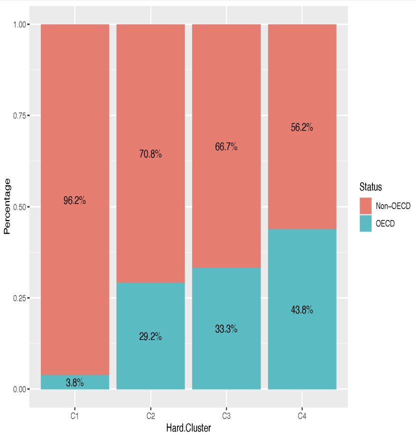


Figure: 4-Cluster solution across OECD and Non-OECD countries

Application: 4-Cluster Solution

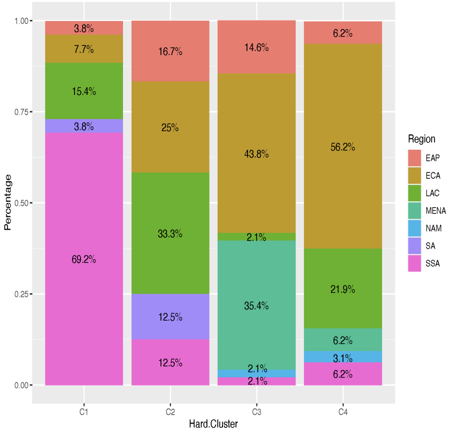


Figure: 4-Cluster solution across regions

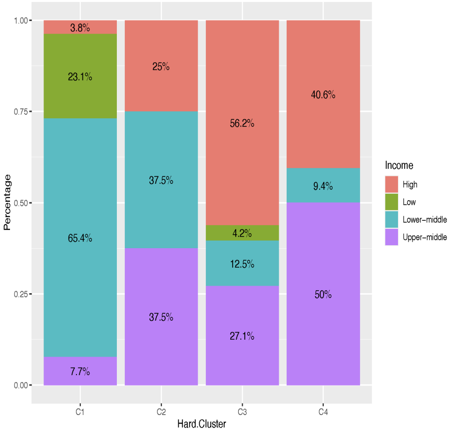
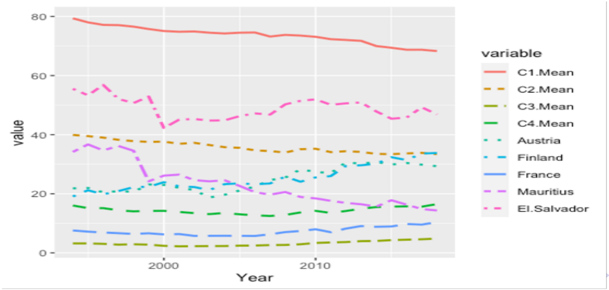


Figure: 4-Cluster solution across income groups

Application: Fuzzy 4-Cluster Solution

Cluster	Number	Countries with Fuzzy membership				C1	C2	C3	C4
C1	26								
C2	23								
C3	48								
C4	28								
C1C2	1	LAC	Lower-middle	Non-OECD	El.Salvador	0.42	0.48	0.03	0.07
C2C4	3	SSA	Upper-middle	Non-OECD	Mauritius	0.01	0.46	0.03	0.51
		ECA	High	OECD	Austria	0.02	0.45	0.07	0.47
		ECA	High	OECD	Finland	0.02	0.42	0.07	0.49
C3C4	1	ECA	High	OECD	France	0.00	0.02	0.46	0.52

Fuzzy 4-Cluster prototypes



Application: 4-Cluster Solution Summary

Mostly non-OECD countries in:

- Low or lower-middle groups (88% combined) which are mainly in SSA and together with countries from the LAC region (84% combined) have the highest level of renewable energy as a percentage of primary energy supply for most of the period between 1995 and 2018.
- Lower-middle and upper-middle income groups (76% combined) from the ECA, LAC region and EAP regions (95% combined) have the second highest level of renewable energy as a percentage of primary energy supply for most of the period between 1995 and 2018.

Application: 4-Cluster Solution Summary

- Mostly OECD countries in the high-income group, and non-OECD countries in the upper-high income groups (83% combined) which are mainly from the ECA, MENA, and the EAP regions (94% combined) have the lowest level of renewable energy as a percentage of primary energy supply for most of the period between 1995 and 2018.
- A combination of OECD in the high-income group and non-OECD countries in the upper-high income groups which are mostly from European and Central Asia (ECA) region have the second lowest level of renewable energy as a percentage of primary energy supply for most of the period between 1995 and 2018.

Application: 4-Cluster Solution Summary

The level of renewable energy as a percentage of primary energy supply of

- El Salvador which is lower-middle income, non-OECD country in the LAC region is simultaneously compatible with that of countries with the highest and second highest levels.

- Mauritius, which is a upper-middle income, non-OECD country in the SSA region, is simultaneously compatible with that of countries with the second highest and second lowest levels.

- Austria and Finland, which are high-income, OECD countries in the ECA region, are simultaneously compatible with that of countries with the second highest and second lowest levels.

- France, which is a high-income, OECD country in the ECA region, is simultaneously compatible with that of countries with the lowest and second lowest levels.

Concluding Remarks

- From the 3- and 4-cluster analyses, it is evident that deployment of renewable sources are different for different income regions.
- The 4-cluster solution provides a better separation of countries.
- In most cases, deployment is higher in low income countries.
  - This is due to the rapid acceleration of renewable deployment in recent years for these countries.
  - Various government support programs and introduction of various renewable sources have been supporting this rapid deployment.
- Deployment of renewables in both low- and high-income countries is now shifting gradually from the fringe to the mainstream of sustainable development.
- This analysis suggests that clustering countries in terms of income could contribute to implementing energy policies in a coordinated manner.
- In future work, we will include in our analysis, renewable energy deployment data for the years after 2018, when they become available.

Published in Social Indicators Research https://doi.org/10.1007/s11205-022-03050-0

References

1. Bezdek, J.C. (1981). *Pattern recognition with fuzzy objective function algorithms*, Kluwer Academic Publishers
2. Bhattacharya M, Inekwe J and Liddle B, (2022), The role of renewable energy sources in residential electricity prices: A club convergence analysis across selected European countries, *Applied Economics*.
3. Campello, R.J.G.B, Hruschka, E. (2006). A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems*, **157(21)**, 2858-2875. Davè, R. (1996). Validating fuzzy partitions obtained through c-shells clustering, *Pattern Recognition Letters*, **17(6)**, 613-623
4. Gugler, K., Haxhimusa, A., Liebensteiner, M., & Schindler, N. (2020). Investment opportunities, uncertainty, and renewables in European electricity markets.*Energy Economics*, **85**, 104575.
5. Maharaj, E.A., De Giovanni, L., D'Urso, P., Bhattacharya M, (2022). Deployment of Renewable Energy Sources: Empirical Evidence in Identifying Clusters with Dynamic Time Warping. *Social Indicators Research* . https://doi.org/10.1007/s11205-022-03050-0.
6. Khan, Z., Ali, S., Umar, M., Kirikkaleli, D., & Jiao, Z. (2020). Consumption-based carbon emissions and international trade in G7 countries: the role of environmental innovation and renewable energy. *Science of the Total Environment*, **730**, 138945.
7. Maharaj E.A., D'Urso, P. (2011). Fuzzy clustering of time series in the frequency domain. *Information Sciences*, **181**, 1187-1211.
8. Xie, X., Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13(8)**, 841-847.