

## Research Article

# Hybrid Unsupervised Exploratory Plots: A Case Study of Analysing Foreign Direct Investment

Álvaro Herrero <sup>1</sup>, Alfredo Jiménez,<sup>2</sup> and Secil Bayraktar<sup>3</sup>

<sup>1</sup>Grupo de Inteligencia Computacional Aplicada (GICAP), Departamento de Ingeniería Civil, Escuela Politécnica Superior, Universidad de Burgos, Av. Cantabria s/n, 09006 Burgos, Spain

<sup>2</sup>Department of Management, KEDGE Business School, 680 Cours de la Libération, 33405 Talence, Bordeaux, France

<sup>3</sup>Department of Corporate Social Responsibility and Human Resources, Toulouse Business School, 20 Boulevard Lascrosses, 31068 Toulouse, France

Correspondence should be addressed to Álvaro Herrero; [ahcosio@ubu.es](mailto:ahcosio@ubu.es)

Received 20 December 2018; Accepted 29 January 2019; Published 2 June 2019

Guest Editor: Jose Garcia-Rodriguez

Copyright © 2019 Álvaro Herrero et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The curse of dimensionality has been an open issue for many years and still is, as finding nonobvious and previously unknown patterns in ever-increasing amounts of high-dimensional data is not an easy task. Advancing in descriptive data analysis, the present paper proposes Hybrid Unsupervised Exploratory Plots (HUEPs) as a new visualization technique to combine the outputs of Exploratory Projection Pursuit and Clustering methods in a novel and informative way. As a case study, HUEPs are validated in a real-world context for analysing the internationalization strategy of companies, by taking into account bilateral distance between home and host countries. As a multifaceted concept, distance encompasses multiple dimensions. Together with data from both the countries and the companies, various psychic distances are analysed by means of HUEPs, to gain deep knowledge of the internationalization strategy of large Spanish companies. Informative visualizations are obtained from the analysed dataset, leading to useful business implications and decision making.

## 1. Introduction

As it is well known, there are many different ways of analysing unlabeled datasets in order to gain knowledge about them. A key challenge in the analysis of high-dimensional unknown data is to identify the patterns that exist across dimensional boundaries. Such patterns may become visible if a change is made to the basis of the space; however, an *a priori* decision as to which basis will reveal most patterns requires prior knowledge of the unknown patterns. This is the main idea behind Exploratory Projection Pursuit (EPP) [1]. As opposed to feature selection, EPP lies within the feature extraction paradigm, as the resulting dimensions are combinations (it could be linear or nonlinear) of the original features in the dataset. On the other hand, clustering [2] consists in the organization of a collection of data items or patterns into clusters based on similarity. Hence, patterns within the same cluster are more similar to each other than they are to a pattern belonging to a different cluster.

Both EPP and clustering methods have been widely applied and combined in previous work. Although it had been stated [3, 4] that dimensionality reduction may identify dimensions that do not enhance the results of a subsequent clustering, some authors have contributed to the main combination stream where dimensionality reduction and clustering methods are sequenced, namely, “tandem” approach. Furthermore, [5] pointed out that “cluster analysis is one of the most frequent contexts in which principal components are derived in order to reduce dimensionality prior to the use of a different multivariate technique”. That is the case of [6] where a canonical transformation is applied to data in order to optimize *k*-means clustering results on functional data. Additionally, [7] have optimized EPP as an initial step and then applied some clustering methods (hierarchical, partitional, and density-based) attaining interesting results. More recently, [8] have proposed Extreme Learning Machine for Joint Embedding and Clustering as a first step to preserve the manifold structure of the data in the original space

while maximizing the class separability of the data in the embedded space at the same time. Similarly, unsupervised dimensionality reduction methods are proposed in [9–12] for subsequent clustering through  $k$ -means.

In [13, 14] EPP and clustering methods are also combined but from a different perspective; dimensionality reduction models, implemented as neural networks, have been applied to add the output of some clustering methods to the obtained projections, though different labels, colours, and symbols. As a result, 2D projections are generated, enriched with information about the number of the cluster assigned to each sample. In those previous works, data from the cybersecurity and environmental fields have been analysed, respectively.

In a third alternative approach, clustering and EPP methods interact. Projection Pursuit Clustering [15] has been proposed to recover clusters in lower dimensional subspaces of the data by simultaneously performing dimension reduction and clustering. The proposed methodology finds both an optimal clustering for a subspace of given dimension and an optimal subspace for this clustering. In order to do that, clustering and projection pursuit methods are adapted in order to interchange information during execution. In a similar way, [16] has proposed a projection pursuit index to identify clusters and other structures in multivariate data, which is obtained from the variance decompositions of the data's one-dimensional projections.

Finally, some other independent uses of these two kinds of methods have been proposed so far [17]. In [18],  $k$ -means clustering method is applied in order to compare its results with those obtained from EPP. Thus, no combination of such methods is proposed but the comparison of their results, instead. On the other hand, both dimensionality reduction and clustering are independently combined in [19] for different tasks under the frame of a hybrid recommender system.

Differentiating from previous work, the present paper proposes the independent application of EPP methods on the one hand and clustering ones on the other. Complete results of the two of them are then combined, together with the glyph metaphor, in a novel way, called Hybrid Unsupervised Exploratory Plots (HUEPs), to support decision making. When compared to the above-mentioned previous work, it can be said that the present paper's proposal is a far more general and simpler approach where any EPP and clustering methods could be combined to generate informative and intuitive 3D visualizations of high-dimensional data. In order to validate this proposal, HUEPs are applied and compared in a case study where internationalization strategies from Spanish Multinational Enterprises (MNEs) are analysed.

In today's business context, management of international operations has been a focal element of company strategies. However, while investing abroad, companies face numerous challenges that, if not taken into consideration, may significantly risk the success of their investment. "Distance" emerges as a major challenge among those. A clear understanding of the differences between idiosyncrasies of the host country and home country may provide opportunities, on one hand, or, just the contrary, ignoring such differences may lead to disruption of the company's activities overseas. Referring to the fundamental role of distance between

countries in the field of international management, [20] has even explicitly stated that "*essentially, international management is management of distances*".

Recent work [21] has conceptualized distance as a multifaceted construct. Along similar lines, various frameworks have investigated the multiple dimensions of distance that may influence a company's international operations. For example the well-known CAGE framework [22] proposed distance to constitute cultural, administrative, geographic, and economic facets. Another framework further posited ten dimensions to capture distance between nations [23]. Moreover, some researchers have dissected these dimensions of distance into further subdimensions with the aim of comprehending this phenomenon better. For instance, cultural distance was proposed to include six dimensions (power distance, uncertainty avoidance, individualism, masculinity, long-term orientation, and indulgence) by the influential work by [24, 25]. The vast number of citations proves that this framework and its operationalization as a single construct [26] became widely popular.

Nevertheless, some recent criticism has raised that an important type of distance, psychic distance, cannot entirely be captured or measured by the current cultural dimensions even though it is a crucial variable influencing managerial decisions in international business [27]. Psychic distance is an extensive framework that goes beyond culture and entails multiple dimensions of distance [28, 29]. It is useful to understand the context in which a manager's perceptions are formed while making a decision. Reference [28] suggests that six macro factors called psychic distance stimuli shape that context [30]. These factors measure the national differences between language, industrial development, social system, democracy, education, and religion [28, 31]. Previous studies have shown that these stimuli significantly impact market selection, performance, entry mode choice, Foreign Direct Investments (FDI), online internationalization, and trade flows [32].

On the other hand, even though researchers [33, 34] confer that combining multiple stimuli into one single construct is problematic in the sense that it may cause an inaccurate view that all components are equally significant, many studies still follow this aggregation approach. This paper, being aware of this potential problem, focuses on one particular stimulus concerning the political system differences between countries, namely, democracy distance, in order to avoid the probable confounding effects of the other stimuli.

While all stimuli may have an important role, we focus on democracy because previous research has emphasized the critical impact of political institutions on FDI decisions [35–39]. The democracy distance variable indicates the level of political rights, civil liberties and checks and balances existing in the country to prevent any opportunistic behavior by the local government to unilaterally modify the rules and laws [28].

According to what has been explained above, the challenging task of analysing the internationalization strategy of companies requires advanced data analysis tools. Up to now, little effort has been devoted to support decision makers with means of getting deep knowledge from such datasets. The

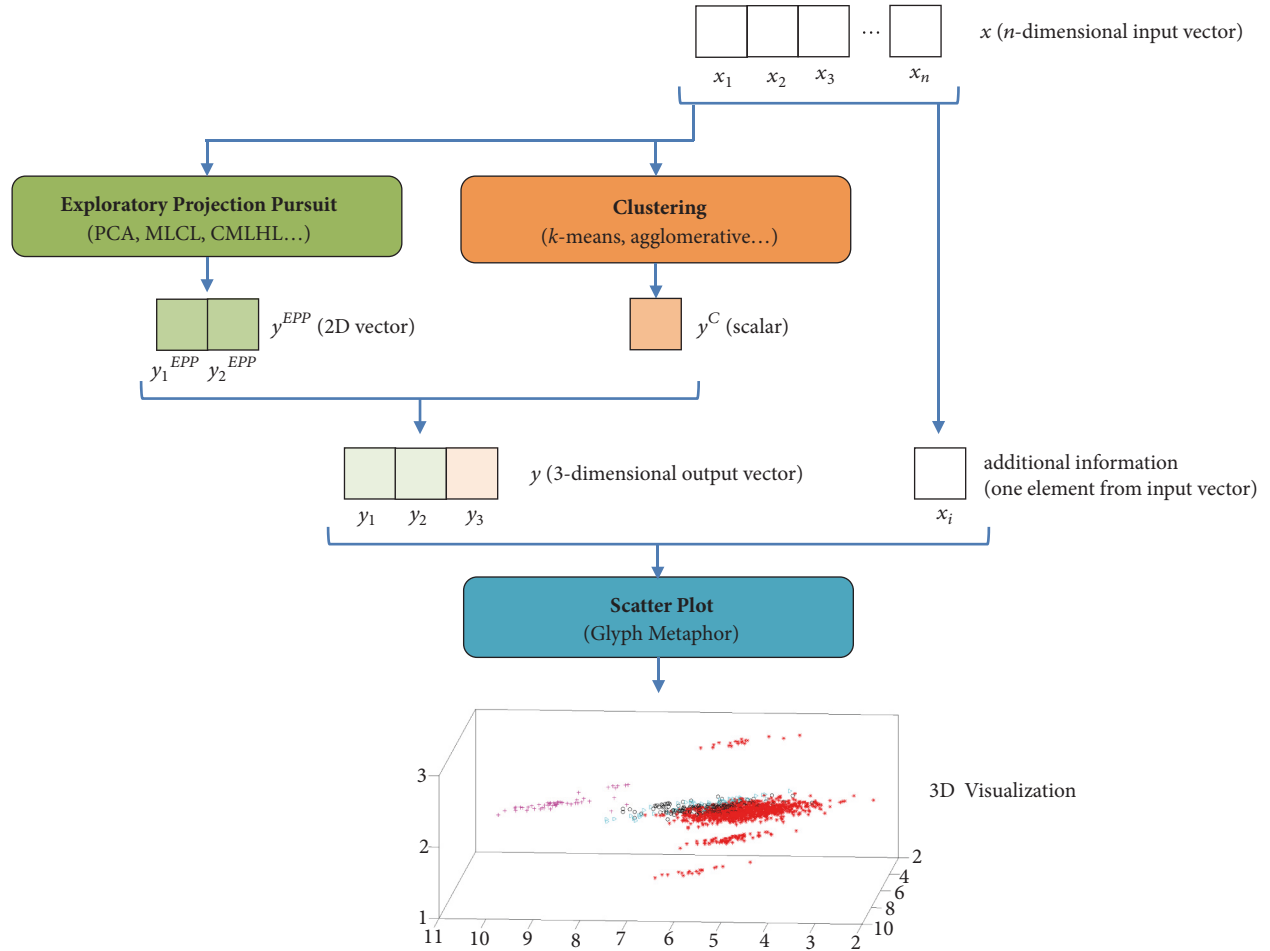


FIGURE 1: Process to obtain a HUEP.

present paper advances previous work by proposing a new visualization tool to ease the analysis of multidimensional datasets related to internationalization. The rest of this paper is organized as follows: the proposed HUEPs and their components are described in Section 2, the case study where HUEPs are validated is introduced in Section 3, together with its associated results that are presented in Section 4 and the conclusions of present study that are stated in Section 4.

## 2. Hybrid Unsupervised Exploratory Plots

As we humans are able to detect anomalies and to recognize different features or patterns through visual inspection, visualization techniques are a viable solution to information seeking. This idea is based on the ability to visualize high-dimensional datasets in a consistent and low-dimensional representation where those anomalies, features, or patterns can be identified. Such depiction of high-dimensional data through visual displays is not easy and cannot be performed immediately in most cases. The difficulty lies in converting raw big-size data into a graphical format that provides a useful insight into the visualized dataset [40]. As previously mentioned, Hybrid Unsupervised Exploratory Plots (HUEPs) are

proposed as a new way of intuitively visualizing data within the field of descriptive datamining.

Visualization techniques have been widely covered in the literature; two of the most relevant works are [41, 42]. Among the wide variety of such techniques, a very popular one is scatter plots, which represent 2D or 3D data as points, with coordinates that correspond to their values. These plots still are one of the most popular and widely used visual representations for multidimensional data [43], due to their simplicity. However, there are some drawbacks, the two main ones being the required low dimensionality of the data to be displayed and the problem of overplotting.

A HUEP is proposed as a scatter plot where each data is considered as a 3D vector. These three-dimensional vectors are obtained from (raw) original data by means of an EPP method and a clustering one, according to what is shown in Figure 1.

As can be seen in Figure 1, a HUEP can be described as a mapping of vectors  $x$  onto vectors  $y$  in an output space. Vectors from the input space ( $n$ -dimensional, being  $n \geq 3$ )  $S_I$  are then mapped into a 3D output space  $S_O$ , according to  $H$  nonlinear transformation:

$$H: S_O \longrightarrow S_I, \quad \text{for any } x \in S_O \quad (1)$$

Resulting vectors  $y$ , from the  $S_O$  space, are defined as

$$y = (y_1, y_2, y_3) \quad (2)$$

$y_1, y_2$  are the output vectors of an EPP method ( $y_1^{EPP}, y_2^{EPP}$ ) and  $y_3$  the output (scalar) of a clustering method. Once obtained, output  $y$  vectors are then plotted in 3D scatter plots. Furthermore, the visualization of each vector is enriched thanks to the glyph metaphor, as can be seen in Section 3 and adding additional information from one of the input features ( $x_i$ ). The widely used glyphs (or multidimensional icons) can be defined as graphical objects that are designed to convey multiple data values [44]. By using different symbols and colours, further information can be added to the 3D visualization of each data point.

Proposed HUEPs are hybrid as they combine both exploratory (dimensionality reduction) methods as well as clustering ones. On the other hand, they are unsupervised as both kinds of methods implement this kind of learning (no target class or value is provided to be reproduced).

The main steps to obtain HUEPs are described in the following subsections.

**2.1. Exploratory Projection Pursuit.** The well-known Exploratory Projection Pursuit (EPP) [1] was proposed as a method to identify structure in a given high-dimensional data. In the case of EPP, this general task is performed by projecting the data onto a low-dimensional subspace. By means of such projection, one can visually identify the structure of the dataset. As not all available projections reveal the data's structure in the same way, EPP defines an index aimed at measuring the "interestingness" of a projection, and then those projections that maximize that index are chosen.

As previously mentioned, EPP initially defines which indices represent interesting directions. When talking about projections, "interestingness" is usually linked to the fact that most projections give almost Gaussian distributions [45]. Consequently, in order to identify the most "interesting" features of the data, the directions generating projections as far from the Gaussian as possible should be found.

Once the most interesting projections are identified, the high-dimensional data are then projected onto a lower dimensional (2D or 3D) subspace, which makes it possible to visually examine the structure of the dataset. From the wide range of EPP projection methods that have been proposed until now, some neural implementations have been selected for HUEPs in present paper, as they have been successfully applied to a wide variety of fields and datasets.

**2.1.1. Principal Component Analysis.** Principal Component Analysis (PCA) is a statistical model that has been widely applied in last decade and still is applied at present time [46]. It was introduced in [47] and describes the variation in a high-dimensional dataset in terms of a set of uncorrelated variables (each one of these variables is a linear combination of the original ones). From a geometrical perspective, it consists of a rotation of the axes of the original coordinate system that generates a new set of orthogonal axes. In the case of PCA the new axes are ordered in terms of the amount

of variance of the original data they account for. As a result, the first axes (those accounting for the highest variance) are the ones selected to obtain the new visualization of data. It should be noted that even if we are able to visualize the data with a few variables, it does not follow that an interpretation will ensue, as it depends on the original dataset. As previously proposed [48, 49], PCA can be performed by means of neural networks.

**2.1.2. Maximum Likelihood Hebbian Learning.** Among all the neural alternatives of performing EPP, Maximum Likelihood Hebbian Learning [50] is one based on the Negative Feedback Network. It associates an input vector ( $x$ ) with an output vector ( $y$ ) computed as

$$y_i = \sum_{j=1}^N W_{ij} x_j, \quad \forall i \quad (3)$$

where  $W_{ij}$  is the weight linking input  $j$  to output  $i$ .

At the training stage, when the output of the neural network is calculated, the activation ( $e_j$ ) is fed back through the same weights and subtracted from the input:

$$e_j = x_j - \sum_{i=1}^M W_{ij} y_i, \quad \forall j \quad (4)$$

Finally, weights are updated according to the specific learning rule:

$$\Delta W_{ij} = \eta \cdot y_i \cdot \text{sign}(e_j) |e_j|^{p-1} \quad (5)$$

where  $\eta$  is the learning rate and  $p$  is a parameter related to the energy function.

**2.1.3. Cooperative Maximum Likelihood Hebbian Learning.** The Cooperative MLHL (CMLHL) model was proposed [51] as an extension of MLHL by adding lateral connections between neurons in the output layer of the network (see Equation (5)). CMLHL can be defined through (4)-(7), where an  $N$ -dimensional input vector ( $x$ ) is processed to obtain an  $M$ -dimensional output vector ( $y$ ).

(1) Feed-forward step:

$$y_i = \sum_{j=1}^N W_{ij} x_j, \quad \forall i \quad (6)$$

(2) Lateral activation passing:

$$y_i(t+1) = [y_i(t) + \tau(b - Ay)]^+ \quad (7)$$

(3) Feedback step:

$$e_j = x_j - \sum_{i=1}^M W_{ij} y_i, \quad \forall j \quad (8)$$

(4) Weight change:

$$\Delta W_{ij} = \eta \cdot y_i \cdot \text{sign}(e_j) |e_j|^{p-1} \quad (9)$$



where  $\tau$  is a parameter to model the “strength” of the lateral connections,  $b$  is the bias parameter,  $A$  is a symmetric matrix used to modify the response to the data,  $\eta$  is the learning rate, and  $p$  is a parameter related to the energy function.

**2.2. Clustering.** EPP has been described in the section above as a method for solving the difficult problem of identifying structure in high-dimensional data. Although for many datasets these dimension reduction methods effectively work to reveal groups, it has been previously highlighted that they are not specifically designed for preserving the clusters and neither the directions of maximum variation of data nor the departure from normality, what may ensure that the reduced space keeps the original structure of groups unaltered [7]. This is one of the main reasons for proposing HUEPs as an advanced visualization technique that provides with EPP visualizations while at the same time keeps information about clustering in the original dataset.

As previously stated, cluster analysis can be defined as the process of organizing data into groups that in some way have similar (or close) members. Data similarity or proximity is measured by a distance function defined on pairs of patterns. Up to now, many different distance measures have been used [52, 53].

On the other hand, all the different approaches to data clustering [2] are classified in two main types of methods: partitional or hierarchical. On the one hand, partitional methods are based on the idea of identifying the partition that optimizes (usually locally) a given clustering criterion. On the other hand, hierarchical methods generate a set of nested partitions that are iteratively merged according to a certain criterion. In present paper, one partitional and one hierarchical method have been applied and are described in following subsections.

**2.2.1. *K-Means.*** *K*-means [54] is a well-known partitional clustering method aimed at grouping data into a given number of clusters. In order to apply it, two parameters must be tuned: the given number of clusters ( $k$ ) and the initial position of centroids. The latter can be chosen by the user or calculated in a preprocessing step. Once initial values are assigned to these parameters, each data in the dataset is assigned to the nearest cluster centroid, attaining the initial allocation of data in clusters. Then, the centroids are iteratively recalculated and a subsequent reallocation of data is made. This step is repeated until no further changes are made to the centroids, when the cluster assigned to each data is generated as the output.

This method heavily relies on its initial parameters; hence, a usual measure of the “goodness” of the grouping is the sum of the proximity Sums of Squared Error (SSE) that it attempts to minimize:

$$SSE = \sum_{j=1}^k \sum_{x \in G_j} \frac{p(x_i, c_j)}{n} \quad (10)$$

where  $c_j$  are the cluster centroids,  $p()$  is the proximity function,  $n$  is the number of rows, and  $k$  is the number of groups.

Similarity or proximity is a key concept for the definition of a cluster. As a result, a measure of the similarity must be carefully chosen as it is crucial to most clustering methods. Among all the available measures of similarity for data whose features are all continuous, some of the most widely used ones are as follows:

- (i) Squared Euclidean distance (sqEuclidean). Each centroid is calculated as the mean of the points in that cluster.
- (ii) Cityblock: sum of absolute differences. Each centroid is calculated as the component-wise median of the points in that cluster.
- (iii) Cosine: one minus the cosine of the included angle between points (treated as vectors). Each centroid is calculated as the mean of the points in that cluster, after normalizing those points to unit Euclidean length.
- (iv) Correlation: one minus the sample correlation between points (treated as sequences of values). Each centroid is calculated as the component-wise mean of the points in that cluster. Previously, those points are centred and normalized to zero mean and unit standard deviation.

As a result of the clustering, a scalar is provided for each input vector, being the number of the cluster to which the vector has been assigned.

**2.2.2. *Hierarchical Methods.*** Differentiating from partitional clustering methods, hierarchical ones can be divided into two types:

- (1) Agglomerative: they begin with each data in a different cluster, and clusters are successively merged together until a stopping criterion is met or until a single cluster is obtained.
- (2) Divisive: they begin with all data assigned to the only cluster, that is split (and its descendants) until a stopping criterion is satisfied or every data is assigned to a different cluster.

In the present study, due to the successful results in initial experiments, agglomerative clustering has been selected in order to be compared to the partitional approach (*k*-means). In the case of agglomerative clustering, there is a variety of linking methods that can be applied. In present study, the following ones have been tested:

- (i) Single: shortest distance.
- (ii) Complete: furthest distance.
- (iii) Ward: inner squared distance (minimum variance algorithm), appropriate for Euclidean distances only.
- (iv) Median: weighted centre of mass distance (WPGMC: Weighted Pair Group Method with Centroid Averaging), appropriate for Euclidean distances only.

- (v) Average: unweighted average distance (UPGMA: Unweighted Pair Group Method with Arithmetic Averaging).
- (vi) Centroid: centroid distance (UPGMC: Unweighted Pair Group Method with Centroid Averaging), appropriate for Euclidean distances only.
- (vii) Weighted: weighted average distance (WPGMA: Weighted Pair Group Method with Arithmetic Averaging).

### 3. Case Study: Internationalization of Spanish SMEs

In order to validate the proposed HUEPs, they are applied to an interesting problem that has not yet been addressed by means of EPP or clustering methods. Hence, HUEPs are generated to analyse the internationalization strategy of companies, what involves a high number of features.

The dataset analysed in the present study is based on a sample of all Spanish MNEs registered with the Foreign Trade Institute (ICEX) and from the Web site <http://www.oficinascomerciales.es>, both managed by the Spanish Ministry of Industry, Tourism, and Trade. In order to analyse a representative sample of companies with sufficient autonomy, we restricted the sample to keep only those large and independent enough to conduct and decide their own internationalization strategy. Thus, following a well-established cutoff point in international business literature, used for example by Eurostat ([http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Enterprise\\_size](http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Enterprise_size)), we dropped from the sample those with less than 250 employees. We also dropped those companies with a foreign majority owner controlling more than half of the capital.

It is also important to note the huge impact of the financial crisis on the Spanish economy, which forced many multinational enterprises to sell or postpone international operations in order to focus on the problems of the home market. To avoid distortions in the results due to this exogenous effect, we took the year 2007 as our base year. Overall, the sample consists of 164 companies investing in 119 countries worldwide. Unfortunately, Afghanistan, Andorra, Puerto Rico, and São Tomé and Príncipe are not included in the sample due to lack of data. In addition, Serbia, Montenegro, and Kosovo are included as a group because at the time of the study they constituted a single country.

For the above-mentioned companies and countries, the following data about each one of the cases of international presence were collected (further details about the different features can be found in [32]):

- (i) Company sector: 5 binary features stating the economy sector the company belongs to (manufacturing, food, construction, regulated, and others).
- (ii) Company product diversification: 3 binary features (nondiversified, related or unrelated diversification).
- (iii) Other company characteristics: assets, number of employees, return on assets (ROA), ROA growth, age, number of countries where the company operates,

and leverage and whether or not the company is included in a stock market.

- (iv) Host country characteristics: GDP, GDP growth, total inward Foreign Direct Investment, population, unemployment, level of corruption, and Economic Freedom Index.
- (v) Geographic and psychic distance stimuli between home and host countries: the data for each psychic distance stimulus is calculated by Dow & Karunaratna [28] based on a principal component analysis of a single factor. The calculations are based on critical factors widely used in the literature to explain cross-national differences at the macro level. Thus, the education distance stimulus is based on differences on literacy rate and enrolment in second and third-level education building on data from the United Nations. The industrial development stimulus takes into account differences in ten dimensions such as in energy consumption, vehicle ownership, employment in agriculture, and number of telephones and televisions. The language stimulus is based on the differences between the dominant languages and the bilateral influence of each country's major language in the other country. The democracy stimulus includes differences in the type of political systems in terms of political rights, civil liberties and POLCON and POLITY IV indices which account for the political constraints of the government of the country based on the existence and alignment of other independent political agents who can keep reducing the government discretionary power. The political ideology stimulus is based on the ideological leanings of the chief executive's political party and the largest political party in the government. Finally, the religion stimulus is calculated based on the differences between the dominant religions and the bilateral influence of each country's dominant religion in the other country.

As a result, a dataset containing 1456 samples and 33 features was obtained and is analysed by means of HUEPs as it is presented in the following section.

### 4. Results and Discussion

Data from the aforementioned real-world case study are shown on low-dimensional spaces, on which they can be visually compared. In this section, the main results (HUEPs) are presented; for comparison purposes, combinations of the three EPP methods (PCA, MLHL, and CMLHL) with two clustering methods (hierarchical clustering and  $k$ -means) are shown. Additionally, Psychic-Democracy information is added through the glyph metaphor. As it is a continuous variable ranging from 0 to 2, it has been discretized in quartiles and data are shown accordingly (see the legend in Figure 2).

Combinations of different values were tested during experimentation for each one of the parameters of the applied models. After that, the best results were selected and are

- + Q1 (1.5 - 1.99)
- ▷ Q2 (1 - 1.49)
- Q3 (0.5 - 0.99)
- \* Q4 (0 - 0.49)

FIGURE 2: Legend for the glyph metaphor when using Psychic-democracy distance.

presented in Section 4 for the sake of brevity. In order to obtain such results, the different parameters were tuned with the following values:

- (i) PCA: number of output dimensions: 2 and 3.
- (ii) MLHL: number of output dimensions: 2 and 3, number of iterations: 3000, learning rate: 0.08009,  $p$ : 0.54.
- (iii) CMLHL: number of output dimensions: 2 and 3, number of iterations: 3000, learning rate: 0.000175,  $p$ : 1.96,  $\tau$ : 0.034.
- (iv)  $k$ -means:  $k$ -means++ algorithm for cluster centre initialization, squared Euclidean distance and values of  $k$  equal to 3 and 6.
- (v) Agglomerative clustering: cosine distance, single linkage method, and a cutoff value adjusted to obtain the same number of clusters as in the case of  $k$ -means (3 and 6).

**4.1. HUEP: EPP + Partitional Clustering.** Firstly, HUEPs generated by the combination of the three EPP methods together with partitional clustering ( $k$ -means) are presented and their most relevant characteristics are discussed.

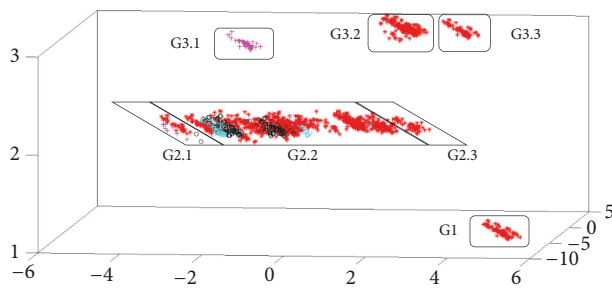
From a general point of view, visualizations in Figure 3 reveal a certain structure in the analysed dataset. The results from Figure 3(a) clearly depict groups at three different levels of the vertical axis (that is, the output of the  $k$ -means clustering method when the  $k$  parameter equals to 3). The first one (labelled as G1) is made of subsidiaries located in the United States. The second one includes three subgroups made of countries sharing specific characteristics. Subgroup G2.1 includes subsidiaries located in countries with economic and political problems such as Venezuela and Bangladesh. Subgroup G2.2 includes subsidiaries in emerging and growing economies, with a more stable environment compared to G2.1, such as Argentina, Brazil, Chile, Colombia, Hungary, Morocco, Mexico, Russia, Thailand, Turkey, Poland, Philippines, Slovakia, and Slovenia. This subgroup also includes some European countries with relatively advanced economies such as Belgium, Ireland, and Portugal. Finally, subgroup G2.3 includes small European countries with advanced economies and stable democracies such as the Netherlands and Norway. The third level includes also three subgroups with particular characteristics. The subgroup G3.1 includes subsidiaries located in China. The subgroup G3.2 includes Japan and some of the largest Western economies such as France, Italy, and Germany. Finally, subgroup G3.3 includes another developed European economy, the UK. Overall, Figure 3(a) offers a very clear determination of a cluster of a country with a very low level of democracy

(China), at the extreme left side of the visualization, compared to democratic societies which appear on the right side. However, the HUEP also clearly distinguishes between advanced societies with a similar pluralistic political system to Spain, such as other geographically closer Western Europe economies of a similar size (France, Italy, and Germany), as opposed to another democratic country but with a different political organization based more on a bipartisan system. Smaller economies are located in the second, intermediate level of the vertical axis, but clearly differentiated according to their level of economic and political development, with those less developed economies at the extreme left side of the graph, stable and growing emerging countries in the middle and more advanced countries at the extreme right side.

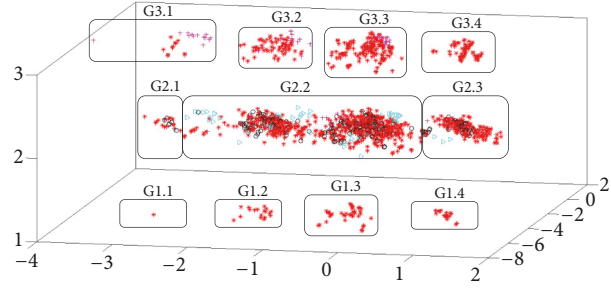
The results from Figure 3(b) exhibit a very similar pattern to those of Figure 3(a). Subgroups G1.1, G1.2, G1.3, and G1.4 gather all of the subsidiaries located in the United States, similar to the group G1 in Figure 3(a). The second level is again a mixed combination of emerging economies from all over the world together with democratic societies of a smaller size of Spain, very similar to what happened in Figure 3(a). In this case, however, it is worth noting that the subgroups of this level are much more heterogeneous and it is not easy to differentiate them according to their level of development as in the previous visualization. Both emerging and advanced countries appear in all subgroups. Finally, the main difference between both figures is that Group 3 is also less clear in Figure 3(b) than in Figure 3(a). While in Figure 3(a) China was clearly identified as an independent subgroup and all the subsidiaries located in this country were included in a single subgroup, in this case they appear simultaneously in subgroups G3.1, G3.2, and G3.3. Besides, subsidiaries, located in the common law based in UK, do not appear in a slightly separated subgroup, but mixed in all G3 subgroups.

As the best results are obtained by CMLHL, HUEP generated by this EPP method together with  $k$ -means is individually shown in Figure 4.

The results from Figure 4 are consistent with the previous ones but offer an insightful nuance. First, G1 is consistent with Figures 3(a) and 3(b) and includes all subsidiaries located in the United States. Next, G2 can be split into two subgroups, the first one (G2.1) is made of subsidiaries in Serbia, Montenegro, and Kosovo. While these three countries used to be a single one, Montenegro held an independence referendum in 2006 and Kosovo declared its unilateral independence in 2008. This particular method, unlike the previous ones, shows the ability to distinguish these historic events taking place around the time the sample was collected. The second one, G2.2, is made of the same mix of emerging economies and democratic advanced countries smaller in size than Spain. Finally, G3 is split into three subgroups. The first one (G3.1) in the left extreme of the graph shows subsidiaries located in China. The second one (G3.2) includes Western European countries close to Spain in terms of geography, size, and democratic systems (France, Italy, and Germany), and the third one (G3.3) includes the subsidiaries located in the UK. According to the Psychic-Democracy information that is also depicted, HUEP generated from CMLHL projection



(a) HUEP: PCA + partitional clustering + Psychic - Democracy information



(b) HUEP: MLHL + partitional clustering + Psychic - Democracy information

FIGURE 3: HUEP samples on partitional clustering ( $k=3$ ).

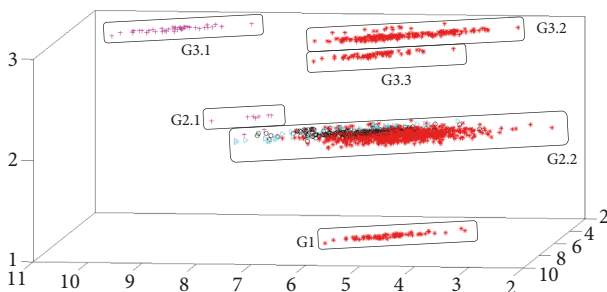


FIGURE 4: HUEP: CMLHL + partitional clustering ( $k=3$ ) + Psychic - Democracy information.

reveals the structure of the dataset in a better way; in Figure 4 subgroups 2.1 and 3.1 only gather countries with highest scores (Q1 - pink crosses). Data with intermediate values of Psychic-Democracy information are all gathered in subgroup 2.2 but in an ordered way, starting from Q2 (blue triangles) close to the data with highest values (Q1) on the left side of the visualization. Then, data from Q3 are visualized (black circles) and data with lowest values (red stars) can be seen on the right side.

To check the effect of increasing the target number of clusters to be identified ( $k$  parameter) by  $k$ -means, some other experiments were run. The results for the three EPP methods when the  $k$  parameter equals to 6 are shown below in Figures 5(a), 5(b), and 6.

In general terms, it can be said that the results from Figure 5(a) are comparable to those from Figure 3(a). At the lower level, on the right extreme of the graph, the visualization identifies two separate large economies. G1 includes subsidiaries in the US and G2 those in Germany. G3, that is more on the left side of the graph, includes subsidiaries in various emerging economies such as Argentina, Chile, Morocco, Poland, and Turkey. G4 is similar to G3 in Figure 2(a). G4.1 identifies subsidiaries located in China. G4.2 includes subsidiaries in France and Italy and G4.3 those in UK. G5.1 includes subsidiaries in three large emerging countries: Brazil, Mexico, and Russia. G5.2 identifies one country in particular, South Korea. G5.3 includes subsidiaries in advanced economies such as Australia, Canada, and the

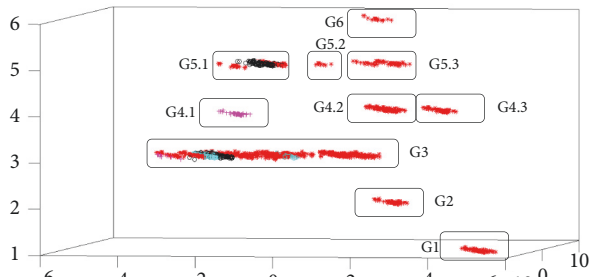
Netherlands. Finally, at the top of the graph, G6 identifies subsidiaries in Japan.

Overall, this HUEP in Figure 5(a) shows consistent results with those from Figure 3(a) as it displays countries according to their level of economic and political development from left to right, and it also identifies specific countries that are relevant and with a particular idiosyncrasy. As in the case of Figure 3(a), US, China, and UK are highlighted given their differences with the continental European system of Spain. However, in this case, Germany, South Korea, and Japan are also identified in specific subgroups. The former may be due to its federal political organization in which the constituent states (*Länder*) retain a measure of sovereignty. The latter two are two stable and advanced countries with a well-functioning democratic, parliament-based, political system. However, their large cultural distance from Spain and political tensions with China and North Korea may explain why this visualization separates them from the rest.

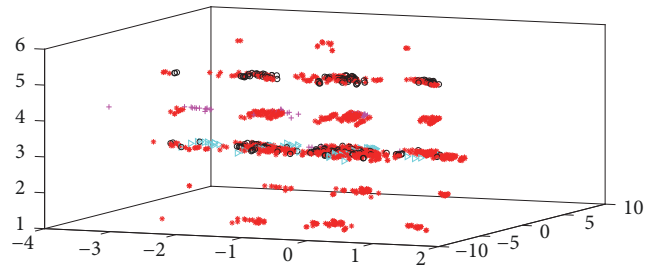
No clear structure is revealed in Figure 5(b): many different subgroups are generated with a heterogeneous mixture of countries. As a result, and for the sake of brevity, results in this figure are not described.

The results from Figure 6 show a very similar pattern to those from Figure 5(a). G1 includes subsidiaries in the US and G2 in Germany. G3.2 includes emerging economies similar to the previous G3 subgroup in Figure 5(a). However, in this case, the visualization separates Serbia in G3.1, due to the previously mentioned events happening in Montenegro and Kosovo. Also similar to G4 in Figure 5, here the subgroup G4.1 located in the left extreme of the graph includes subsidiaries in China, whereas G4.2 includes those in France and Italy and G4.3 those in the UK. Finally, this visualization identifies Japan in G6, but contrary to the previous visualization, South Korea is included in a larger and more heterogeneous group with other countries in G5. This group includes all the countries that in Figure 5(a) were part of subgroups G5.1 (Brazil, Mexico, and Russia) and G5.3 (Australia, Canada, and the Netherlands). Overall, while this visualization offers the advantage of identifying Serbia separately as Figure 4, it shows a less clear picture in G5 compared to Figure 5(a).





(a) HUEP: PCA + partitional clustering + Psychic - Democracy information



(b) HUEP: MLHL + partitional clustering + Psychic - Democracy information

FIGURE 5: HUEP samples on partitional clustering ( $k=6$ ).

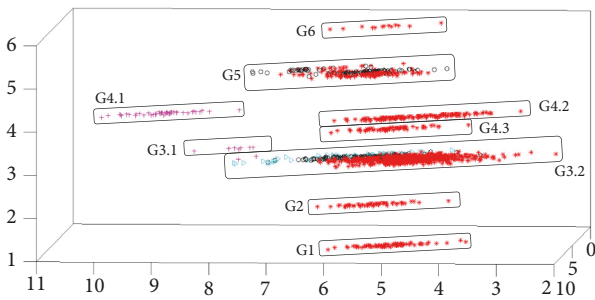


FIGURE 6: HUEP: CMLHL + partitional clustering ( $k=6$ ) + Psychic - Democracy information.

When analysing Psychic-Democracy information depicted in Figure 6, it can be said that once again, groups are coherently organized according to such criteria. Only 2 subgroups (out of 9) contain data from more than one quartile. Furthermore, there is a global and decreasing ordering from left (data in Q1, depicted as pink crosses) to right (data in Q4, depicted as red stars).

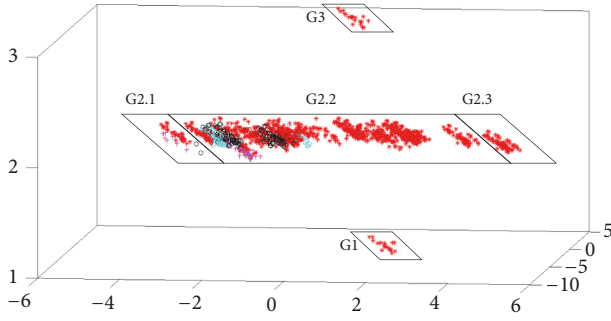
**4.2. HUEP: EPP + Hierarchical Clustering.** In order to check the validity of proposed HUEPs to combine results from different clustering methods, results from hierarchical clustering (combined with the 3 different EPP methods, namely, PCA, MLHL, and CMLHL) are shown in this subsection.

Figure 7(a) shows some interesting differences compared to previous visualizations. While also organized in three levels (3 output clusters) as Figures 3(a) and 3(b), the countries uniquely identified in separate groups are different. In the lower level, G1 identifies Australia and in the upper level G3 identifies Ireland. In the middle level, three subgroups are identified. In this case, consistent with Figures 3(a) and 3(b), countries on the left show lower levels of economic and political development. Thus, G2.1 includes Venezuela and Bangladesh, while G3 identifies the US. G2.2 is a very heterogeneous group including all other countries in the world. In this case, the HUEP underlines the particular situation of Ireland, a location where the laws of the country offer very favourable conditions given the low corporate tax,

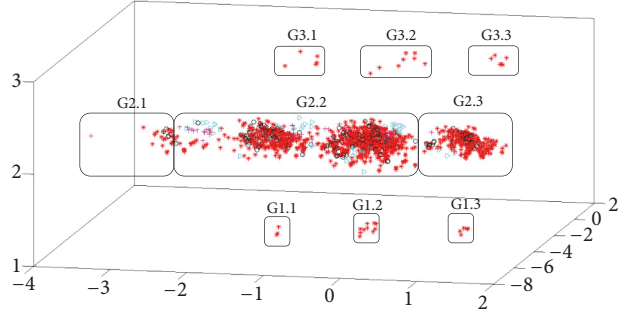
noticeably lower than in the rest of Europe. As a result, many MNEs have located their subsidiaries, often leading to controversial debates and loss of legitimacy. For example, Zara's owner Inditex has been accused of tax evasion (<https://www.independent.ie/business/irish/zara-owner-used-ireland-to-slash-its-tax-bill-meps-claim-35279873.html>), reporting millions of euros in turnover but having no employees on the payroll (<https://fashionunited.uk/news/business/inditex-accused-of-dodging-585-million-euros-in-taxes/2016120822765>). The case of Australia might be due to the fact that it is a country perceived as distant both in terms of geography and culture, which represents an obstacle to FDI, and pertaining to the Commonwealth and therefore based on a common law system with relevant similarities with the UK.

Figure 7(b) is also structured in three levels and shows very consistent results with the previous one. However, in this case, G1 and G3 are split into three subgroups. Similar to Figure 7(a), G1 includes subsidiaries in Australia and G3 includes subsidiaries in Ireland. However, in this visualization it is possible to observe differences based on the specific sector of the firms. Subgroups on the left (G1.1 and G3.1) include companies in the infrastructure sector such as ACS, Ferrovial or Indra, and other highly regulated sectors such as airlines (Iberia). Subgroups in the middle (G1.2 and G3.2) include large companies in manufacturing such as Inditex and Mango. Finally, subgroups on the right side of the graph include smaller (albeit also MNEs) companies such as Teka, Tamisa, or Valdepeña. While this visualization is more precise about the sectors of these two particular countries, the subgroups in the middle level of the vertical axis (G2.1, G2.2, and G2.3) are heterogeneous and it is not easy to identify groups based on their level of economic or political development compared to G2 in Figure 7(a).

The results of Figure 8 are quite similar to those in Figure 7(a). G1 includes subsidiaries in Australia and G3 includes those in Ireland. However, G2.1 includes two countries that have been repeatedly identified in independent subgroups in previous visualization (China and Serbia), although in this case they form a group together due to the perspective. Finally, in G2.3 the UK is identified as a slightly separate group compared to the larger G2.2 which includes the majority of countries of the world. It is worth highlighting that the ordering (from left to right) of Psychic-Democracy



(a) HUEP: PCA + hierarchical clustering + Psychic - Democracy information



(b) HUEP: MLHL + hierarchical clustering + Psychic - Democracy information

FIGURE 7: HUEP samples on hierarchical clustering.

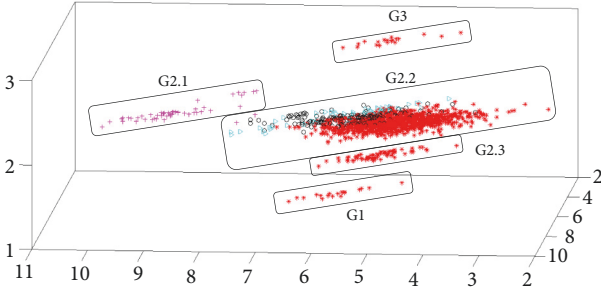


FIGURE 8: HUEP: CMLHL + hierarchical clustering + Psychic - Democracy information.

information is preserved in Figure 8, with a more precise definition than in Figure 4.

From previous Sections 4.1 and 4.2 it can be concluded that HUEPs successfully combine the output from different EPP (PCA, MLHL, and CMLHL) and clustering (partitional and hierarchical) methods.

**4.3. Comparison to Alternative Visualizations.** Up to the authors' knowledge, there is not any validation method to test HUEPs with quantitative metrics. As a consequence, the obtained results are visually compared with some other visualizations of the same dataset. For a fair comparison, 3D scatterplots have also been generated.

Initially, HUEPs are compared to a combination of EPP together with partitional clustering, without using the glyph metaphor with any additional information as it has been used in Figures 3–8 (Psychic-Democracy).

In Figure 9 the same structure that has already been described in the case of Figure 4 is revealed. The same data are located in the same groups, but obviously adding further details through the glyph metaphor makes HUEPs more informative. Thanks to the different colours and shapes, it is easy to get an idea of the global ordering that has been previously mentioned (for Figures 4 and 8) and to know which countries are located in some of the groups.

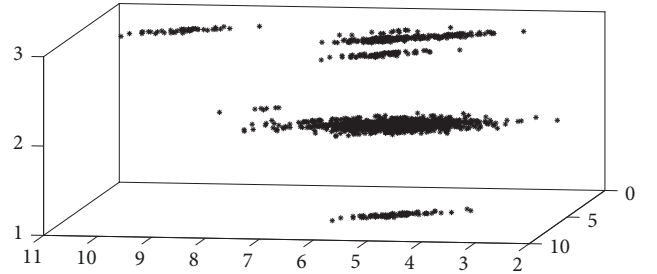


FIGURE 9: CMLHL + partitional clustering.

**4.3.1. Alternative Distance Information.** In order to check the results of the proposed HUEPs when visualizing some other distance criteria (different from Psychic-Democracy) through the glyph metaphor, Figures 10–12 are provided. As these also are continuous variables, they have been discretized in quartiles, as in the case of Figures 3–8.

Yet, in order to provide a comprehensive analysis, we also conduct the analysis of all psychic dimensions altogether. To do so we rely on the operationalization suggested by [26] as this method has been proven superior to the simple average of dimensions since it also takes into account the differences in variance of the dimensions. Algebraically, this method can be expressed as

$$KS_j = \frac{\sum_{i=1}^6 \left( (I_{ij} - I_{iu})^2 / V_i \right)}{6} \quad (11)$$

where  $I_{ij}$  is country  $j$ 's score on the  $i$ th cultural dimension,  $I_{iu}$  is the score for Spain on this dimension, and  $V_i$  is the variance of the score on the dimension.

While in the previous visualization we focused on the democracy distance, we also controlled the visualizations of other psychic distance stimuli and also that of the overall psychic distance aggregated into a single construct using the Kogut & Singh's formula. For the sake of parsimony, we show here only those with a clearer visualization of the different groups, in particular the one using education distance, the one using Religion distance, and the one using the aggregated

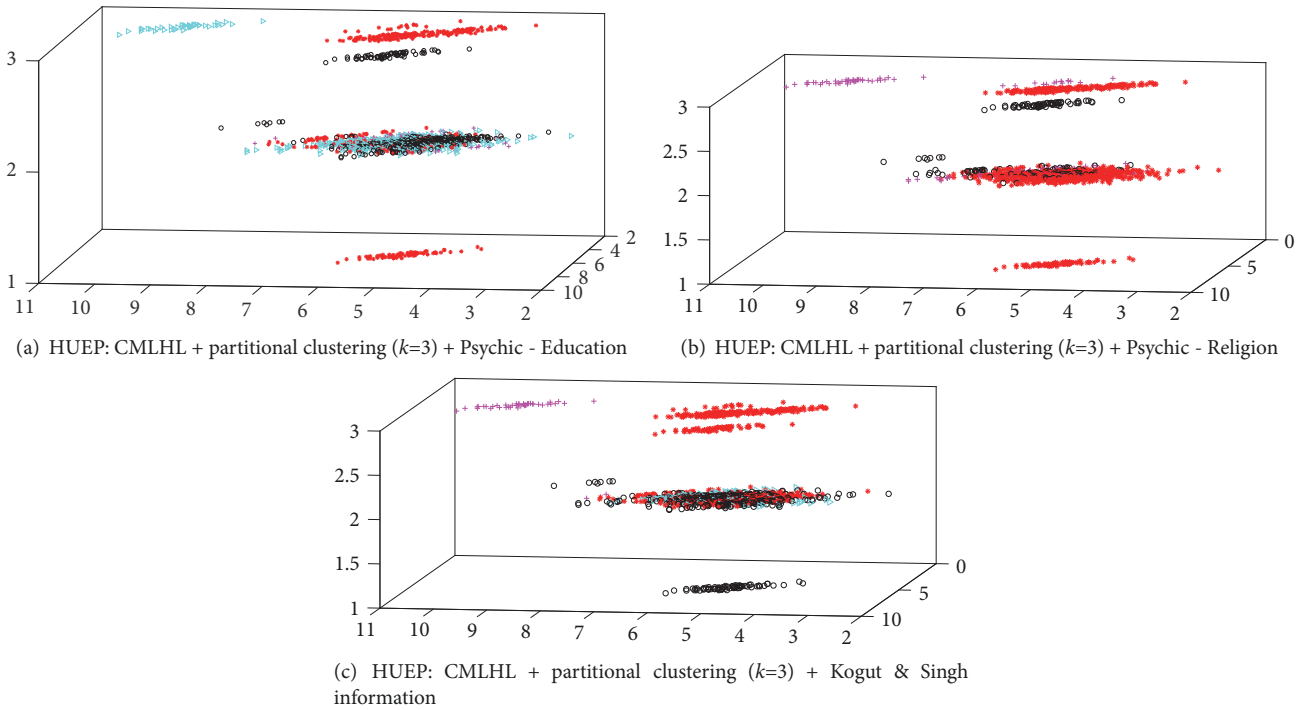


FIGURE 10: Comparison of HUEPs when visualizing different distance criteria through the glyph metaphor.

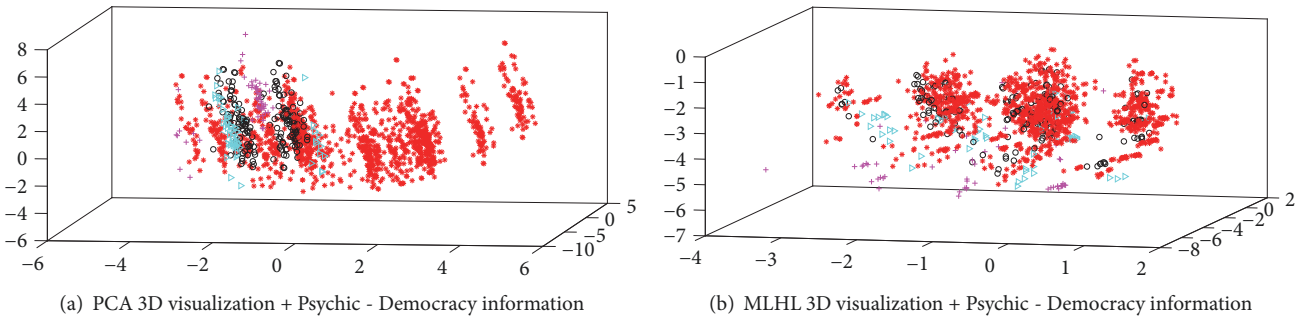


FIGURE 11: 3D plots from EPP methods.

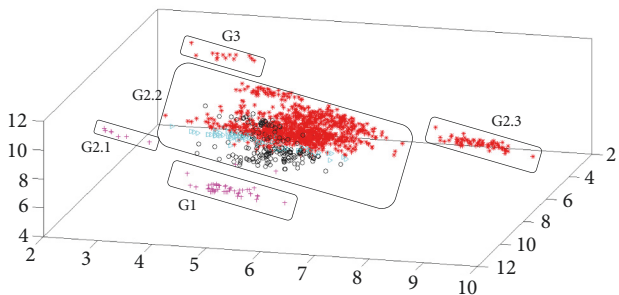


FIGURE 12: CMLHL 3D visualization + Psychic - Democracy information.

Kogut & Singh’s formula. However, as it can be seen in Figures 10(a), 10(b), and 10(c), the visualization is less clear as it is observed in the more fuzzy combination of colours in some

of the groups. Although different criteria can be visualized by the HUEPs, not all of them are equally informative for a given projection. In the case of the criteria visualized in Figure 10 it can be seen that some of the data from the same quartile are gathered in the same groups but some others are not. Furthermore, a global ordering is not revealed as in the case of Psychic-Democracy (Figures 4 and 8).

**4.3.2. 3D EPP Projections.** Finally, the comprehensive comparison of visualizations also comprises simpler 3D plots where only the output (3 first components) of the EPP methods is depicted, together with the glyph metaphor.

When compared with previous HUEPs that combine the outputs of corresponding EPP methods (PCA or MLHL), Figure 11 does not reveal the structure of the whole dataset in a sparse and clear way, although some subgroups could be identified.

In the case of Figure 12, the 3D CMLHL visualization reveals more clearly defined groups than those from Figure 11. In this case, the majority of countries are included in a very heterogeneous group in G2.2. However, the method identifies the subsidiaries located in China in group G1, the subsidiaries located in Serbia in the subgroup G2.1, and the subsidiaries located in the UK in the subgroup G2.2, countries showing some specific features as already described. Finally, G3 includes the subsidiaries in South Africa, a country that was never represented in its own separate group. As in the case of Australia, that was singled out in some previous visualizations, this is a country that is both geographically and culturally distant to Spain and with a political system based on the UK's common law system, as a former colony and part of the Commonwealth. While this visualization identifies specific countries such as China, Serbia, and South Africa, the very heterogeneous nature of countries included in G1 makes the visualization less clear than previous ones such as those of Figures 4 and 8. When compared to the corresponding HUEPs, it can be said that adding the clustering information makes the visualization more precise, as data are split in a larger number of more separated groups, which let us gain deeper knowledge of the case of study.

## 5. Conclusions and Future Work

From the results presented in Section 4, it can be concluded that HUEPs are a useful technique to visually analyse internationalization data in order to better understand it. More specifically, the presented visualizations provide insightful information about the geographical distribution of Spanish subsidiaries. They also allow for the identification of specific countries exhibiting specific political and legal characteristics or going through particular historic events (e.g., China, the UK, US, Serbia, etc.). This type of data represents a valuable source of information for managers in enterprises who can learn from vicarious experience (i.e., the knowledge that companies can obtain from the actions of other firms sharing a common characteristic, such as nationality) [32]. By observing the behavior of other companies, firms can imitate best practices and avoid previous mistakes [55]. Besides, the data is also very relevant for policy-makers interested in attracting larger volumes of foreign investors, as these investments can provide key technology or managerial talent missing in the country and also positive spillovers in the form of a boost for the competitiveness of other related industries in the economy [56].

When considering the different EPP methods that have been applied, it can be said that CMLHL provides the more sparse projections, what is consistent with previous work. On the other hand, both clustering methods generate meaningful outputs and it is worth mentioning that HUEPs greatly accommodate to a varying number of clusters (higher than 1). According to the glyph metaphor comparison, adding Democracy (Psychic) distance let us better understand the nature of the analysed dataset by HUEPs. Thanks to the more precise definition and higher number of groups in the visualizations, HUEPs contribute to overcome some of the

drawbacks of scatter plots: overplotting and overlapping. All in all, it has been proven that HUEPs are a valid proposal to combine the outputs from different EPP and clustering methods. Additionally, the 3D scatterplots can be enriched with information from different sources (distance criteria in the present case study).

In future work, HUEPs will be applied to some other multidimensional datasets, comprising companies from other countries apart from Spain and thus comparing the internationalization strategies of companies from different countries.

## Data Availability

Previously reported ICEX data were used to support this study and are available at <http://www.oficinascomerciales.es>. These prior datasets are cited at relevant places within the text.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The work was conducted during Álvaro Herrero's research stay at KEDGE Business School in Bordeaux (France). Some results of this ongoing research, from the same dataset, have been presented in the 13th International Conference on Soft Computing Models in Industrial and Environmental Applications, as a paper entitled "Visualizing Industrial Development Distance to Better Understand Internationalization of Spanish Companies".

## References

- [1] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Transactions on Computers*, vol. 23, pp. 881–890, 1974.
- [2] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.
- [3] W. Desarbo, K. Jedidi, K. Cool, and D. Schendel, "Simultaneous multidimensional unfolding and cluster analysis: An investigation of strategic groups," *Marketing Letters*, vol. 2, no. 2, pp. 129–146, 1991.
- [4] M. Vichi and H. A. L. Kiers, "Factorial k-means analysis for two-way data," *Computational Statistics & Data Analysis*, vol. 37, no. 1, pp. 49–64, 2001.
- [5] I. T. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 2002.
- [6] T. Tarpey, "Linear transformations and the  $k$ -means clustering algorithm: applications to clustering curves," *The American Statistician*, vol. 61, no. 1, pp. 34–40, 2007.
- [7] G. Menardi and N. Torelli, "Reducing data dimension for cluster detection," *Journal of Statistical Computation and Simulation*, vol. 83, no. 11, pp. 2047–2063, 2013.
- [8] T. Liu, C. K. L. Lekamalage, G. Huang, and Z. Lin, "Extreme learning machine for joint embedding and clustering," *Neurocomputing*, vol. 277, pp. 78–88, 2018.



- [9] S. Jun, S.-S. Park, and D.-S. Jang, "Document clustering method using dimension reduction and support vector clustering to overcome sparseness," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3204–3212, 2014.
- [10] K. K. Bharti and P. K. Singh, "Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3105–3114, 2015.
- [11] J. Guo, X. Zhao, X. Yuan, Y. Li, and Y. Peng, "Discriminative unsupervised 2D dimensionality reduction with graph embedding," *Multimedia Tools and Applications*, vol. 77, no. 3, pp. 3189–3207, 2018.
- [12] C.-L. Liu, W.-H. Hsaio, and T.-H. Chang, "Locality sensitive K-means clustering," *Journal of Information Science and Engineering*, vol. 34, no. 1, pp. 289–305, 2018.
- [13] R. Sánchez, Á. Herrero, and E. Corchado, "Visualization and clustering for SNMP intrusion detection," *Cybernetics and Systems*, vol. 44, no. 6-7, pp. 505–532, 2013.
- [14] Á. Arroyo and Á. Arroyo, "A hybrid intelligent system for the analysis of atmospheric pollution: a case study in two european regions," *Logic Journal of the IGPL*, vol. 25, no. 6, pp. 915–937, 2017.
- [15] R. J. Bolton and W. J. Krzanowski, "Projection pursuit clustering for exploratory data analysis," *Journal of Computational and Graphical Statistics*, vol. 12, no. 1, pp. 121–142, 2003.
- [16] Y. G. Yatracos, "Detecting clusters in the data from variance decompositions of its projections," *Journal of Classification*, vol. 30, no. 1, pp. 30–55, 2013.
- [17] R. Redondo, "Neural visualization for the analysis of energy and water consumptions in the automotive industry," in *Proceedings of the International Joint Conference SOCO18-CISIS18-ICEUTE18*, Springer International Publishing, Cham, Switzerland, 2019.
- [18] S. L. Marie-Sainte, "Detection and visualization of non-linear structures in large datasets using exploratory projection pursuit laboratory (EPP-Lab) software," *Journal of King Saud University - Computer and Information Sciences*, vol. 29, no. 1, pp. 2–18, 2017.
- [19] M. Nilashi, O. Ibrahim, and K. Bagherifard, "A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques," *Expert Systems with Applications*, vol. 92, pp. 507–520, 2018.
- [20] S. Zaheer, M. S. Schomaker, and L. Nachum, "Distance without direction: Restoring credibility to a much-loved construct," *Journal of International Business Studies*, vol. 43, no. 1, pp. 18–27, 2012.
- [21] A. Jiménez, D. Benito-Osorio, J. Puck, and P. Klopff, "The multifaceted role of experience dealing with policy risk: The impact of intensity and diversity of experiences," *International Business Review*, vol. 27, no. 1, pp. 102–112, 2018.
- [22] P. Ghemawat, "Distance still matters. The hard reality of global expansion.," *Harvard Business Review*, vol. 79, no. 8, pp. 137–142, 2001.
- [23] H. Berry, M. F. Guillén, and N. Zhou, "An institutional approach to cross-national distance," *Journal of International Business Studies*, vol. 41, no. 9, pp. 1460–1480, 2010.
- [24] S. L. Merker, *Culture's Consequences: International Differences in Work-Related Values*, vol. 27, Sage Publications, Beverly Hills, Ca, USA, 1980.
- [25] G. Hofstede, G. J. Hofstede, and M. Minkov, *Cultures and Organizations: Software of the Mind*, McGraw-Hill, New York, NY, USA, 2010.
- [26] B. Kogut and H. Singh, "The effect of national culture on the choice of entry mode," *Journal of International Business Studies*, vol. 19, no. 3, pp. 411–432, 1988.
- [27] R. L. Tung and A. Verbeke, "Beyond hofstede and GLOBE: improving the quality of cross-cultural research," *Journal of International Business Studies*, vol. 41, no. 8, pp. 1259–1274, 2010.
- [28] D. Dow and A. Karunaratna, "Developing a multidimensional instrument to measure psychic distance stimuli," *Journal of International Business Studies*, vol. 37, no. 5, pp. 578–602, 2006.
- [29] L. Håkanson and B. Ambos, "The antecedents of psychic distance," *Journal of International Management*, vol. 16, no. 3, pp. 195–210, 2010.
- [30] P. A. Brewer, "Operationalizing psychic distance: A revised approach," *Journal of International Marketing*, vol. 15, no. 1, pp. 44–66, 2007.
- [31] J. Johanson and J.-E. Vahlne, "The internationalization process of the firm: a model of knowledge development and increasing foreign market commitments," *Journal of International Business Studies*, vol. 8, no. 1, pp. 23–32, 1977.
- [32] A. Jiménez and D. de la Fuente, "Learning from others: the impact of vicarious experience on the psychic distance and FDI relationship," *Management International Review*, vol. 56, no. 5, pp. 633–664, 2016.
- [33] O. Shenkar, "Cultural distance revisited: towards a more rigorous conceptualization and measurement of cultural differences," *Journal of International Business Studies*, vol. 32, no. 3, pp. 519–535, 2001.
- [34] B. L. Kirkman, K. B. Lowe, and C. B. Gibson, "A quarter century of culture's consequences: A review of empirical research incorporating Hofstede's cultural values framework," *Journal of International Business Studies*, vol. 37, no. 3, pp. 285–320, 2006.
- [35] A. Delios and W. J. Henisz, "Political hazards, experience, and sequential entry strategies: The international expansion of Japanese firms, 1980-1998," *Strategic Management Journal*, vol. 24, no. 11, pp. 1153–1164, 2003.
- [36] G. L. Holburn and B. A. Zelner, "Political capabilities, policy risk, and international investment strategy: Evidence from the global electric power generation industry," *Strategic Management Journal*, vol. 31, no. 12, pp. 1290–1315, 2010.
- [37] A. Jiménez, "Does political risk affect the scope of the expansion abroad? Evidence from Spanish MNEs," *International Business Review*, vol. 19, no. 6, pp. 619–633, 2010.
- [38] T. Lawton, T. Rajwani, and J. Doh, "The antecedents of political capabilities: A study of ownership, cross-border activity and organization at legacy airlines in a deregulatory context," *International Business Review*, vol. 22, no. 1, pp. 228–242, 2013.
- [39] A. Jiménez, I. Luis-Rico, and D. Benito-Osorio, "The influence of political risk on the scope of internationalization of regulated companies: insights from a Spanish sample," *Journal of World Business*, vol. 49, no. 3, pp. 301–311, 2014.
- [40] G. Conti, *Security Data Visualization: Graphical Techniques for Network Analysis*, No Starch Press, 2007.
- [41] L. Wilkinson, *The Grammar of Graphics. Statistics and Computing*, Springer, New York, 2nd edition, 2005.
- [42] D. A. Keim, "Information visualization and visual data mining," *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 1, pp. 1–8, 2002.
- [43] N. Elmqvist, P. Dragicevic, and J.-D. Fekete, "Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1141–1148, 2008.

- [44] C. Ware, *Information Visualization: Perception for Design*, Morgan Kaufmann, Boston, Mass, USA, 3rd edition, 2013.
- [45] D. Williams and D. Freedman, "Asymptotics of graphical projection pursuit," *The Annals of Statistics*, vol. 12, no. 3, pp. 793–815, 1984.
- [46] F. Segovia, J. M. Górriz, J. Ramírez, and F. J. Martínez-Murcia, "Using deep neural networks along with dimensionality reduction techniques to assist the diagnosis of neurodegenerative disorders," *Logic Journal of the IGPL. Interest Group in Pure and Applied Logics*, vol. 26, no. 6, pp. 618–628, 2018.
- [47] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 11, pp. 559–572, 1901.
- [48] E. Oja, "Principal components, minor components, and linear neural networks," *Neural Networks*, vol. 5, no. 6, pp. 927–935, 1992.
- [49] C. Fyfe, "A neural network for PCA and beyond," *Neural Processing Letters*, vol. 6, no. 1-2, Article ID 1009606706736, pp. 33–41, 1997.
- [50] E. Corchado, D. MacDonald, and C. Fyfe, "Maximum and minimum likelihood Hebbian learning for exploratory projection pursuit," *Data Mining and Knowledge Discovery*, vol. 8, no. 3, pp. 203–225, 2004.
- [51] E. Corchado and C. Fyfe, "Connectionist techniques for the identification and suppression of interfering underlying factors," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 17, no. 8, pp. 1447–1466, 2003.
- [52] B. Andreopoulos, A. An, X. Wang, and M. Schroeder, "A roadmap of clustering algorithms: finding a match for a biomedical application," *Briefings in Bioinformatics*, vol. 10, no. 3, pp. 297–314, 2009.
- [53] W. Zhuang, Y. Ye, Y. Chen, and T. Li, "Ensemble clustering for internet security applications," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1784–1796, 2012.
- [54] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, p. 14, University of California Press, Berkeley, Calif, USA, 1967.
- [55] A. Terlaak and Y. Gong, "Vicarious learning and inferential accuracy in adoption processes," *Academy of Management Review (AMR)*, vol. 33, no. 4, pp. 846–868, 2008.
- [56] K. E. Meyer and E. Sinani, "When and where does foreign direct investment generate positive spillovers? A meta-analysis," *Journal of International Business Studies*, vol. 40, no. 7, pp. 1075–1094, 2009.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

