

Cluster of Vulnerable Municipalities in Mexico to facilitate the Creation of Coordination Mechanisms in the Humanitarian Responses through Machine Learning Techniques

*Diana Sánchez-Partida*¹ *, Maria Beatriz Bernábe Loranca*² *, Jorge A. Ruiz-Vanoye*³ *, Ricardo A Barrera Camara*⁴

¹ Department of Logistics and Supply Chain Management, Faculty of Engineering, UPAEP University. E-mail[: diana.sanchez@upaep.mx](mailto:diana.sanchez@upaep.mx)

² Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla. E-mail[: beatriz.bernabe@gmail.com](mailto:beatriz.bernabe@gmail.com)

³ Universidad Politecnica de Pachuca, DIIP E-mail jorge@ruizvanoye.com

⁴ Universidad Autónoma del Carmen, Facultad de Ciencias de Información. E-mail rbarrera@pampano.unacar.mx

1 Introduction

Humanitarian Logistics (HL) is the process of planning, implementing, and controlling the efficient, cost-effective flow and storage of goods and materials, as well as related information, from the point of origin to the point of consumption to alleviate the suffering of vulnerable people. It involves the management of supply chains in humanitarian contexts, such as natural disasters, conflicts, or other crises, to ensure that essential goods and services reach those in need in a timely and effective manner (Regis-Hernandez et al. 2021).

The HL encompasses various activities, including procurement, transportation, warehousing, distribution, and information management, to assist affected populations during emergencies. It involves coordinating the efforts of multiple stakeholders,

such as government agencies, non-governmental organizations (NGOs), international organizations, donors and funding agencies, volunteers and community groups, and the private sector, to deliver aid efficiently and effectively to those in crises (Barojas-Payán et al. 2021; Fon-Galvez et al. 2021).

Therefore, Humanitarian Logistics Management (HLM) is a critical factor in humanitarian operations because it enables organizations to respond quickly and effectively to emergencies, optimize the use of resources, coordinate efforts among stakeholders, manage risks, and maximize the impact of aid delivery to vulnerable populations.

Stakeholders are crucial in humanitarian logistics, as they coordinate their expertise, resources, and support. This collaboration is essential for overcoming logistical challenges, maximizing the impact of relief efforts, and ultimately supporting affected populations' well-being during crises (Fon-Galvez et al. 2021).

Fon-Galvez et al. (2021) developed an analysis of a case study about the 19S earthquake in Puebla, Mexico, where the final results showed a lack of coordination between organizations involved in the relief efforts. This lack of coordination led to inefficiencies and duplication of efforts, impacting the overall effectiveness of the response. Also, this analysis concluded the lack of knowledge of cultural needs, lack of information, lack of psychosocial attention, and mismanagement of government resources. These challenges and mismanagement underscore the complexities and difficulties organizations face in supporting those affected by natural disasters, which could lead to delays in the recovery process and hinder the overall effectiveness of the response efforts.

This study delves into the complexity of the situation in Mexico, where various natural phenomena require efficient logistics systems capable of providing aid in a rapid and coordinated manner. The lack of cooperation and coordination between organizations emerges as a latent challenge, potentially resulting in substantial losses of human and material resources in humanitarian logistics. Contingencies arising from natural disasters underline the urgency of optimizing logistics processes, and this study addresses this need through an innovative proposal.

In this work, an approach based on unsupervised learning is proposed, specifically on the medoid participation method to identify and segment highly vulnerable Mexican municipalities to improve final allocations and strengthen the efficiency of logistics systems in emergencies. This article is structured as follows. This introduction frames the importance of why supply chain actors must be coordinated. In the theoretical framework, we can grasp unsupervised learning and the PAM algorithm used in this proposal. In this context, the problem is to explain the concept of vulnerability and how this vulnerability affects the Mexican municipalities. In addition, a background of some similar proposals is described. In the methodology and results, it is possible to see the algorithms applied and the results obtained. Moreover, in the conclusions section, we see the advantages and disadvantages of the algorithm proposed and future research lines.

2 Theoretical Framework: Unsupervised Learning

Unsupervised learning is a paradigm within machine learning in which the model is trained without labels or categories previously assigned to the data.

Unlike supervised learning, where input and output pairs are provided for the model to learn the relationship between them, in unsupervised learning, the algorithm faces unlabeled data and must find patterns, structures, or relationships independently. A prevalent method is clustering, the main objective of which is to organize data into homogeneous groups or "clusters" based on similarities between them. Clustering algorithms seek to identify inherent patterns in the data to group them meaningfully but are not provided with information about the categories a priori.

In general, unsupervised learning algorithms are valuable for discovering patterns in unlabeled data. However, they may be less accurate than supervised algorithms in some applications because they do not use labeled training data. Despite these possible drawbacks, unsupervised algorithms remain an essential tool in machine learning and have a wide range of practical applications, such as in logistics processes, to identify the distribution center and its potential customers.

Partitioning is expected to get trapped in local optima, such as traditional k-means; therefore, many efforts have been made to escape from them with excellent results. However, because the grouping obeys an optimization model and is solved with combinatorial algorithms to reach the global optimum, at this point, a medoid partitioning algorithm has been developed that achieves a global optimum for grouping our INNEC data generated by the National Institute of Ecology and Climate Change

(INECC) retrieved on 2017 (INECC, 2017). We need the coordinates and the distances, which were obtained through an Automated Data Collection (ADC) in real-time using Geocode Google's API.

The algorithm in question, called PAM, is based on the philosophy of partitioning by medoids and respects the structure of traditional partitioning in data entry: the distance matrix and the number of groups (K), in such a way that it is generated in the answer the value of the objective function, the cost in solution time and the analytical structure of the groups indicating the objects that belong to each group and its centroid, in this case medoid as a center (Loranca Bernabe, 2022).

The PAM algorithm has been compared with the P-Median to check the efficiency of the global optimum, a model widely known in the literature as a combinatorial optimization problem that is used in particular in logistics to find the optimal locations of P "medians" or distribution centers to minimize the total distance between customer points to be assigned to their respective medians. This problem is commonly used in resource allocation, facility design, and location logistics contexts. In the P-median problem, we seek to find a set of P locations (where P is an integer) to minimize the sum of the distances between each client or destination and the closest location (Loranca, 2014).

The PAM model implemented in the tests of this work is shown below:

PAM is a non-hierarchical method where the data set is partitioned into a previously specified number of k conglomerates (groups or clusters). Observations are iteratively assigned to the clusters until some stopping criterion (function to be optimized) is satisfied, such as the sum of squares within the clusters is minimal.

As already repeated, the PAM function is based on the search for k representative objects called medoids among the objects in the data set. These medoids are calculated such that the total dissimilarity of all objects towards their nearest medoid is minimal: that is, the goal is to find a subset ${m_1 \dots m_k} \subset {1 \dots n}$ which minimizes the objective function:

$$
\sum_{i=1}^n min_{t=1,\dots,k} d(i, m_t)
$$

Each object is assigned to the corresponding cluster of the nearest centroid. That is, the object i is placed in the cluster v_i when the medoids m_{vi} is closer to i than any other medoid m_{wi} , o:

$$
d(i, m_{vi}) \leq d(i, m_w)
$$
para todo $w = 1, ..., k$

Currently, the PAM algorithm consists of two steps:

1.- Build the initial centroids

 m_1 is the object with the smallest $\sum_{i=1}^n d(i, m_1)$

. .

 m_k decrease target (1) as much as possible

2.- Exchange

Repeat until convergence is achieved and all pairs of objects are considered (i, j) with $i \in \{m_1, ..., m_k\}$ and $j \in \{m_1, ..., m_k\}$ and make the exchange $i \leftrightarrow j$ for anyone who decreases the target more.

The PAM objective function depends on the dissimilarities between the objects; therefore, this function only needs the distance matrix as input.

On the other hand, the P-Median problem considers the following: It is required to partition a finite set of objects into precisely *p* groups. Each of said groups will be characterized by one of the objects, which is selected as the median of the group and the subset of objects assigned to said median. A distance is specified for each pair of objects, and it is required to minimize the sum of distances between the objects and the medians to which they are assigned.

Be $N = \{1, \ldots, n\}$ the set of objects.

For each (i, j) , $i \in N$, $j \in N$, where d_{ij} means the distance (similarity) between objects i and j; given the number p that denotes the number of groups, it is necessary to partition the set *N* into *p* disjoint subsets, for example, $N = \bigcup_{k=1}^{p} N_k y N_r \cap N_s = \emptyset$, for all $r, s \in \{1, ..., p\}$, $r \neq s$. Next, we consider the following mathematical programming model for the problem. The following decision variables are defined:

$$
y_i = \left\{ \begin{array}{cl} 1, & \text{if object } i \text{ is selected as median,} \\ 0, & \text{otherwise.} \end{array} \right.
$$

and
 $x_{ij} = \begin{cases} 1, \\ 0, \end{cases}$ If object j is assigned to median,i,
otherwise.

The problem can be modeled as follows:

$$
\min \sum_{i \in N} \sum_{j \in N} d_{ij} x_{ij} \tag{1}
$$

$$
bound\ to\ \sum_{i\in N} x_{ij} = 1\ \ j\in N\tag{2}
$$

$$
\sum_{i \in N} y_i = p \tag{3}
$$

 (4)

 (5)

 (6)

$$
x_{ij} \le y_i \quad i \in N, j \in N
$$

$$
x_{ij} \leq \{0,1\} \quad i \in N, j \in N \tag{3}
$$

$$
y_i \le \{0,1\} \quad i \in N \tag{0}
$$

Constraints (2) ensure that each object is assigned to one of the medians. Constraint (3) ensures that *p* objects are selected as medians. Finally, constraints (4) to (6) ensure that objects can only be assigned to selected medians (Church, 2008).

3 Context of the Problem: Vulnerable Municipalities in Mexico with INECC data

Vulnerability refers to the susceptibility of individuals, communities, or systems to harm or damage from external stresses, shocks, or hazards. In humanitarian logistics and disaster management, vulnerability is a critical concept that helps understand and address the risks faced by populations in crises. Here are some critical points about vulnerability following (Sánchez-Partida et al. 2022):

- Exposure to Risk: Vulnerability is often associated with the exposure of a system or community to potential risks, such as natural disasters, conflicts, or health emergencies.
- Sensitivity: Vulnerability also reflects a system's or population's sensitivity to the impacts of these risks. Socioeconomic status, infrastructure, and resource access can influence hazard sensitivity.
- Capacity to Adapt: The capacity of a system or community to adapt, cope, and recover from adverse events is another aspect of vulnerability. More substantial adaptive capacities can reduce vulnerability and enhance resilience.
- Multi-dimensional: Vulnerability is a multi-dimensional concept encompassing social, economic, environmental, and institutional factors that influence the ability of individuals or communities to withstand and recover from shocks.
- Dynamic Nature: Vulnerability is dynamic and can change over time due to various factors such as climate change, urbanization, or socio-political developments. Understanding these dynamics is essential for effective risk management.
- Assessment and Mitigation: Assessing vulnerability helps identify at-risk populations or areas, enabling targeted interventions and mitigation strategies to reduce the impact of disasters and emergencies.
- Policy Implications: Recognizing vulnerability informs the development of policies, programs, and interventions aimed at building resilience, reducing risks, and enhancing the overall well-being of vulnerable populations.

In summary, vulnerability is a complex concept involving individual's and communities' exposure, sensitivity, and adaptive capacity to risks and hazards. Addressing vulnerability through coordinated efforts in humanitarian logistics and disaster management is essential for ensuring effective response and recovery during crises.

The research proposes a partitioning approach to identify and segment vulnerable Mexican municipalities. This approach involves applying statistical techniques and decision-making models, such as the random decision-based model, to improve final allocations. This study is critical because the frequency of natural disasters in Mexico is evident, which requires efficient logistics systems to offer immediate support. Without solutions like the one in this document, the lack of logistics organization can result in significant losses of human and material resources in humanitarian logistics. For this purpose, the INECC data was used; the National Institute of Ecology and Climate Change obtained it. It is a government agency responsible for environmental protection, climate change research, and developing policies and strategies related to ecological sustainability and climate resilience. The INECC data contains climate change data, environmental indicators, vulnerability assessments, policy research, scientific studies, and reports. As a result of one study of this institution, the Special Climate Change Program 2014–2018 contains a list of 319 municipalities with high vulnerability (PECC 2014–2018).

3.1 Background

A specific example of the need for coordination is presented, highlighting the earthquake and tsunami in South Asia in 2004, which led to a massive humanitarian response. Despite warnings from aid agencies against sending unsolicited items, many well-intentioned donors sent inappropriate supplies that created significant logistical problems. It highlights the importance of planning and efficient coordination to avoid obstacles in distributing humanitarian aid. An earthquake in Gujarat, India, in 2001 is mentioned to highlight the challenges of working in politically sensitive areas with a military presence. Effective coordination was essential to deliver supplies despite logistical difficulties.

The study by Sánchez-Partida et al. (2022) uses statistical methods, specifically clustering, to analyze the vulnerability of Mexican municipalities to meteorological events. Firstly, vulnerability data was identified and collected, and secondly, the following clustering techniques were used: the NbClust, the Cross-Entropy Clustering (CEC), Hierarchical and Partitioning Statistical Methods like hierarchical clustering and k-means clustering, and Cluster Analysis Indexes. By employing a combination of these methods and indexes, the study aimed to identify and segment vulnerable zones in Mexico based on factors such as climate change vulnerability and disaster risk. Using diverse clustering methods and indexes helps provide comprehensive data analysis and supports informed decision-making for disaster response and risk management strategies. The importance of collaborative networks in disaster planning and response is highlighted, allowing the participation of multiple actors from the public, private, and non-profit sectors. The effectiveness of these networks lies in their flexibility and ability to adapt to provide solutions in circumstances of high complexity and uncertainty. (The Concise Encyclopedia of Statistics | WorldCat.org, 2006). At the local level, horizontal collaborations between local governments and organizations are crucial to strengthening disaster response capacity. The study proposes a clustering analysis to accurately segment vulnerable areas, aiming to support national strategic decisions to plan and coordinate humanitarian aid at the local level. The need to improve the capacity of local governments to generate collaborative networks is highlighted and focuses specifically on horizontal collaborations between local governments (Sánchez-Partida et al. 2022).

4 Methodology Applied and Results: Tests with the INECC Matrix

In this section, tests were developed with the integrated INECC Matrix of 319 elements. For comparison purposes, the Brunch and Bound (B&B) method on Lingo software was used for P-median and the PAM software developed by the authors since PAM and P-median are equivalent in the results they yield. Medians are identified as centroids with the P-Median, while in PAM, we will use medians. The results achieved were recorded, such as the centroids, the computational cost, the value of the objective function, and the objects that belong to each group (see Table 1).

	P-Median with $(B&B)$			PAM		
K	COST	TIME IN SECONDS	MEDIANS	COST	IN TIME SECONDS	MEDOIDS
2	1209.299	24.82	61.273	1209.299	θ	273,61
	703.4600	23.56	41,236,270,319	703.4601	θ	41,236,270,319
5	617.9420	23.19	44,236,270,295,319	617.9419	θ	44,236,270,319,295
9	411.4370	59.79	33, 68, 107, 136, 177, 18 4,217,295,311	412.745	Ω	22, 111, 187, 144, 129, 33, 230, 295, 217

Table 1. shows the results achieved of the execution test

Table 2 below reveals, for each K, the centroids that coincide and where they do not coincide to see the similarities between these models.

Table 3. P-Median with its medians and objects

As can be seen, in more than 50%, the medoids and medians are the same; in the case of K less than 5, they coincide 100%, although their costs per configuration vary minimally, and the parity of results is notable. The following Figure 1 indicates the differences in the cost of the objective function.

Fig. 1. Differences in the cost of the objective function

P-Median's cost was the best but with a minimal difference. The following Figure 2 shows the variation in the computational cost of Lingo and PAM.

Fig. 2. Computational Cost Difference.

As we can see, the variation in seconds of the P-Median compared to PAM is very high. While PAM takes 0 seconds for K=2, Lingo takes 24.82 seconds to execute to reach the global optimum. Although the computational cost is much higher for a more significant amount of data, in the case of the INECC matrix, the computational cost did not exceed 59.79 seconds for Lingo with the P-Median and only 4 seconds with PAM. In terms of computational cost, PAM is much more efficient.

To better understand our data, a median with its respective objects and a medoid with its respective clusters were chosen, and both coincided. The median and medoid 58 were chosen with the objects 55, 57, 60, 63, 64,65,67 that coincide with the clusters of the medoid.

With the latitude and longitude data, we look for which municipalities they were with the Google Earth tool, as detailed in Table 5. On the other hand, the median or medoid 58 corresponding to the municipality of Bocoyna Municipality, Chihuahua, Mexico, will serve the other municipalities described in Table 5.

5 Conclusions

Disaster management literature has considered "Collaborative Public Management" a suitable approach to respond effectively to natural disasters. This approach involves facilitating and operating in multi-organizational arrangements to address problems that individual organizations cannot solve. Unlike the traditional top-down hierarchical governance system, collaborative public management is formed through network structures that facilitate vertical and horizontal cooperation between organizations.

In conclusion, the research proposes an innovative approach using the clustering method to identify and segment vulnerable Mexican municipalities to understand better vulnerable municipalities and how they can be managed in times of crisis. Thus helping actors in the humanitarian supply chain make more informed, structured decisions.

This Project's scope is to apply any of the PAM and P-median models since they are equivalent. However, PAM is a better proposal given its scope in computing time.

On the other hand, it is necessary to incorporate metaheuristic methods for large instances, both for this problem and for similar problems.

References

Barojas-Payán, E., Sánchez-Partida, D., Gibaja-Romero, D.E., Martínez-Flores, J.L., Cabrera-Rios, M. (2021). Optimization Model to Locate Pre-positioned Warehouses and Establish Humanitarian Aid Inventory Levels. In: Disaster Risk Reduction in Mexico. Springer, Cham. https://doi.org/10.1007/978-3-030-67295-9_8

Church, R. L. (2008). BEAMR is an exact and approximate model for the p-median problem. *Computers and Operation Research*, 35(2), 417- 26. https://doi.org/10.1016/j.cor.2006.03.006

Fon-Galvez, MJ., Sánchez-Partida, D., Gibaja-Romero, DE., Caballero-Morales, SO. (2021). 19S Earthquake in Puebla, Mexico: Intervention of the Different Actors in Humanitarian Aid. In: Disaster Risk Reduction in Mexico. Springer, Cham. https://doi.org/10.1007/978-3-030- 67295-9_7

¿Qué es el aprendizaje no supervisado? (n.d.). TIBCO Software. https://www.tibco.com/es/reference-center/what-is-unsupervised-learning

¿Qué es el aprendizaje no supervisado? | IBM. (n.d.). https://www.ibm.com/mx-es/topics/unsupervised-learning

Loranca, M. B. B., Velázquez, R. G., & ME, A. (2015). The p-median problem: A tabu search approximation proposal applied to districts. *J Math Syst Sci*, 5(3).

National Institute of Ecology and Climate Change (INECC). Retrieved July 1st, 2017, from https://www.gob.mx/inecc

PDF visualización del fichero Loranca Bernábe, 2022-Libro Introducción al Particionamiento.pdf. (n.d.). https://librosutp.com/index.php/editorialutp/catalog/view/96/79/432

Regis-Hernández, F., Mora-Vargas, J., Ruíz, A., Sánchez-Partida, D. (2021). Reliable Network Design: Case Study. In: Disaster Risk Reduction in Mexico. Springer, Cham. https://doi.org/10.1007/978-3-030-67295-9_12

Sánchez-Partida, D., Mora-Vargas, J., Smith, N.R., Castillo-Villar, F. (2022). Clustering of Highly Vulnerable Mexican Municipalities to Develop Humanitarian Public Policies. In: Regis-Hernández, F., Mora-Vargas, J., Sánchez-Partida, D., Ruiz, A. (eds) Humanitarian Logistics from the Disaster Risk Reduction Perspective . Springer, Cham. https://doi.org/10.1007/978-3-030-90877-5_2

Special Climate Change Program (PECC) 2014–2018. Retrieved July 1st, 2017, from http://dof.gob.mx/nota_detalle.php?codigo=5342492&fecha=28/04/2014

The concise encyclopedia of statistics | WorldCat.org. (2006). https://search.worldcat.org/es/title/1017267126