www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 17(3), 2026, 182-189. ISSN: 2007-1558. https://doi.org/ 10.61467/2007.1558.2026.v17i3.1358

Technology Applied to Identifying Areas without Access to Drinking Water: Challenges and Opportunities

Miguel Angel Torres González [1], *Rosalía América Juárez Aguirre* [1], *Manuel Antonio Vivanco Cabecera* [1], *Sonia Gómez Gómez* [1]

[1] Universidad Politécnica de Pachuca

Email address(es): *mtorres@upp.edu.mx, america@upp.edu.mx, vivanco@upp.edu.mx, sonia@upp.edu.mx*

✉Corresponding author: *america@upp.edu.mx*

Abstract. Safe access to drinking water remains a critical priority for governments and international organizations. The World Health Organization estimates that more than 2.2 billion people lack safely managed water services, which has serious consequences for health and economic development. In this context, different technologies, and especially artificial intelligence (AI), are emerging as innovative tools that can help identify, more accurately and quickly, areas with inadequate drinking water service coverage. (WHO & UNICEF, 2021). This study analyzes the applications of machine learning and statistical techniques to identify sectors of society with a lack of drinking water supply. It also describes methodological approaches that integrate sociodemographic, environmental, and water infrastructure data with predictive models based on neural networks and decision trees. (Ye et al., 2026). Finally, the challenges associated with data quality, the replicability of models in different contexts, and the need for public policies that promote the responsible adoption of these cutting-edge technologies are discussed (Sang et al., 2023**)**.

Keywords: Areas without drinking water, Applied technology, Main components, Artificial Intelligence, Public policies for drinking water supply.

Article information

Received: April 9, 2026

Accepted: May 9, 2026

1. Introduction

The global problem of drinking water shortages mainly affects rural populations and informal settlements.

Bain et al, (2014) They conducted a comprehensive assessment of the risk of fecal contamination in drinking water, based on a systematic review of field studies and surveys, finding an inequality between the population that reports having access to a water connection to the distribution network (domestic connection), protected well, or safe community source, compared to a proportion of that same population that consumes water contaminated with fecal matter. They apply the concept of a water gap in situations where there is not only a lack of infrastructure, but also a failure in the microbiological quality of water that is supposed to be safe, which in reality exposes millions of people to gastrointestinal diseases.

Research shows that the mere availability of infrastructure does not guarantee safety, and therefore the water gap reflects the distance between reported access and the true quality of the water consumed.

Aitken et al., 2020, argues that traditional monitoring methodologies, based on censuses and field visits, are costly, slow, and limited in spatial coverage. For this reason, AI has begun to be used to automate the identification of water scarcity patterns by analyzing large volumes of geospatial and sociodemographic data.

The objective of this article is to analyze the different technologies used to detect areas without access to drinking water, propose a replicable methodological model, and reflect on the ethical and political implications of its adoption. (Fatemeh et al., 2025)

2. Problem statement and state of the art

Over the past decade, the search for more effective methods to identify populations without safe access to drinking water has led to significant research combining satellite data sources, traditional censuses, and artificial intelligence techniques. (Ye et al., 2026).

The work of Bain et al. (2014) marked an important starting point by showing that water coverage statistics published by governments, especially those in underdeveloped countries, underestimate the actual magnitude of exposure to microbiological risks. In their study, they demonstrated that a considerable proportion of the supply systems classified as improved were in fact contaminated, suggesting that the gap between reported access and actual quality does not match census records, casting doubt on the reliability of methodologies based exclusively on household surveys and field visits.

Several studies have proposed the use of satellite imagery as a complementary resource to reduce the spatial and temporal coverage limitations of censuses. These technologies allow for the generation of more frequently updated data with higher spatial resolution, which is especially useful in contexts where censuses are infrequent or outdated, according to Aitken, Rivera, McIntyre y Siegel (2020), “satellite imagery offers an unprecedented opportunity to complement and improve traditional census-based estimates of access to drinking water by providing continuous spatial data that can be updated more frequently”.

In their study, Aitken et al. (2020) They conducted one of the most comprehensive analyses by developing a machine learning model that integrated multispectral images from Landsat and Sentinel-2 with georeferenced water source records. their study confirmed that supervised classification algorithms, such as Random Forest, could identify patterns of water vulnerability with a level of accuracy that surpassed that of traditional methods, particularly in rural and peri-urban regions where surveys are more costly and difficult to implement.

In this context, Ye, Dong, McCright & Gasteyer (2026),conducted a critical review of machine learning applications in water resource management. these authors highlighted that artificial intelligence techniques not only provide advantages in identifying communities without access to drinking water services, but also allow for the modeling of prospective water availability scenarios in the face of climate change. According to their analysis, the combination of satellite data and sociodemographic variables improves predictive capacity and the prioritization of public investments, although they emphasize the need for databases of sufficient quality.

Sang, Hou, Wang & Ding (2023) they explored the use of convolutional neural networks (CNN) to predict access to drinking water in developing regions. Their research showed that these types of models achieve over 90% accuracy in classifying areas at high risk of water scarcity, provided that high-resolution satellite images and geolocated data are available. However, they note that these systems require advanced technological infrastructure and technical capabilities that are not always present in all countries, particularly in developing countries.

The identification of populations without safe access to drinking water has evolved considerably over the last two decades. From the mid-2000s onwards, approaches began to take hold that recognized the limitations of traditional censuses and field visits, particularly in contexts of high geographical dispersion. (Kunz et al., 2007), these approaches highlighted that coverage indicators often overestimated effective access to water, as they did not consider microbiological quality or the continuous availability of the resource.

With advances in remote sensing and the accessibility of medium- and high-resolution satellite images, there has been a proliferation of studies proposing the integration of remote data with census records and spatial models. MacDonald & Calow (2009) they pointed out that the use of satellite imagery to estimate the location of water points and assess drought patterns could provide crucial input for water planning in sub-Saharan Africa and South Asia. This approach was pioneering in demonstrating that combining remote sensing with geographic information systems (GIS) could reduce data collection costs and time.

Currently, the incorporation of artificial intelligence techniques has amplified the capabilities for detecting areas with high marginalization of drinking water service, Vervoort et al. (2016) they documented that supervised learning models, such as Random Forest and artificial neural networks, began to be used experimentally to estimate groundwater availability and the vulnerability of rural communities. They emphasized that the integration of satellite data, climate variables, and socioeconomic factors increased the accuracy of the models, although it also posed challenges related to information quality and database interoperability.

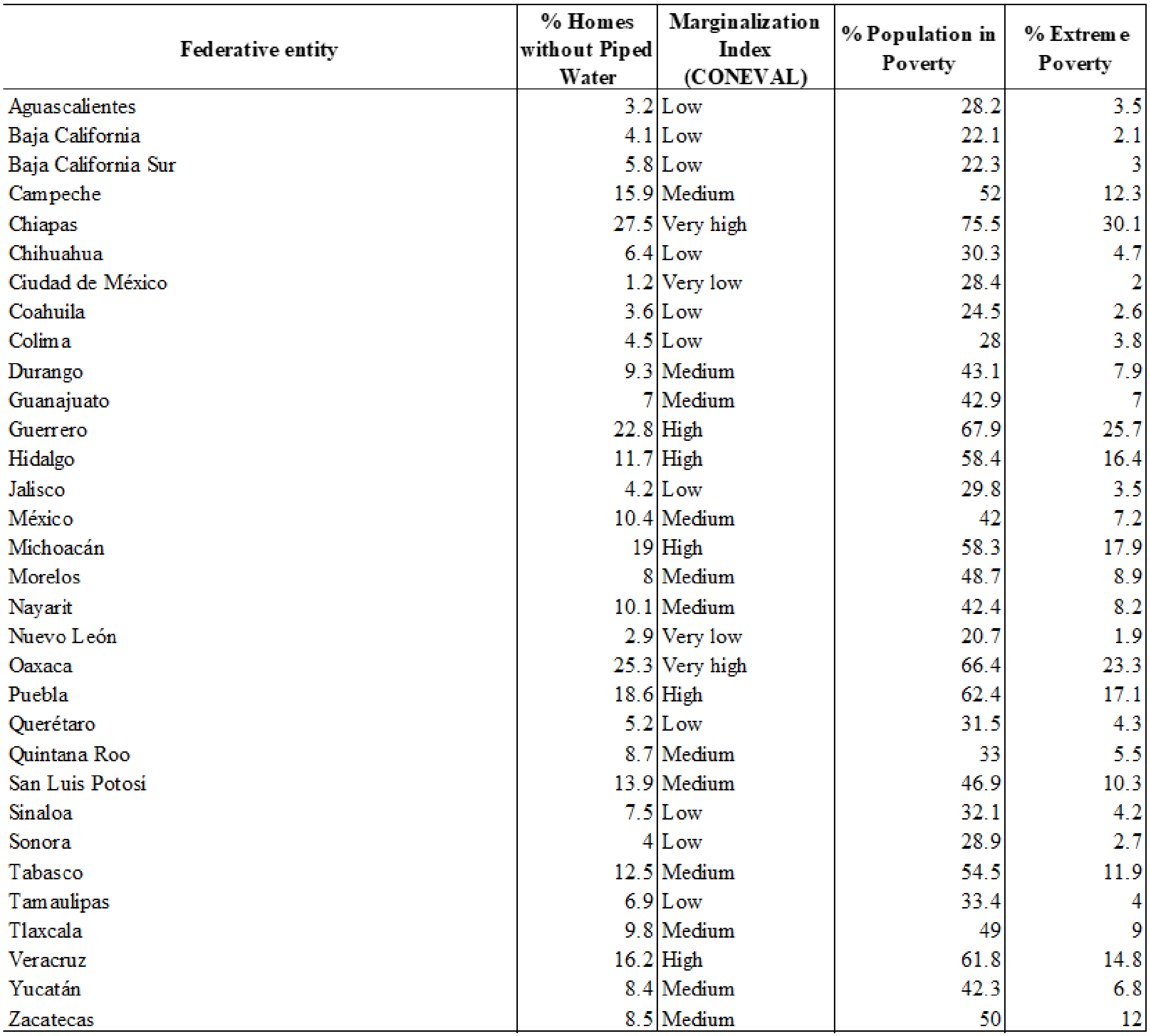
Currently, other research has reinforced the relevance of remote data combined with advanced algorithms, Jiaxin et al. (2020) They highlighted that new-generation satellites offer images with higher temporal frequency and spatial resolution, which facilitates monitoring changes in land use and the evolution of water bodies. By applying deep learning techniques, the authors were able to identify patterns of deterioration in surface sources that were not traditionally captured by field surveys. On their part, Layth, Najah, Ala Hassan, Farhan Lafta, Ahmed and Al-Dujaild (2025) addressed the relationship between water scarcity and social factors, proposing that artificial intelligence can also contribute to strategic decision-making and the targeting of public interventions.

The convergence of satellite imagery, censuses, and statistical and artificial intelligence algorithms represents a substantial methodological advance over conventional monitoring schemes. However, its effective adoption depends on the availability of technological infrastructure, the training of local teams, and the integration of multisectoral data (Vervoort et al., 2016; Jiaxin et al., 2020).

2.1 The case of Mexico

Table 1 shows the information proposed for constructing the water vulnerability index using the techniques proposed in this study.

Table 1. Information for constructing the Water Vulnerability Index



Federative entity	% Homes without Piped Water	Marginalization Index (CONEVAL)	% Population in Poverty	% Extreme Poverty
Aguascalientes	3.2	Low	28.2	3.5
Baja California	4.1	Low	22.1	2.1
Baja California Sur	5.8	Low	22.3	3
Campeche	15.9	Medium	52	12.3
Chiapas	27.5	Very high	75.5	30.1
Chihuahua	6.4	Low	30.3	4.7
Ciudad de México	1.2	Very low	28.4	2
Coahuila	3.6	Low	24.5	2.6
Colima	4.5	Low	28	3.8
Durango	9.3	Medium	43.1	7.9
Guanajuato	7	Medium	42.9	7
Guerrero	22.8	High	67.9	25.7
Hidalgo	11.7	High	58.4	16.4
Jalisco	4.2	Low	29.8	3.5
México	10.4	Medium	42	7.2
Michoacán	19	High	58.3	17.9
Morelos	8	Medium	48.7	8.9
Nayarit	10.1	Medium	42.4	8.2
Nuevo León	2.9	Very low	20.7	1.9
Oaxaca	25.3	Very high	66.4	23.3
Puebla	18.6	High	62.4	17.1
Querétaro	5.2	Low	31.5	4.3
Quintana Roo	8.7	Medium	33	5.5
San Luis Potosí	13.9	Medium	46.9	10.3
Sinaloa	7.5	Low	32.1	4.2
Sonora	4	Low	28.9	2.7
Tabasco	12.5	Medium	54.5	11.9
Tamaulipas	6.9	Low	33.4	4
Tlaxcala	9.8	Medium	49	9
Veracruz	16.2	High	61.8	14.8
Yucatán	8.4	Medium	42.3	6.8
Zacatecas	8.5	Medium	50	12

In Table 1, we can see that Mexico City reports the lowest percentage of homes without drinking water service (1.2 percent). Chiapas, on the other hand, has the highest percentage at 27.5 percent. In terms of poverty, Chiapas has the highest indicator at 75.5 percent, while Nuevo León reports 20.7 percent in this case. In terms of extreme poverty, Chiapas also has the worst scenario at 30.1 percent, while Nuevo León reports 1.9 percent of the population in extreme poverty.

Source: Prepared by the authors

3. The Model

Principal Component Analysis (PCA) is proposed for the treatment of the Mexican case, which allows for the reduction of informational redundancy between highly correlated variables, the identification of latent patterns in the data structure, and the optimization of the performance of subsequent predictive algorithms. PCA made it possible to transform an initial set of quantitative variables (related to sociodemographic characteristics, access to infrastructure, housing conditions, and geographic location) into an indicator that preserved as much of the total variance as possible (Jolliffe & Cadima, 2016).

3.1 Principal component analysis

Mathematically, Principal Component Analysis (***PCA***) is an orthogonal transformation that converts a set of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The first principal component (***PC1***) is defined as the linear combination of the original variables that captures the largest proportion of the total variance of the data:

$$PC1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p$$

Where the coefficients a_{1j} (known as loadings) are chosen in such a way that the variance of ***PC1*** is maximized, subject to the constraint that the sum of their squares equals unity (Jolliffe & Cadima, 2016). By using the first component as an index, it is guaranteed that the resulting metric represents the most robust 'signal' of the data structure, discarding the statistical noise from the minor components.

Unlike equal weighting methods, Principal Component Analysis (***PCA***) assigns weights (loadings) endogenously. The first principal component (***PC1***) is extracted such that it maximizes the explained variance, acting as the best statistical summary of the latent phenomenon. As noted by Müller and Schmidt (2022), this approach ensures that variables contributing the most variability to the system have a proportional impact on the final index, thereby enhancing the discriminatory power between the observation units.

Principal Component Analysis (***PCA***) enables the reduction of a system of ***p*** variables into a significantly simpler structure without sacrificing interpretability. Recently, its efficacy has been demonstrated in the development of social vulnerability and sustainability indices, where it identifies patterns that simple additive methods overlook (Dahal, 2021; Singh & Patel, 2021). This underscores that this statistical technique remains the standard of choice for transforming highly correlated variables into synthetic indicators with minimal loss of variance.

4. Model results

The water vulnerability index for the states of Mexico can be seen in table 2. The criterion for selecting the states with the greatest lack of drinking water service is identified with those states that have a “PCA index” with the highest factor load.

The entities that meet the selection criteria established according to the methodology are Chiapas, Guerrero, Michoacán, Oaxaca, Puebla, and Veracruz, which, for our study, represent the states of the Mexican Republic that are most behind in terms of drinking water coverage.

Table 2. Water vulnerability index

Federative entity	% Homes without Piped Water	Marginalization Index (CONEVAL)	% Population in Poverty	% Extreme Poverty	Index PCA
Aguascalientes	3.2	Low	28.2	3.5	-1.5859
Baja California	4.1	Low	22.1	2.1	-1.8515
Baja California Sur	5.8	Low	22.3	3	-1.6263
Campeche	15.9	Medium	52	12.3	1.1075
Chiapas	27.5	Very high	75.5	30.1	4.4117
Chihuahua	6.4	Low	30.3	4.7	-1.1361
Ciudad de México	1.2	Very low	28.4	2	-1.8695
Coahuila	3.6	Low	24.5	2.6	-1.7636
Colima	4.5	Low	28	3.8	-1.4580
Durango	9.3	Medium	43.1	7.9	-0.1469
Guanajuato	7	Medium	42.9	7	-0.4236
Guerrero	22.8	High	67.9	25.7	3.3697
Hidalgo	11.7	High	58.4	16.4	1.3161
Jalisco	4.2	Low	29.8	3.5	-1.4394
México	10.4	Medium	42	7.2	-0.1500
Michoacán	19	High	58.3	17.9	2.0584
Morelos	8	Medium	48.7	8.9	0.0334
Nayarit	10.1	Medium	42.4	8.2	-0.0809
Nuevo León	2.9	Very low	20.7	1.9	-2.0235
Oaxaca	25.3	Very high	66.4	23.3	3.3363
Puebla	18.6	High	62.4	17.1	2.1158
Querétaro	5.2	Low	31.5	4.3	-1.2254
Quintana Roo	8.7	Medium	33	5.5	-0.7725
San Luis Potosi	13.9	Medium	46.9	10.3	0.5831
Sinaloa	7.5	Low	32.1	4.2	-1.0132
Sonora	4	Low	28.9	2.7	-1.5544
Tabasco	12.5	Medium	54.5	11.9	0.8785
Tamaulipas	6.9	Low	33.4	4	-1.0313
Tlaxcala	9.8	Medium	49	9	0.2073
Veracruz	16.2	High	61.8	14.8	1.7039
Yucatán	8.4	Medium	42.3	6.8	-0.3421
Zacatecas	8.5	Medium	50	12	0.3725

Source: Prepared by the authors

The implementation of the principal component variable reduction technique in Orange AI facilitated the visual interpretation of the main axes, the detection of outliers, and the preliminary grouping of regions with similar conditions, which was key to identifying states with a high probability of lacking drinking water service.

5. Proposed public policies

- Based on the results obtained, the following public policies are proposed for the use of technology in detecting areas lacking drinking water:
- Establish a national strategy that coordinates government, academic, and private efforts for the use of AI in planning, detecting, and mitigating areas without access to drinking water.
- Implement a technical entity specializing in the consolidation, processing, and analysis of census, climate, geospatial, and satellite data using AI, with the aim of identifying critical areas of water scarcity.
- Allocate budgetary resources within programs to promote projects for detecting and predicting water service deficiencies using AI, prioritizing marginalized areas.
- Promote agreements between state governments, technology companies, and universities to implement pilot projects using technology based on statistical models and AI in municipalities with high water vulnerability.

- Incorporate predictive analysis and automated visualization modules into information systems in locations with inadequate drinking water supplies, in order to generate real-time maps that identify and prioritize areas without water supplies.
- Establish legal and ethical guidelines for the use of statistical methods and AI in public services, considering algorithmic transparency, data protection, and equity in the distribution of water resources.
- Implement specialized training programs in statistical techniques and AI applied to water for state and municipal officials, as well as for community committees.
- Develop mobile platforms that enable citizens to report water supply failures, provide feedback on AI models, and participate in decision-making processes.

6. Conclusions

Research shows that the application of technology for the efficient identification of areas without access to drinking water should be considered as a tool that allows decision-makers to better allocate resources to reduce the gap in drinking water service coverage in marginalized areas. (Gelila et al., 2025)

Taken together, these findings support the argument that the use of statistics, statistical methods, geographic information systems, and artificial intelligence constitutes an indispensable methodological evolution for overcoming the limitations of conventional monitoring strategies. However, the bibliographic references consulted agree that their successful implementation depends both on the availability of open and updated data and on the strengthening of local capacities in the management of advanced digital tools.

It is recommended that government agencies strengthen the generation of high-quality data and promote inter-institutional collaboration strategies. It is also essential to establish regulatory frameworks that guarantee the ethical use of information (Fatemeh et al., 2025)

However, challenges remain, such as the need to improve information banks, the legal framework for the use of information, training for managers of the various agencies in the drinking water sector in statistical and AI techniques. On the other hand, the adoption of predictive models raises ethical questions about privacy and the risk of stigmatization of communities identified as priorities. (Sang et al., 2023)

It is emphasized that artificial intelligence and statistical prediction methods do not replace traditional mechanisms of community participation, but rather complement them, optimizing government decision-making (Ye et al., 2026)

References

Abdulameer, L., Al Maimuri, N. M. L., Nama, A. H., Rashid, F. L., & Al-Dujaili, A. N. (2025). The role of artificial intelligence in managing sustainable water resources: A review of smart solution implementations. *Water Conservation & Management, 9*(2), 181–191. https://doi.org/10.26480/wcm.02.2025.281.291

Bain, R., Cronk, R., Hossain, R., Bonjour, S., Onda, K., Wright, J., Yang, H., Slaymaker, T., Hunter, P., Prüss-Ustün, A., Bartram, J., & Fewtrell, L. (2014). Global assessment of exposure to faecal contamination through drinking water based on a systematic review. *Tropical Medicine & International Health, 19*(8), 917–927. https://doi.org/10.1111/tmi.12334

Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., & Zupan, B. (2013). Orange: Data mining toolbox in Python. *Journal of Machine Learning Research, 14*, 2349–2353.

Javanbakht-Sheikhahmad, F., Rostami, F., Azari, A., & Veisi, H. (2025). Mapping and managing water poverty indicators: A spatio-temporal analysis. *Environmental and Sustainability Indicators, 28*, Article 101005. https://doi.org/10.1016/j.indic.2025.101005

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2065), Article 20150202. https://doi.org/10.1098/rsta.2015.0202

Kunz, R., Eisenberg, J. N. S., Hubbard, A., & Colford, J. M. (2007). The contribution of epidemiology to risk assessment: Drinking water and intestinal protozoa. *Risk Analysis, 27*(2), 355–374. https://doi.org/10.1111/j.1539-6924.2007.00887.x

Li, J., Ma, R., Cao, Z., Xue, K., Xiong, J., Hu, M., & Feng, X. (2022). Satellite Detection of Surface Water Extent: A Review of Methodology. *Water, 14*(7), Article 1148. https://doi.org/10.3390/w14071148

MacDonald, A. M., & Calow, R. C. (2009). Developing groundwater for secure rural water supplies in Africa. *Desalination, 248*(1–3), 546–556. https://doi.org/10.1016/j.desal.2008.05.100

Sang, J., Hou, B., Wang, H., & Ding, X. (2023). Prediction of water resources change trend in the Three Gorges Reservoir Area under future climate change. *Journal of Hydrology, 617*, Article 128881. https://doi.org/10.1016/j.jhydrol.2022.128881

World Health Organization & United Nations Children's Fund. (2021). *Progress on household drinking water, sanitation and hygiene 2000–2020: Five years into the SDGs*. WHO/UNICEF Joint Monitoring Programme. https://www.who.int/publications/i/item/9789240030848

Ye, L., Dong, Q., McCright, A., & Gasteyer, S. (2026). An Innovative Approach to Predict Drinking Water Risks Using System, Community, and Regulatory Characteristics. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 63*. https://doi.org/10.1177/00469580251411440

Yitageasu, G., Chekole Adane, K., Hailu Tesfaye, A., Gizaw, Z., Kifle, T., & Demoze, L. (2025). Geospatial analysis of limited access to drinking water services in East African countries: Insights from recent demographic and health surveys. *Frontiers in Water, 7*, Article 1487795. https://doi.org/10.3389/frwa.2025.1487795