electronics

Article

A Fuzzy Entropy-Based Thematic Classification Method Aimed at Improving the Reliability of Thematic Maps in GIS Environments

Barbara Cardone [1] and Ferdinando Di Martino [1,2,*]

1 Dipartimento di Architettura, Università degli Studi di Napoli Federico II, Via Toledo 402, 80134 Napoli, Italy
2 Centro di Ricerca Interdipartimentale "Alberto Calza Bini", Università degli Studi di Napoli Federico II, Via Toledo 402, 80134 Napoli, Italy
* Correspondence: fdimarti@unina.it; Tel.: +39-081-253-8907

Abstract: Thematic maps of spatial data are constructed by using standard thematic classification methods that do not allow management of the uncertainty of classification and, consequently, evaluation of the reliability of the resulting thematic map. We propose a novel fuzzy-based thematic classification method applied to construct thematic maps in Geographical Information Systems. An initial fuzzy partition of the domain of the features of the spatial dataset is constructed using triangular fuzzy numbers; our method finds an optimal fuzzy partition evaluating the fuzziness of the fuzzy sets by using a fuzzy entropy measure. An assessment of the reliability of the final thematic map is performed according to the fuzziness of the fuzzy sets. We implement our method on a GIS framework, testing it on various vector and image spatial datasets. The results of these tests confirm that our thematic classification method provide thematic maps with a higher reliability with respect to that obtained through fuzzy partitions constructed by expert users.

Keywords: fuzzy entropy; fuzzy partition; fuzzy reliability; GIS; thematic map

Citation: Cardone, B.; Di Martino, F. A Fuzzy Entropy-Based Thematic Classification Method Aimed at Improving the Reliability of Thematic Maps in GIS Environments. *Electronics* **2022**, *11*, 3509. https://doi.org/10.3390/electronics11213509

Academic Editors: Moad Kissai, Bruno Monsuez and Barys Shyrokau

Received: 11 October 2022
Accepted: 26 October 2022
Published: 28 October 2022

1. Introduction

Thematic classification methods are an important tool used in various application fields that make use of spatial analysis processes in Geographical Information Systems (for short, GIS) such as: environmental sciences, urban planning, climate risk analysis, geomorphological studies of territory, etc.

Thematic maps are built in GIS environments by adopting thematic classification techniques to partition elements of a spatial dataset (the theme). These techniques perform a partitioning into equivalence classes of the domain of one or more features of a spatial dataset using specific rules for class construction. Each class, called a thematic class, is assigned a linguistic label and a symbol with which the spatial elements belonging to it will be displayed on the map.

A well-known thematic classification method partitions the theme into a number of classes equal to the number of unique values of a set of one or more features; this method, called Unique Value, is applicable only when the number of unique values of the features considered is much less than the number of entities of the theme to be represented on the map. When this does not happen, for example when the domain of a feature is a real interval, it is necessary to resort to other classification methods that segment this numerical interval into contiguous and disjoint sub-intervals. Each sub-interval, characterized by the values of its bounds, called breaks, is associated with a thematic class to which the entities whose value of the considered feature belongs to the sub-interval [1].

One of the best-known thematic classification methods is the Natural Breaks method, introduced by Jenks [2]. Natural Breaks is an optimization method that determines the values of the breaks of each thematic class in order to minimize the sum of the standard deviations of the values of the features of the elements belonging to each class [2]. This method is called natural as it provides partitioning of the theme in such a way that, with respect to the selected feature, all the elements of a class are similar among them and dissimilar to those of other classes.

The main flaw of the canonical thematic classification methods is the impossibility of attributing uncertainty to the classification. In order to manage this uncertainty, it is necessary to use a method based on the fuzzy partitioning of the domain of the selected feature, assigning a membership degree of an element to each thematic class. This approach has the advantage of modeling the expert's reasoning more accurately and of providing measures for assessing the reliability of the resulting thematic map.

In [3], a fuzzy multilevel comprehensive model is applied to assess the reliability of thematic maps.

A fuzzy classification method applied to spatial image data is proposed in [4], where the Morisita index [5] is used to assess the reliability of the thematic map. This index measures the similarity between two sets of data; it is used in [4] to assess the accuracy of the thematic map obtained using a fuzzy partition of the feature. This index is computationally complex to calculate and is difficult to interpret as a measure of the reliability of the resulting thematic map.

In [6], a logistic function proposed in [7] is used to fuzzify the domain of a numerical feature and to create a fuzzy classification of this feature. The critical point of this approach is that the resulting thematic classes can be difficult to interpret by the user.

A fuzzy thematic classification method based on fuzzy partitioning of a feature is proposed in [8]. The study area is partitioned into subzones and in each subzone is applied a method based on the calculus of the term frequency–inverse document frequency index (TF–IDF) to measure the relevance of a type of problem in a sub-area, extracted from the reports made by citizens over a period of time. The authors create a Ruspini fuzzy partition [9] of the measured relevance to construct thematic maps of the relevance of the type of problem. This approach has the advantage of producing thematic classifications connected to the user's reasoning, creating thematic maps that can be easily interpreted by the user.

The main shortcoming of the fuzzy-based thematic classification methods proposed in the literature is their difficult interpretability by the user; in fact, these methods do not take into account the user's approximate reasoning in the construction of thematic classes. Another critical point of these approaches is their inadequacy in assessing the reliability of the resulting thematic map.

In this research, we propose a fuzzy classification method based on the fuzzy partitioning of the selected feature, measuring the reliability of the resulting thematic map using the Fuzzy Entropy technique introduced by De Luca and Termini [10,11].

Fuzzy Entropy was used in [12] to measure the reliability of clusters detected in hotspot analysis; the Extended Fuzzy C-Means algorithm [13], an extended version of the Fuzzy C-Means algorithm [14–16], is applied to detect hotspots on the map; Fuzzy Entropy measures the fuzziness of the final fuzzy clusters.

In [17] is proposed a new validity index based on Fuzzy Entropy to find the best number of clusters in Fuzzy C-Means. In [18], Fuzzy Entropy is applied to find the optimal values of the cluster centers in Fuzzy C-Means.

We use Fuzzy Entropy to measure the fuzziness of the fuzzy sets composing the fuzzy partition of the domain of the selected feature.

The goal of our method is to determine the most suitable fuzzy partition to build the thematic map of the selected feature by measuring the fuzzy entropy of the fuzzy sets that make it up. Our method iteratively builds finer fuzzy partitions starting from an initial fuzzy partition. At each iteration, the fuzzy sets with the highest fuzzy entropy are split. The process ends when the fuzzy entropy of the fuzzy partition is less than a predetermined threshold, where the fuzzy entropy of the fuzzy partition is given by the average of the fuzzy entropy of its fuzzy sets.

The final fuzzy partition will be used to create the thematic map showing the spatial distribution of the feature selected in the study area; to this thematic will be assigned a reliability according to the fuzzy entropy of the fuzzy partition used.

We call our method Thematic Fuzzy Entropy Partition (for short, TFEP).

Initially, the user, based on their experience, creates a Ruspini fuzzy partition, setting its cardinality and using triangular fuzzy numbers to create the fuzzy sets.

Then, the fuzzy entropies of each fuzzy set are measured and the fuzzy entropy of the fuzzy partition is computed. If it is less than or equal to a prefixed threshold, the algorithm stops, and the thematic map is constructed assigning to it a reliability measure, a function of the fuzzy entropy of the fuzzy partition. Otherwise, the fuzzy partition is reconstructed, splitting the fuzzy set having the highest fuzzy entropy to form two fuzzy sets; then, the fuzzy entropies of all the fuzzy sets are recalculated; this process is iterated until the fuzzy entropies of the fuzzy relation are less than or equal to the threshold.

These steps are schematized in Figure 1.

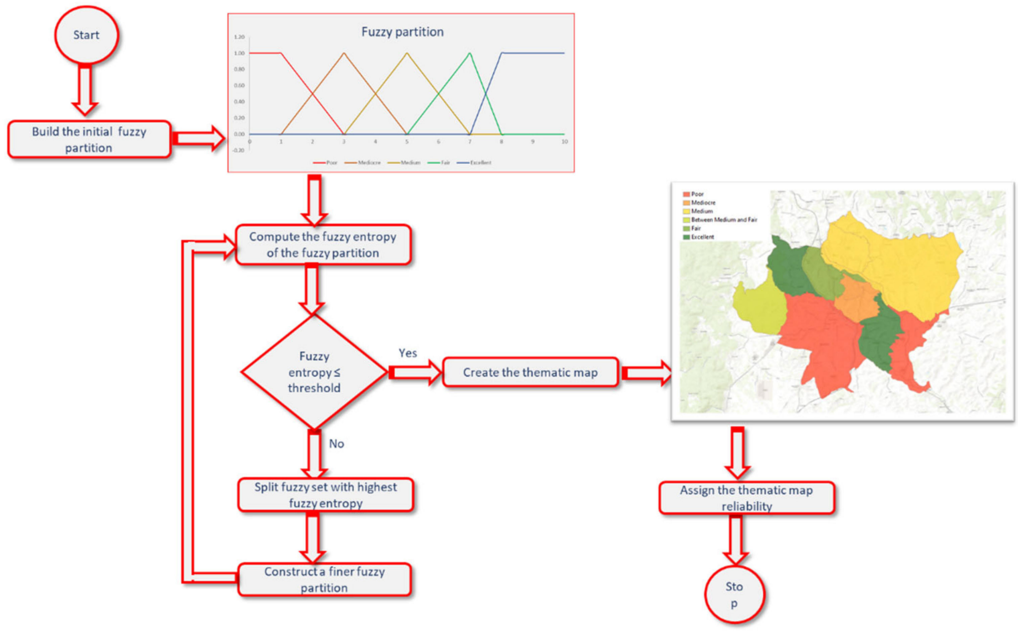

Figure 1. Schema of the Thematic Fuzzy Entropy Partition method (TFEP).

The main benefits of TFEP over the thematic classification methods well-known in the literature are summarized below:

- TFEP is applicable to any type of spatial data and builds the resulting thematic map from an initial fuzzy partition of the selected feature created by the user;
- The reliability of the resulting thematic map is assessed by using the fuzzy entropy. The reliability estimate allows the user to evaluate to what extent the thematic classification of the analyzed feature provides a spatial distribution of the entities included in the study area on the map adhering to the values assumed by the feature;
- The use of an iterative process of measurement of the fuzzy entropy of the fuzzy partition allows generation of a fuzzy partition whose fuzzy entropy is less than or equal to a specified threshold, in order to build a thematic map with an acceptable level of reliability.

Section 2 briefly describes the definition of fuzzy partition according to Ruspini and the measure of fuzzy entropy by De Luca and Termini. The TFEP method is presented in Section 3; Section 4 shows the results of its application in an area of study given by the municipalities of the province of Florence, in Italy. Final considerations and future perspectives are included in Section 5.

2. Preliminaries

2.1. Ruspini Fuzzy Partition

Let $F = \{A_1, A_2, \ldots, A_n\}$, a family of fuzzy sets defined on a universe of the discourse X, where A_i: $X \to [0, 1]$ is the membership function of the ith fuzzy set.

According to Ruspini [9], F is a fuzzy partition of X where each fuzzy set is not empty and the union of the membership degrees of an element to the fuzzy sets of F is the unity.

These two constraints are given by Equations (1) and (2):

$$\forall A_i \in F \;\; \exists x \in X \; : A_i(x) \neq 0 \tag{1}$$

$$\sum_{i=1}^{n} A_i(x) = 1 \; \forall x \in X \tag{2}$$

The Ruspini condition does not require that fuzzy sets must be disjoint.

Fuzzy numbers can be used to construct fuzzy partition. As an example, let x be a variable representing the percentage of cloud cover in a day. The universe of the discourse X is given by the interval [0%, 100%].

Let $F = \{A_1, A_2, A_3\}$, labeled, respectively, Sunny, Partly cloudy, and Overcast, and described by triangular fuzzy sets whose membership function is given as follows.

$$A_1(x) = \begin{cases} 1 & x < 20 \\ \frac{(x-20)}{20} & 20 \le x \le 40 \\ 0 & x > 40 \end{cases} \quad A_2(x) = \begin{cases} 0 & x < 20 \\ \frac{(x-20)}{20} & 20 \le x \le 40 \\ \frac{80-x}{40} & 40 \le x \le 80 \\ 0 & x > 80 \end{cases} \quad A_3(x) = \begin{cases} 0 & x < 40 \\ \frac{(x-40)}{40} & 40 \le x \le 80 \\ 1 & x > 80 \end{cases} \tag{3}$$

The first fuzzy set is an R-function, the second fuzzy set a triangular function, and the third fuzzy set an L-function.

Figure 2 shows the three fuzzy sets.

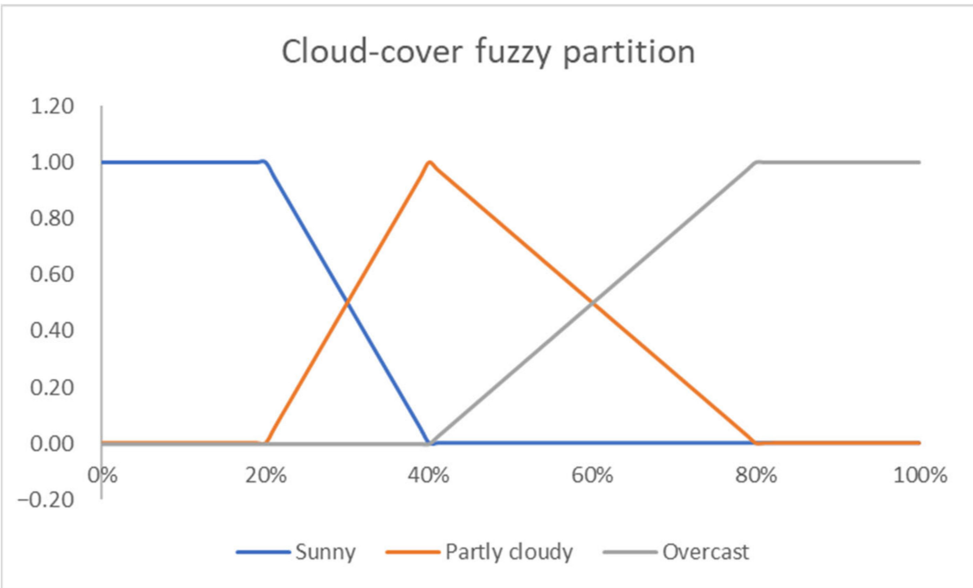

Figure 2. Example of fuzzy partition built using fuzzy numbers.

F is a fuzzy partition as the constraints (1) and (2) are respected.

2.2. Fuzzy Entropy Function and Fuzziness of a Fuzzy Set

The De Luca and Termini fuzzy entropy measure was introduced in [10,11] to evaluate the fuzziness of a fuzzy set.

Let h: $[0, 1] \to [0, 1]$ be a continuous function; it is called a fuzzy entropy function if:

1. h(1) = 0
2. h(u) = h(1 − u)
3. is monotonically increasing in in $[0, \frac{1}{2})$
4. is monotonically decreasing in in $[\frac{1}{2}, 1]$

In [10,11], the following fuzzy entropy function, called Shannon's function, is proposed:

$$h(u) = \begin{cases} 0 & \text{if } u = 0 \\ -u \cdot \log_2(u) - (1-u) \cdot \log_2(1-u) & \text{if } 0 < u < 1 \\ 0 & \text{if } u = 1 \end{cases} \tag{4}$$

Shannon's function is shown in Figure 3. It has the minimum value 0 when u = 0 or u = 1 and the maximum value 1 when $u = \frac{1}{2}$.

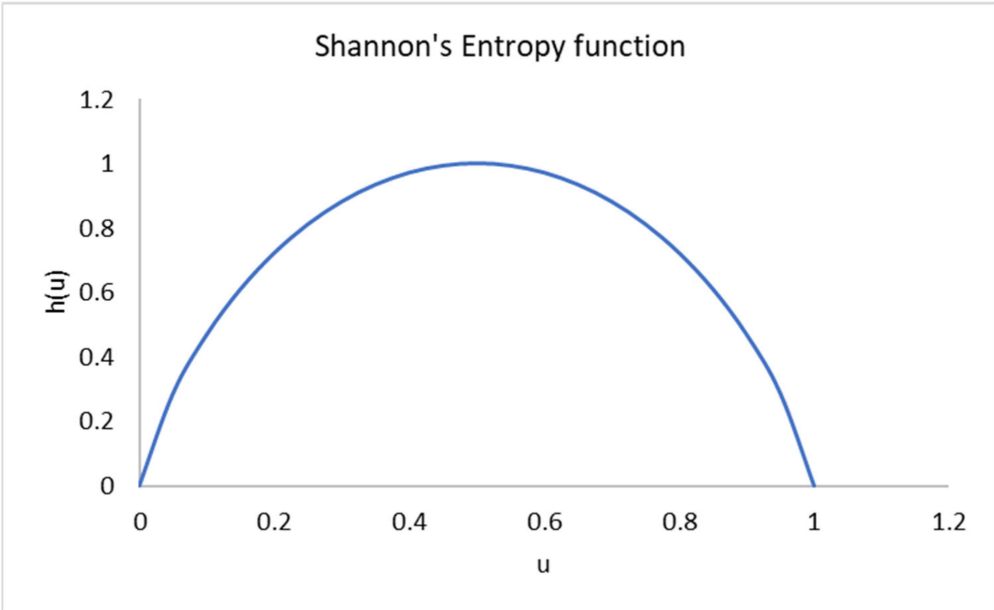

Figure 3. Shannon's Entropy function.

In [19], the connection between fuzzy entropy and conventional classical entropy is discussed and the use of fuzzy entropy in the management of uncertainty in decision-making applications is shown. Applications of fuzzy entropy in decision-making problems are proposed in [20–22].

The De Luca and Termini fuzzy entropy technique can be used to express the fuzziness of a fuzzy set.

Formally, let $X = \{x_1, x_2, \ldots, x_N\}$ be a discrete set of values; the fuzziness of the fuzzy set A defined on X is given by the formula:

$$H(A) = K \sum_{j=1}^{N} h(A(x_j)) \tag{5}$$

where K is a multiplicative constant set to normalize the measure of the fuzziness. If we set the maximum value of H(A) = 1 (all the objects belong to A with membership degree $A(x_j) = \frac{1}{2}$), then K is equal to 1/N, obtaining:

$$H(A) = \frac{1}{N} \sum_{j=1}^{N} h(A(x_j)) \tag{6}$$

The fuzziness of the fuzzy set A in (5) corresponds to the average of the membership degrees of the elements of X to A. If A is a crisp set, then for any element x_j, j = 1, ... , N $A(x_j) = 0$ or $A(x_j) = 1$ and H(A) = 0.

The fuzziness measure (5) is used in [17,18] to evaluate the fuzziness of the fuzzy clusters in Fuzzy C-means and to determine the optimal number of clusters.

The reliability of a fuzzy set is measured in [12,23] to evaluate the reliability of fuzzy clusters detected in hot spot analysis. It is given by the formula:

$$\mathrm{R(A) = 1 - H(A)} \tag{7}$$

where the fuzziness H(A) of the fuzzy set A is measured by (5).

As an example, let X = $\{x_1, x_2, \ldots, x_6\}$ be a discrete set of six elements having values in the Real domain U = [0, 100]. Let A: U→ [0, 1] be a triangular fuzzy set given by:

$$\mathrm{A(x)} = \begin{cases} 0 & x < 10 \\ \frac{(x-10)}{10} & 10 \le x \le 20 \\ \frac{40-x}{20} & 20 \le x \le 40 \\ 0 & x > 40 \end{cases} \tag{8}$$

In Table 1 are shown, for each element of X, its value, the membership degree to the fuzzy set A, and the entropy function given by (8).

Table 1. Example of calculus of the fuzziness of a fuzzy set.

X	x	A(x)	h(A(x))
x_1	61.19	0.00	0.00
x_2	15.50	0.55	0.99
x_3	29.88	0.51	1.00
x_4	31.01	0.45	0.99
x_5	28.76	0.56	0.99
x_6	8.33	0.00	0.00

The fuzziness of the fuzzy set A, given by (5), is: H(A) = 0.66. The reliability is R(A) = 1 − H(A) = 0.33.

3. The Thematic Fuzzy Entropy Partition Method

We propose an iterative fuzzy thematic classification method to find the best fuzzy partition of the domain of the selected feature; we use the fuzzy entropy measure to assess the fuzziness of a fuzzy partition and to evaluate the reliability of the final thematic map.

Let E = $\{e_1, e_2, \ldots, e_N\}$ be a theme formed by N georeferenced entities e_i $i = 1, \ldots, N$. The aim of the TFEP method is to find the best thematic classification of the entities of E based on a selected feature.

Let X = $\{x_1, x_2, \ldots, x_N\}$ be the numerical discrete set of the feature values, where x_i is the feature value taken by the entity e_i. X is a subset of a real close interval U.

Let F = $\{A_1, A_2, \ldots A_M\}$ be a fuzzy partition of U, where A_k: U→ [0, 1], k = 1, 2, ... M. To the set A_k is assigned by the user a label C_k.

F is the initial fuzzy partition created by the user to construct a thematic classification of the entities of the theme E using triangular fuzzy sets. The choice of triangular fuzzy sets is motivated by their ease of representation and modeling as fuzzy numbers expressed by triplets of numbers.

TFEP measures the fuzziness of the fuzzy sets $A_1, A_2, \ldots, A_M$ using (9) and computes the fuzziness of the fuzzy partition F given by the average of the fuzziness of M fuzzy sets:

$$\mathrm{H}(F) = \frac{1}{\mathrm{M}} \sum_{\mathrm{k}=1}^{\mathrm{M}} \mathrm{H}(\mathrm{A_k}) \tag{9}$$

If $H(F)$ is greater than a fuzziness threshold H_{Th}, the fuzzy set having the highest fuzziness is split into two fuzzy sets and a new finer fuzzy partition formed by M + 1 fuzzy sets is created. If A_k is the fuzzy set with the highest fuzziness, it is split into two fuzzy sets to which the following labels are assigned, respectively: almost C_k and more than C_k.

The splitting of the fuzzy set A consisting of a fuzzy number (a, b, c) in two fuzzy sets is achieved by constructing two fuzzy numbers (a_1, b_1, c_1) and (a_2, b_2, c_2), where:

$$\begin{array}{cc} a_1 = a & a_2 = b - \frac{b-a}{2} = b_1 \\ b_1 = b - \frac{b-a}{2} & b_2 = b + \frac{c-b}{2} = c_1 \\ c_1 = b + \frac{c-b}{2} & c_2 = c \end{array} \tag{10}$$

Moreover, in order to respect the Ruspini condition:

- If the previous fuzzy set exists and it consists of the fuzzy number $(a_{Prev}, b_{Prev}, c_{Prev})$, the value of c_{Prev}, previously set equal to b, is changed to b_1;
- If the next fuzzy set exists and it consists of the fuzzy number $(a_{Next}, b_{Next}, c_{Next})$, the value of a_{Next}, previously set equal to b, is changed to b_2.

The process is iterated until the fuzziness of the current fuzzy partition F is less than or equal to H_{Th}.

Finally, the thematic map of E is built using the fuzzy partition F found. An entity is assigned to the class corresponding to the fuzzy set to which it belongs with the highest degree.

The user can change the labels of the fuzzy sets in the final fuzzy partition if the user intends to combine them into semantically more meaningful names.

To the thematic map is assigned a reliability given by the formula:

$$R = 1 - H(F) \tag{11}$$

The TFEP algorithm is schematized below in Algorithm 1.

Algorithm 1: Thematic Fuzzy Entropy Partition (TFEP)

1. *Initialize* the fuzzy partition F given by M fuzzy sets.
2. Set the fuzziness threshold H_{Th}
3. $H \leftarrow 1$ //variable containing the fuzziness of the fuzzy partition
4. **While** $H > H_{Th}$
5.   **For** k = 1 to M //for each fuzzy set of F
6.   *Compute* the fuzziness of A_k $H(A_k)$ by (5)
7.   **Next** k
8.   *Compute* $H(F)$ by (7)
9. $H \leftarrow H(F)$
10.   **If** $H > H_{Th}$
11.   *Split* the fuzzy sets with highest fuzziness in two fuzzy sets by (8)
12.   *Assign* the labels of the two fuzzy sets using the prefixes *almost* and *more than*
13. $M \leftarrow M + 1$
14.   *Construct* the new fuzzy partition F having M fuzzy sets
15.   **End If**
16. **End While**
17. *Create* the thematic map using the fuzzy partition F
18. *Compute* the reliability of the thematic map by (9)

We now show an example of application of the TFEP method, considering the discrete set $X = \{x_1, x_2, \ldots, x_6\}$ shown in the example in Table 1.

Initially we construct a fuzzy partition with the four fuzzy sets in Figure 4. The four fuzzy sets A_1, A_2, A_3, and A_4 are labeled, respectively, Low, Medium low, Medium high, and High. A_1 is an R-function, A_2 and A_3 are triangular fuzzy sets, and A_4 is an L-function. Their membership degree functions are given by:

$$A_1(x) = \begin{cases} 1 & x < 10 \\ \frac{(x-10)}{10} & 10 \le x \le 20 \\ 0 & x > 20 \end{cases} \quad A_2(x) = \begin{cases} 0 & x < 10 \\ \frac{(x-10)}{10} & 10 \le x \le 20 \\ \frac{40-x}{20} & 20 \le x \le 40 \\ 0 & x > 40 \end{cases}$$

$$A_2(x) = \begin{cases} 0 & x < 10 \\ \frac{(x-20)}{20} & 20 \le x \le 40 \\ \frac{70-x}{30} & 40 \le x \le 70 \\ 0 & x > 70 \end{cases} \quad A_3(x) = \begin{cases} 0 & x < 40 \\ \frac{(x-40)}{30} & 40 \le x \le 70 \\ 1 & x > 70 \end{cases}$$

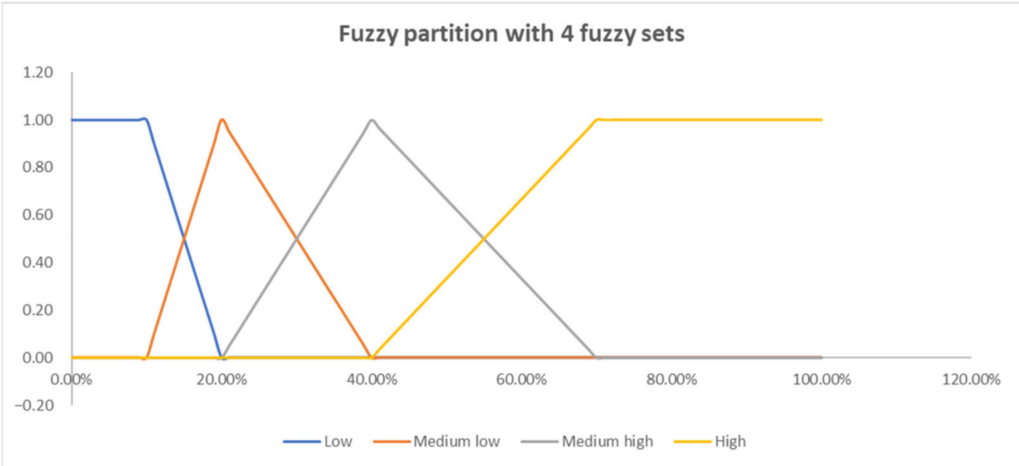

Figure 4. Example of fuzzy partition of the domain [0%,100%] with 4 fuzzy sets.

We set the fuzziness threshold H_{Th} to 0.60, considering that the average degree of belonging of an element to a fuzzy set with a fuzziness equal to 0.6 is equal to 0.15 or 0.85. In Table 2 are shown the values of the fuzziness calculated for the four fuzzy sets. The fuzzy set having the highest fuzziness is the fuzzy set A_2; its fuzziness is 0.66, greater than the threshold.

Table 2. Calculus of the fuzziness of the 4 fuzzy sets in Figure 3.

X	x	A(x)				h(A(x))			
		Low	**Medium Low**	**Medium High**	**High**	**Low**	**Medium Low**	**Medium High**	**High**
x_1	61.19%	0.00	0.00	0.29	0.71	0.00	0.00	0.87	0.87
x_2	15.50%	0.45	0.55	0.00	0.00	0.99	0.99	0.00	0.00
x_3	29.88%	0.00	0.51	0.49	0.00	0.00	1.00	1.00	0.00
x_4	31.01%	0.00	0.45	0.55	0.00	0.00	0.99	0.99	0.00
x_5	28.76%	0.00	0.56	0.44	0.00	0.00	0.99	0.99	0.00
x_6	8.33%	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
			H(A)			0.17	**0.66**	0.64	0.15

In the next step, the fuzzy set $A_{2;}$ is split into two fuzzy sets by (8). Table 3 show the fuzziness of all fuzzy sets. Now, all fuzzy sets have a fuzziness less than the threshold and the process stops.

Table 3. Calculus of the fuzziness of the 5 fuzzy sets after splitting the fuzzy set A_2.

X	x	A(x)					h(A(x))				
		Low	Almost Medium Low	More than Medium Low	Medium High	High	Low	Almost Medium Low	More than Medium Low	Medium High	High
x_1	61.19%	0.00	0.00	0.00	0.29	0.71	0.00	0.00	0.00	0.87	0.87
x_2	15.50%	0.00	0.95	0.05	0.00	0.00	0.00	0.29	0.29	0.00	0.00
x_3	29.88%	0.00	0.00	0.67	0.33	0.00	0.00	0.00	0.91	0.91	0.00
x_4	31.01%	0.00	0.00	0.60	0.40	0.00	0.00	0.00	0.97	0.97	0.00
x_5	28.76%	0.00	0.00	0.75	0.25	0.00	0.00	0.00	0.81	0.81	0.00
x_6	8.33%	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
				H(A)			0.00	0.05	0.50	0.59	0.15

Figure 5 shows the final fuzzy partition for this example. The user, assigning the labels to the five thematic classes, can decide to use the label of the corresponding fuzzy set or to use other labels that the user deems semantically more suitable.

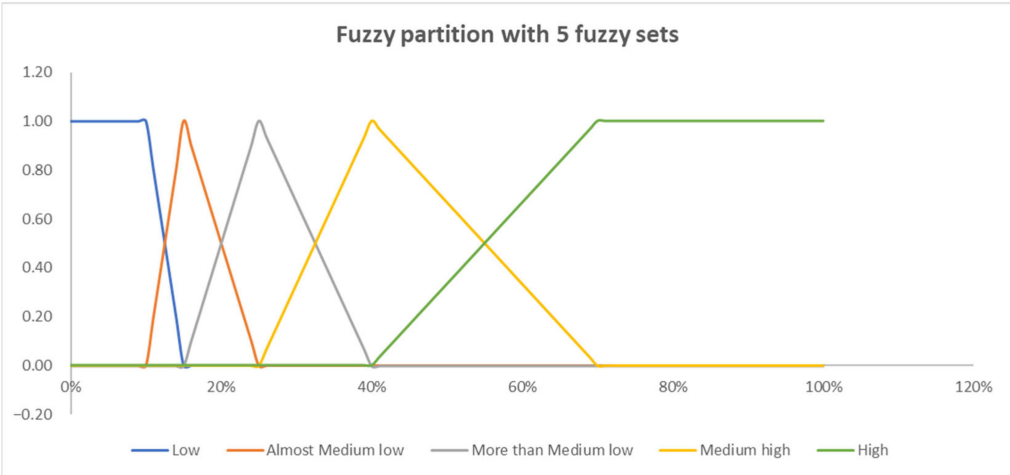

Figure 5. Final fuzzy partition with 5 fuzzy sets.

4. Experimental Results

We implemented the TFEP algorithm in the Tool GIS ArcGIS Desktop 10.8, using the *ArcPy* Python libraries.

We test TFEP, executing it on over 100 vector and image spatial datasets with different cardinalities, setting the fuzzy entropy threshold H_{Th} to 0.60. For brevity, we show the results obtained for a vector and an image spatial dataset.

In this first test, we execute TFEP to extract a thematic map showing the spatial distribution of the number of inhabitants by residential building in an urban study area given by the 44 municipalities of the province of Florence (Italy). The source spatial dataset is a vector dataset consisting of census data provided by the Italian National Institute of Statistics (ISTAT) (https://www.istat.it/it/archivio/104317 (accessed on 1 July 2022).

The study area is shown in Figure 6.

We asked an expert to create the initial fuzzy partition on the domain given by the numerical interval in which the number of inhabitants of a residential building varies; this fuzzy partition, composed by three triangular fuzzy numbers, is shown in Figure 7.

Each fuzzy set is expressed by a triplet (a, b, c); in Table 4 are shown the label of each fuzzy set, the type of triangular fuzzy number used, the values of the parameters in the triplet, and the measured fuzziness.

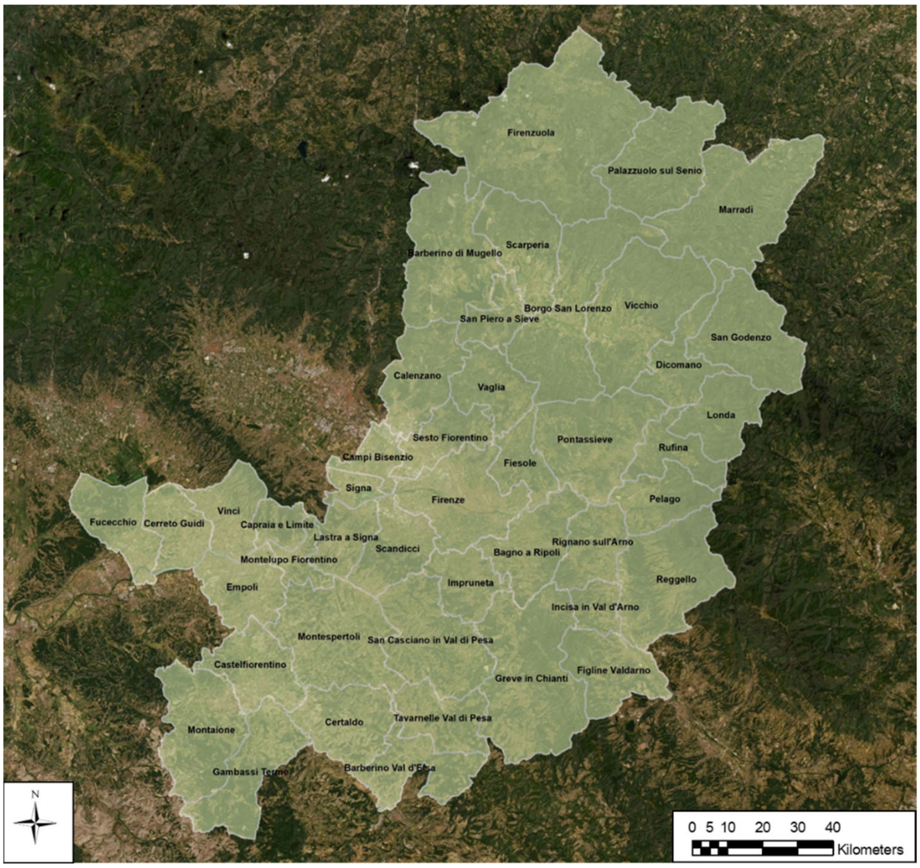

Figure 6. The study area: the municipalities of the province of Florence (Italy).

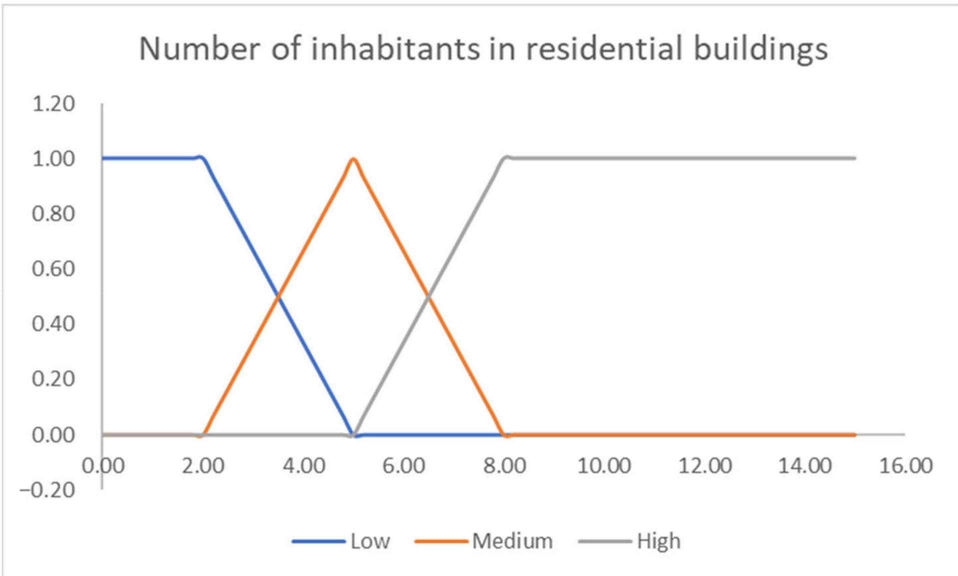

Figure 7. The initial fuzzy partition given by 3 fuzzy numbers.

Table 4. Fuzzy sets in the fuzzy partition in Figure 7 and their fuzziness.

A	Fuzzy Number				H(A)
	Type	**a**	**b**	**c**	
Low	R-function		2.00	5.00	0.17
Medium	Triangular	2.00	5.00	8.00	**0.62**
High	L-function	5.00	8.00		0.45

Figure 8 shows the thematic map obtained using the fuzzy partition in Figure 7.

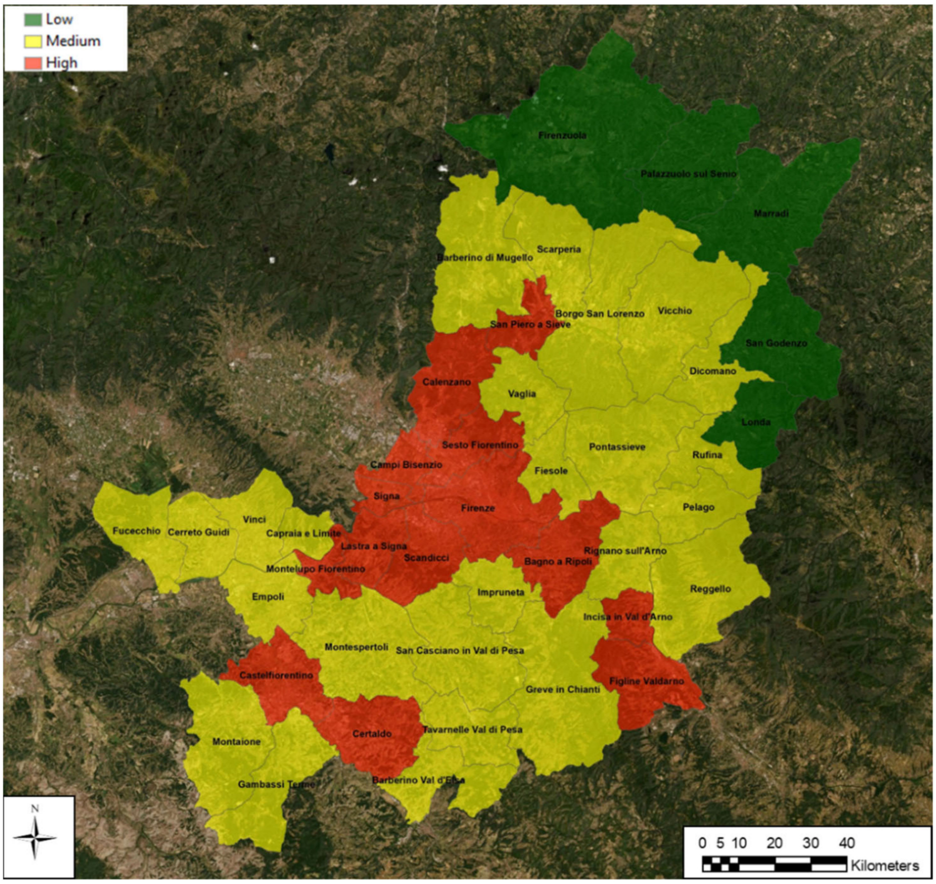

Figure 8. The thematic map of the number of inhabitants in residential buildings using the initial fuzzy partition.

The fuzziness $H(F)$ of this fuzzy partition, computed by (7) is 0.41; then, the reliability of the thematic map in Figure 8 is 0.59.

Since the fuzziness of the fuzzy set Medium is higher than the threshold, TFEP splits this fuzzy set into two fuzzy sets, called, respectively, Almost Medium and More than medium. In Figure 9 is shown the new fuzzy partition having four fuzzy sets.

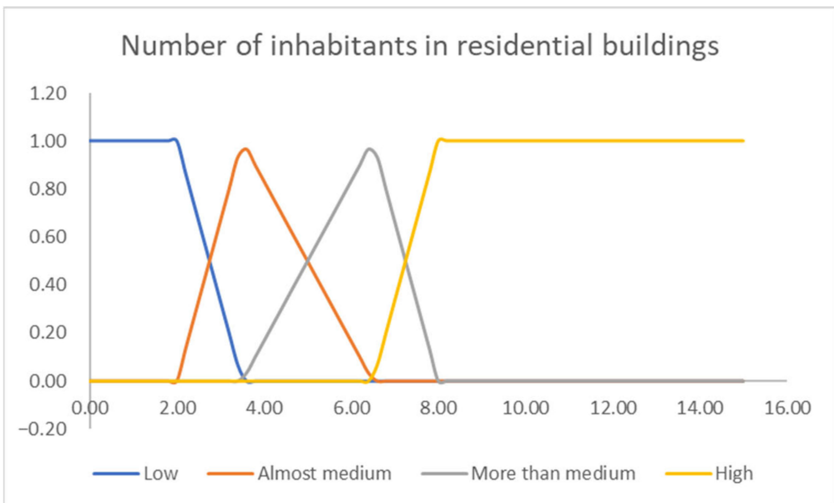

Figure 9. The new fuzzy partition given by 4 fuzzy numbers.

In Table 5 are shown the label of each fuzzy set, the type of triangular fuzzy number used, the values of the parameters in the triplet, and the measured fuzziness.

Table 5. Fuzzy sets in the fuzzy partition in Figure 9 and their fuzziness.

A	Fuzzy Number				H(A)
	Type	a	b	c	
Low	R-function		2.00	3.50	0.05
Almost medium	Triangular	2.00	3.50	6.50	0.48
More than medium	Triangular	3.50	6.50	8.00	0.59
High	L-function	6.50	8.00		0.17

Now, the fuzziness of all fuzzy sets is below the threshold H_{Th}. TFEP stops, generating the final thematic map shown in Figure 10.

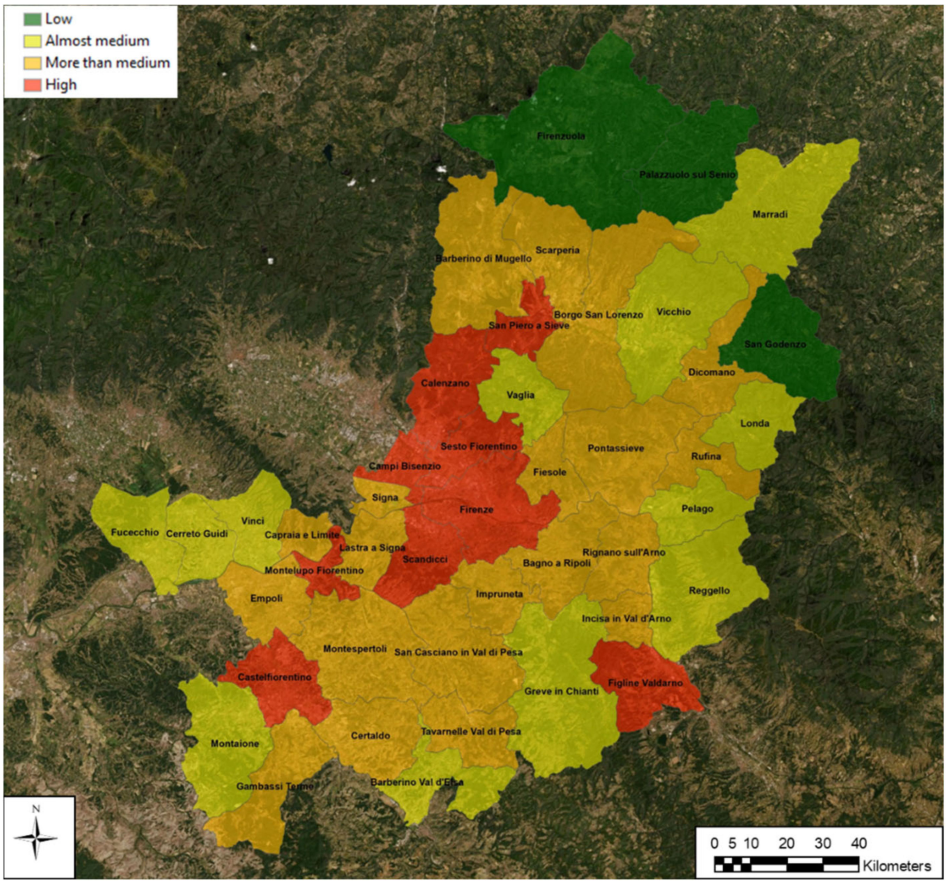

Figure 10. Final thematic map of the number of inhabitants in residential buildings.

The fuzziness H(F) of the final fuzzy partition, computed by (7), is 0.32 and the reliability of the thematic map in Figure 10 is 0.68.

Now, we show the results of another test executed to obtain a thematic map of a spatial image dataset. It is given by a 1 m × 1 m remote sensing image dataset of the Normalized Difference Vegetation Index (for short, NDVI) obtained in June 2022 by the Sentinel 2 satellite on the study area of the municipality of Naples (Italy).

The NDVI is obtained by a combination of the satellite Near-Infrared (NIR) and Red (R) bands, using the formula: $\mathrm{NDVI} = \frac{\mathrm{NIR}-\mathrm{R}}{\mathrm{NIR}+\mathrm{R}}$.

This allows evaluation of some characteristics of the vegetation. It varies between -1 and 1; values between -1 and 0 are typical of uncultivated and anthropogenic areas. Values between 0 and 1 correspond to cultivated or wooded areas; the closer these values are to 1, the greater the soil vegetation cover and the evapotranspiration capacity of the vegetation.

In Figure 11 is shown the NDVI image dataset on the study area.

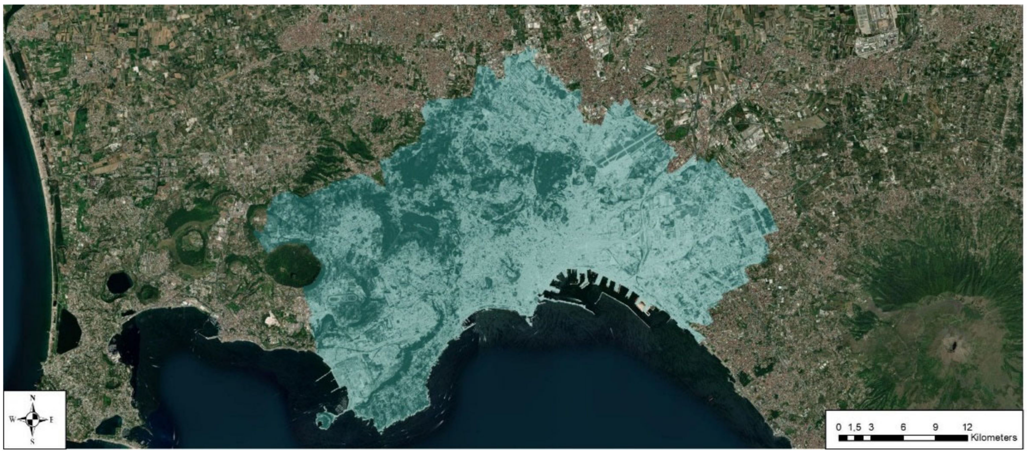

Figure 11. The study area: satellite image of the NDVI index on the municipality of Naples (Italy).

The expert sets the initial fuzzy partition on the domain [−1, 1], given by four triangular fuzzy sets as in Figure 12.

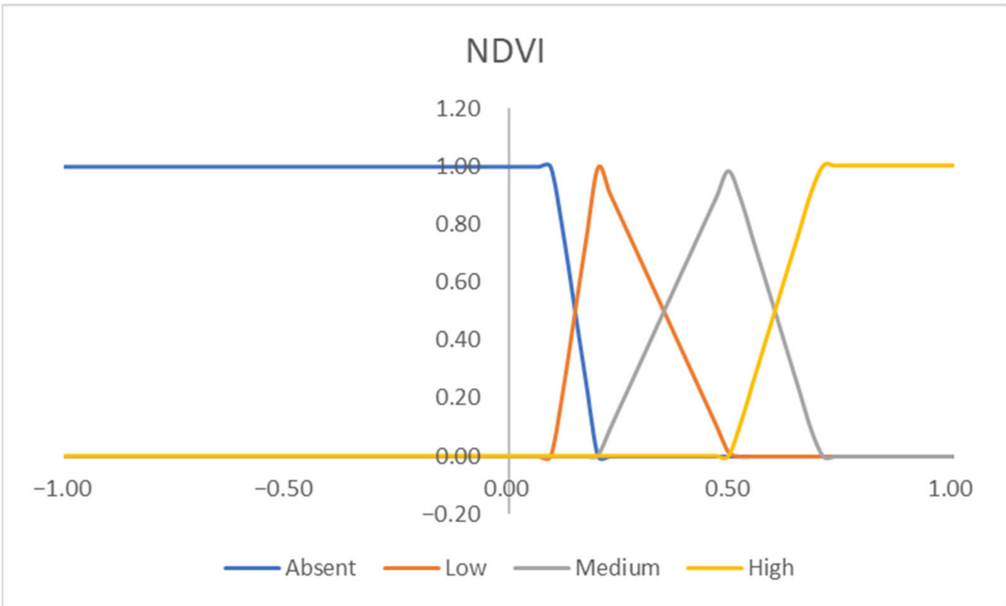

Figure 12. The initial fuzzy partition of the NDVI index given by 4 fuzzy numbers.

In Table 6 are shown the label of each fuzzy set, the type of triangular fuzzy number used, the values of the parameters a, b, and c in the triplet, and the measured fuzziness. In bold is shown the highest fuzziness.

Table 6. Fuzzy sets in the fuzzy partition in Figure 12 and their fuzziness.

A	Fuzzy Number				H(A)
	Type	a	b	c	
Absent	R-function		0.10	0.20	0.22
Low	Triangular	0.10	0.20	0.50	0.57
Medium	Triangular	0.20	0.50	0.70	**0.66**
High	L-function	0.50	0.70		0.41

Figure 13 shows the thematic map obtained using the fuzzy partition in Figure 12.

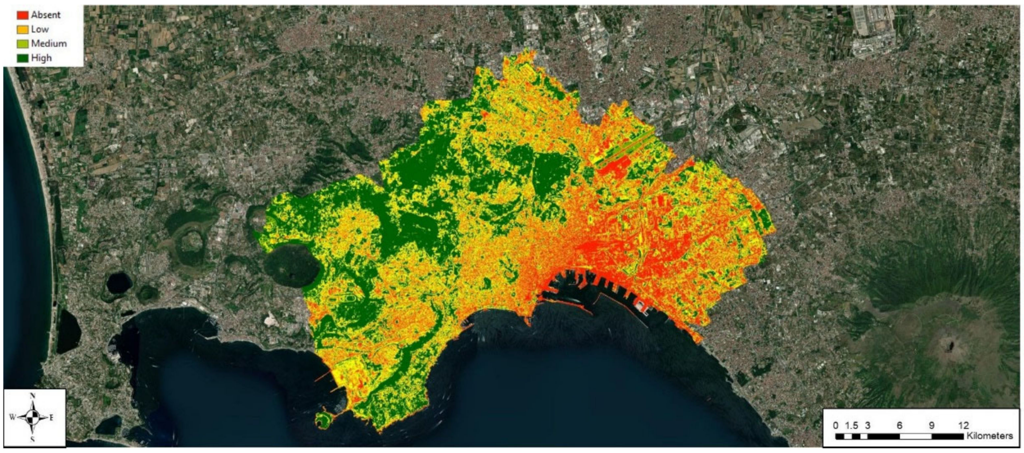

Figure 13. The thematic map of the NDVI index using the initial fuzzy partition.

The fuzziness $H(F)$ of this fuzzy partition, computed by (7) is 0.47; then, the reliability of the thematic map in Figure 13 is 0.53.

Since the fuzziness of the fuzzy set Medium is higher than the threshold, TFEP splits this fuzzy set into two fuzzy sets, called, respectively, Almost Medium and More than medium. In Figure 14 is shown the new fuzzy partition having five fuzzy sets.

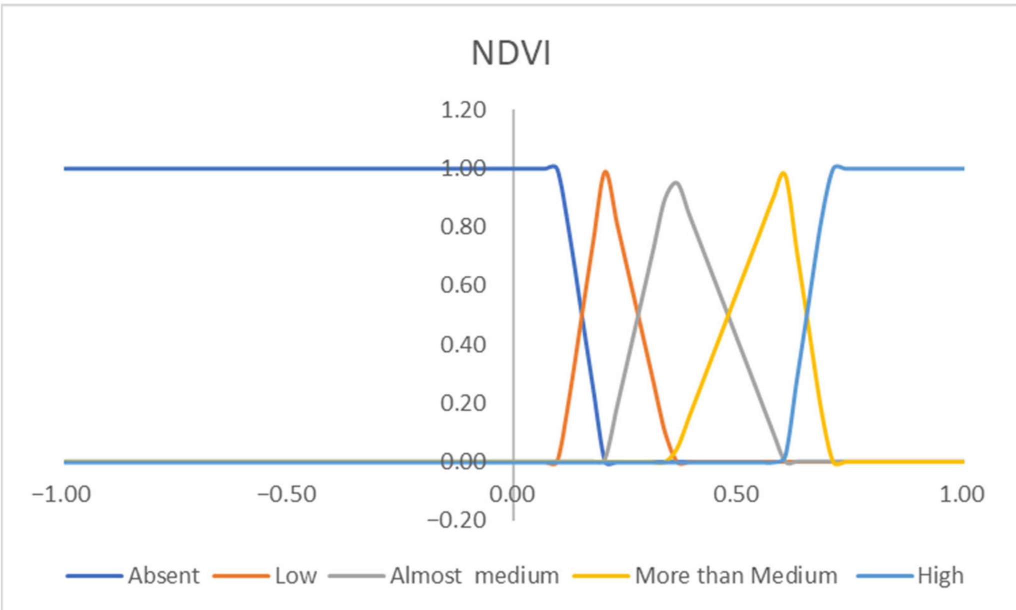

Figure 14. The second fuzzy partition of NDVI given by 5 fuzzy numbers.

In Table 7 are shown the label of each fuzzy set, the type of triangular fuzzy number used, the values of the parameters in the triplet, and the measured fuzziness. In bold is shown the highest fuzziness.

Figure 15 shows the thematic map obtained using the fuzzy partition in Figure 14.

The fuzziness $H(F)$ of this fuzzy partition, computed by (7) is 0.46; then, the reliability of the thematic map in Figure 15 is 0.54.

Since the fuzziness of the fuzzy set More than medium is higher than the threshold, TFEP splits this fuzzy set into two fuzzy sets, called, respectively, Almost more than medium and More more than medium. In Figure 16 is shown the new fuzzy partition having six fuzzy sets.

Table 7. Fuzzy sets of the fuzzy partition in Figure 14 and their fuzziness.

A	Fuzzy Number				H(A)
	Type	a	b	c	
Absent	R-function		0.10	0.20	0.22
Low	Triangular	0.10	0.20	0.35	0.52
Almost medium	Triangular	0.20	0.35	0.60	0.53
More than medium	Triangular	0.35	0.60	0.70	**0.62**
High	L-function	0.60	0.70		0.38

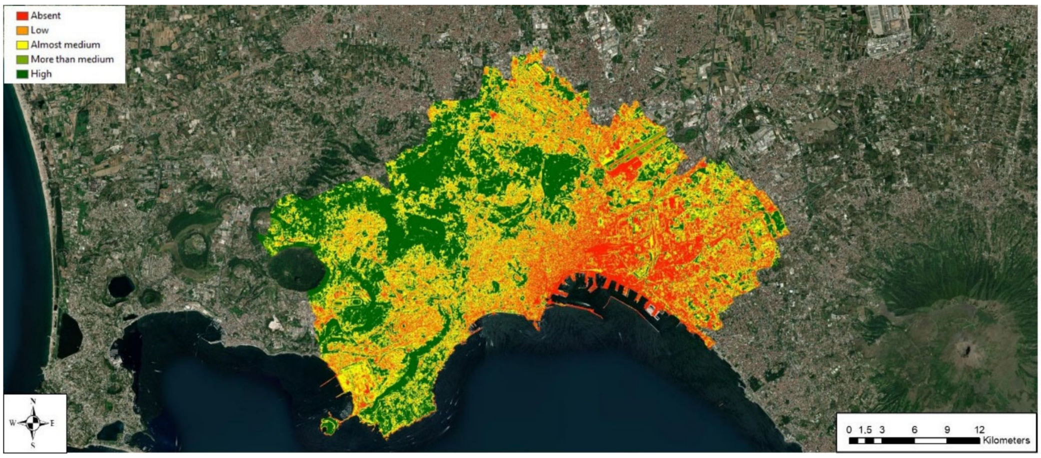

Figure 15. The thematic map of the NDVI index using fuzzy partition in Figure 14.

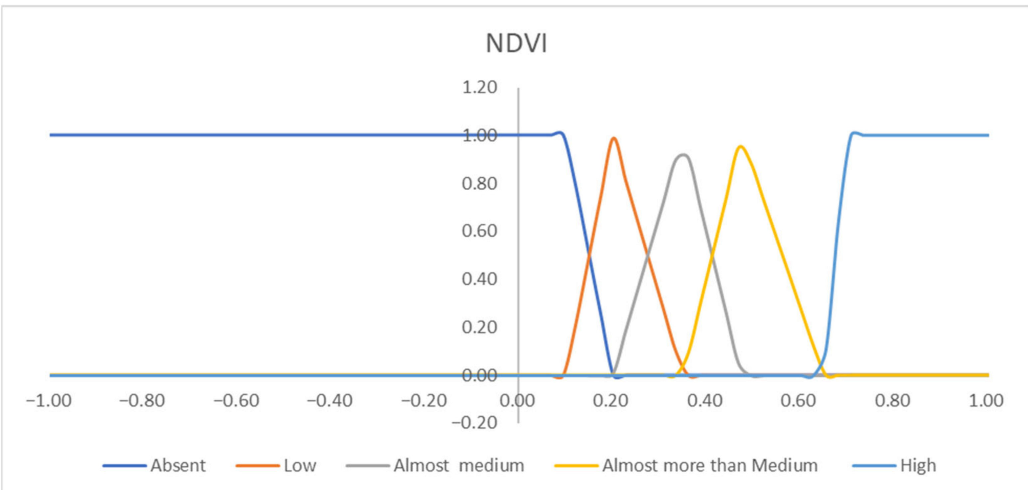

Figure 16. The third fuzzy partition of NDVI given by 6 fuzzy numbers.

In Table 8 are shown the label of each fuzzy set, the type of triangular fuzzy number used, the values of the parameters in the triplet, and the measured fuzziness.

Table 8. Fuzzy sets of the fuzzy partition in Figure 14 and their fuzziness.

A	Fuzzy Number				H(A)
	Type	**a**	**b**	**c**	
Absent	R-function		0.10	0.20	0.22
Low	Triangular	0.10	0.20	0.35	0.52
Almost medium	Triangular	0.20	0.35	0.48	0.48
Almost more than medium	Triangular	0.35	0.48	0.65	0.49
More more than medium	Triangular	0.48	0.65	0.70	0.50
High	L-function	0.65	0.70		0.33

The fuzziness of all fuzzy sets is below the threshold H_{Th}. TFEP stops. The expert decided to rename the labels of the six fuzzy sets in order to semantically represent the final thematic map more clearly. The new labels assigned by the expert to the final fuzzy sets are shown in Table 9.

Table 9. New labels assigned by the expert to the final fuzzy sets.

Label Final Fuzzy Set	New Label Final Fuzzy Set
Absent	Absent
Low	Low
Almost medium	Medium low
Almost more than medium	Medium
More more than medium	Medium high
High	High

The final thematic map is shown in Figure 17.

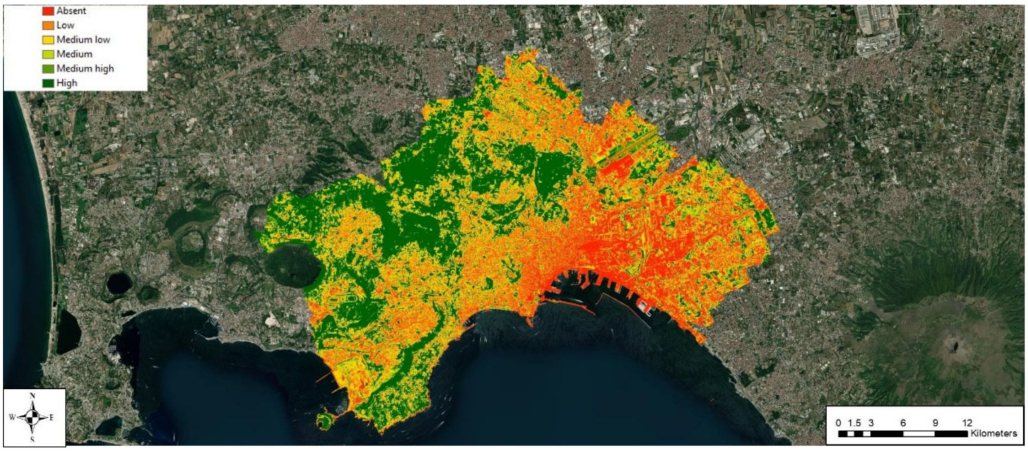

Figure 17. The final thematic map of the NDVI index.

The fuzziness $H(F)$ of the fuzzy partition, computed by (7) is 0.42; then, the reliability of the final thematic map is 0.58.

To assess the benefits of our fuzzy thematic classification method in terms of increase in the map reliability, we measure the difference between the reliability of the final thematic map and the reliability of the initial thematic map, calling this difference Map-reliability gain.

In Figure 18 is shown the trend of the map-reliability gain with respect to the map reliability of the initial thematic map, obtained by measuring the map-reliability gain in all tests.

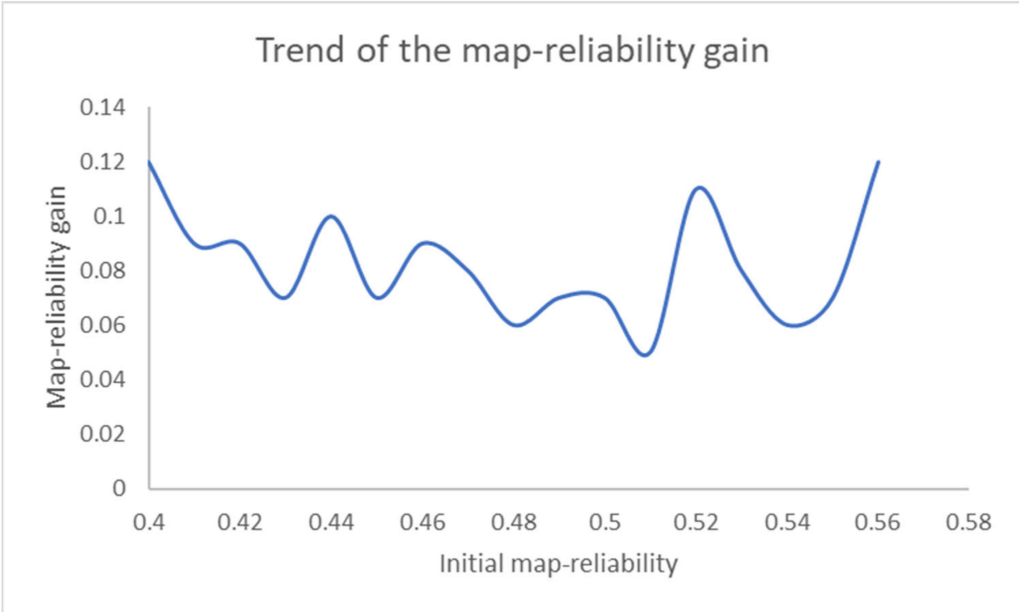

Figure 18. Trend of the map-reliability gain.

This trend shows that TFEP always increases the reliability of the initial thematic map, with an average gain between 0.04 and 0.12.

In order to compare the performances of TFEP with respect to those of other fuzzy thematic classification methods, we measured, for each of the spatial datasets used in the tests, the reliability of the final thematic classification obtained using the methods proposed in [4,6,8]. Table 10 shows the average, standard deviation, minimum, and maximum values of the map-reliability gain with respect to the map reliability obtained with each of the other fuzzy thematic classification methods. The reliability gain is always positive and, on average, varies between 0.08 and 0.10.

Table 10. TFEP reliability gain with respect to other fuzzy thematic classification methods.

Method	Reliability Gain			
	Average	**StD**	**Min**	**Max**
[4]	0.10	0.05	0.05	0.16
[6]	0.09	0.04	0.04	0.12
[8]	0.08	0.04	0.03	0.10

These results show that TFEP produces thematic maps whose reliability is higher than that of thematic maps generated using other fuzzy thematic classification methods.

In a nutshell, the results of our tests show that, regardless of the type and cardinality of the theme, TFEP allows you to build reliable thematic maps, building fuzzy partitions whose fuzzy sets have a fuzziness not exceeding a predetermined threshold. In fact, the static assignment of the fuzzy partition can generate unreliable thematic maps due to the high fuzziness of fuzzy sets. Assigning the spatial element to the fuzzy set with the highest membership degree results in classification uncertainty, as membership degrees to other fuzzy sets are not considered; when the fuzziness of a fuzzy set is high, this uncertainty becomes non-negligible. TFEP's strategy in these cases is to build a finer fuzzy partition by splitting the fuzzy set with the highest fuzziness.

5. Conclusions

We propose a novel fuzzy thematic classification method implemented in GIS environments to produce thematic maps and to assess their reliabilities. Initially, a fuzzy partition of the selected feature of the theme is performed by the user using triangular fuzzy numbers; at each iteration is measured the fuzziness of the fuzzy sets measured by the De Luca and Termini fuzzy entropy measure, splitting the fuzzy set with the highest fuzziness. The process stops when the fuzziness of each fuzzy set is below a prefixed threshold. The reliability of the thematic map is measured in functions of the fuzziness of the final fuzzy partition.

The results of our tests in which TFEP is applied on both vector and image spatial datasets show that our method increases the reliability of the initial thematic map, regardless of the type and cardinality of the theme.

In the future, we intend to extend TFEP to Interval Type-2 fuzzy sets, in order to allow the user to evaluate the uncertainty of the fuzzy sets of the initial fuzzy partition and also to consider this uncertainty to refine the assessment of the final thematic map reliability. A further line of research we intend to pursue refers to the construction of an evolutionary fuzzy-based thematic classification approach that employs the use of a fitness function correlated with classification reliability.

Author Contributions: The contributions of the two authors B.C. and F.D.M. are summarized below: conceptualization, B.C. and F.D.M.; methodology, B.C. and F.D.M.; software, B.C. and F.D.M.; validation, B.C. and F.D.M.; formal analysis, B.C. and F.D.M.; investigation, B.C. and F.D.M.; resources, B.C. and F.D.M.; data curation, B.C. and F.D.M.; writing—original draft preparation, B.C. and F.D.M.; writing—review and editing, B.C. and F.D.M.; visualization, B.C. and F.D.M.; supervision, B.C. and F.D.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schowengerdt, R.A. Chapter 9—Thematic Classification. In *Remote Sensing (Third Edition)*; Academic Press: Cambridge, MA, USA, 2007; pp. 387–456. [CrossRef]
2. Jenks, G.F. The Data Model Concept in Statistical Mapping. *Int. Yearb. Cartogr.* **1967**, *7*, 186–190.
3. He, Z. The principle of fuzzy mathematical multilayer synthetic evaluation and its application of quality evaluation in thematic mapping. *Acta Geod. Cartogr. Sin.* **1989**, *18*, 266–276.
4. Ricotta, C. Evaluating the classification accuracy of fuzzy thematic maps with a simple parametric measure. *Int. J. Remote Sens.* **2004**, *25*, 2169–2176. [CrossRef]
5. Morisita, M. Iδ-Index, A Measure of Dispersion of Individuals. *Res. Popul. Ecol.* **1962**, *4*, 1–7. [CrossRef]
6. Taheri, F.; Jafari, H.; Rezaei, M.; Bagheri, R. The use of continuous fuzzy and traditional classification models for groundwater potentiality mapping in areas underlain by granitic hard-rock aquifers. *Environ. Earth Sci.* **2020**, *79*, 91. [CrossRef]
7. Yousefi, M.; Carranza, E.J.M. Prediction–area (P–A) plot and C–A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. *Comput. Geosci.* **2015**, *79*, 69–81. [CrossRef]
8. Cardone, B.; Di Martino, F. A GIS-Based Fuzzy Multiclassification Framework Applied for Spatiotemporal Analysis of Phenomena in Urban Contexts. *Information* **2022**, *13*, 248. [CrossRef]
9. Ruspini, E.H. A new approach to clustering. *Inf. Control* **1969**, *15*, 22–32. [CrossRef]
10. De Luca, A.; Termini, S. Entropy and energy measures of fuzzy sets. In *Advances in Fuzzy Set Theory and Applications*; Gupta, M.M., Ragade, R.K., Yager, R.R., Eds.; North-Holland: Amsterdam, The Netherlands, 1979; pp. 321–338.
11. De Luca, A.; Termini, S. A Definition of Non-probabilistic Entropy in the Setting of Fuzzy Sets Theory. *Inf. Control* **1972**, *20*, 301–312. [CrossRef]
12. Di Martino, F.; Sessa, S. Fuzzy Entropy-Based Spatial hot spot Reliability. *Entropy* **2021**, *23*, 531. [CrossRef] [PubMed]
13. Kaymak, U.; Setnes, M. Fuzzy Clustering with Volume Prototype and Adaptive Cluster Merging. *IEEE Trans. Fuzzy Syst.* **2002**, *10*, 705–712. [CrossRef]
14. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Plenum Press: New York, NY, USA, 1981; p. 272. ISBN 978-0306406713.
15. Bezdek, J.C.; Ehrlich, R.; Full, W. The fuzzy C-means Clustering Algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [CrossRef]
16. Bezdek, J.C.; Pal, S.K. *Fuzzy Models for Pattern Recognition: Methods That Search for Structure in Data*; IEEE Press: New York, NY, USA, 1992; p. 544. ISBN 978-0780304222.
17. Di Martino, F.; Sessa, S. A New Validity Index Based on Fuzzy Energy and Fuzzy Entropy Measures in Fuzzy Clustering Problems. *Entropy* **2020**, *22*, 1200. [CrossRef] [PubMed]

18. Cardone, B.; Di Martino, F. A Novel Fuzzy Entropy-Based Method to Improve the Performance of the Fuzzy C-Means Algorithm. *Electronics* **2020**, *9*, 554. [CrossRef]
19. Aggarwal, M. Bridging the Gap Between Probabilistic and Fuzzy Entropy. *IEEE Trans. Fuzzy Syst.* **2020**, *28*, 2175–2184. [CrossRef]
20. Lee, H.M.; Chen, C.M.; Chen, J.M.; Jou, Y.L. An efficient fuzzy classifier with feature selection based on fuzzy entropy. *IEEE Trans. Syst. Man Cybern. B Cybern.* **2001**, *31*, 426–432. [PubMed]
21. Wang, X.; Dong, C. Improving generalization of fuzzy if–then rules by maximizing fuzzy entropy. *IEEE Trans. Fuzzy Syst.* **2009**, *17*, 556–567. [CrossRef]
22. Raghu, S.; Sriraam, N.; Kumar, G.P.; Hegde, A.S. A novel approach for real-time recognition of epileptic seizures using minimum variance modified fuzzy entropy. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 2612–2621. [CrossRef] [PubMed]
23. Cardone, B.; Di Martino, F. Fuzzy-Based Spatiotemporal Hot Spot Intensity and Propagation—An Application in Crime Analysis. *Electronics* **2022**, *11*, 370. [CrossRef]