Using grey clustering to evaluate nitrogen pollution in estuaries with limited data

Regina Temino-Boes a,⁎, Rabindranarth Romero-Lopez b, Sara Patricia Ibarra-Zavaleta c, Inmaculada Romero a

a Instituto de Ingeniería del Agua y del Medio Ambiente, Universitat Politècnica de València, Camino de Vera s/n, Valencia 46022, Spain. b Departamento de Ingeniería Civil, Universidad Veracruzana, Lomas del Estadio s/n, Xalapa 91000, Mexico c Laboratoire Hydrologique, Climat et Changement, École de Technologie Supérieure, 1100 Rue Notre-Dame Ouest, Montreal, QC, Canada

A B S T R A C T

Many techniques exist for the evaluation of nutrient pollution, but most of them require large amounts of data and are difficult to implement in countries where accurate water quality information is not available. New methods to manage subjectivity, inaccuracy or variability are required in such environments so that water managers can invest the scarce economic resources available to restore the most vulnerable areas. We propose a new methodology based on grey clustering which classifies monitoring sites according to their need for nitrogen pollution management when only small amounts of data are available. Grey clustering focuses on the extraction of information with small samples, allowing management decision making with limited data. We applied the entropy-weight method, based on the concept of information entropy, to determine the clustering weight of each criterion used for classification. In order to reference the pollution level to the anthropogenic pressure, we developed two grey indexes: the Grey Nitrogen Management Priority index (GNMP index) to evaluate the relative need for nitrogen pollution management based on a spatiotemporal analysis of total nitrogen concentrations, and the Grey Land Use Pollution index (GLUP index), which evaluates the anthropogenic pressures of nitrogen pollution based on land use. Both indexes were then confronted to validate the classification. We applied the developed methodology to eight estuaries of the Southern Gulf of Mexico associated to beaches, mangroves and other coastal ecosystems which may be threatened by the presence of nitrogen pollution. The application of the new method has proved to be a powerful tool for decision making when data availability and reliability are limited. This method could also be applied to assess other pollutants.

Keywords:
Coastal ecosystems
Coastal management
Entropy weighting
Grey clustering
Nitrogen pollution
Water quality

HIGHLIGHTS

Grey clustering helps coastal pollution management under low data availability.
Entropy weighting identifies the main sources of water pollution from land use.
Urban development over mangroves is a relevant source of nitrogen pollution.
The limited available data in poor regions can be useful for water management.

G R A P H I C A L A B S T R A C T
1. Introduction

Many methods have been developed to identify nutrient pollution (Andersen et al., 2011; Ferreira et al., 2011; Garmendia et al., 2012; Lundberg et al., 2005; Primpas et al., 2010). However, most of these methods were developed in countries where large water quality datasets are available and nutrient emissions are regulated. Consequently, they consider many variables and tend to require large amounts of data. These tools become difficult to implement for those countries where environmental monitoring and policy is less developed. The progress towards the implementation of nutrient pollution management tools entails a challenge for these countries and requires a gradual implementation (Garmendia et al., 2015). It has been recognized that results of environmental management should not be generalized worldwide as economic development plays an important role in sustainable practices (Sánchez-Hernández et al., 2017). The development of tools to be implemented in areas of limited data is necessary and has proven to contribute to bringing useful nutrient management information and prioritize problems for attention (Do-Thu et al., 2011; Firmanyah et al., 2017; Montangero and Belevi, 2007, 2008). New methods are necessary for a rapid overall evaluation of coastal water quality with the scarce and sometimes unreliable data available (Shaban et al., 2010; Xianyu et al., 2017), while governments switch to more restrictive environmental policies.

Multivariate statistical techniques such as cluster analysis are widely used for the spatiotemporal variation analysis of water quality (Hajigholizadeh and Melesse, 2017; Kitsiou and Karydis, 2011; Shaban et al., 2010; Vadde et al., 2018). But in developing countries decision making often relies on limited data (Schärer et al., 2006), and traditional methods cannot deal with the uncertainty from data collection, storage, processing and interpretation (Lermontov et al., 2009). Conventional water quality indexes are weak in dealing with inaccuracies or vagueness (Azarnivand, 2017), and thus, new methods to assess uncertainty are required. Fuzzy logic has proven to be a useful and robust method under these circumstances (Schärer et al., 2006), which enables the processing of imprecise information (Adriaenssens et al., 2004). However, most fuzzy indexes require many parameters and are not reliable when information of a single pollutant is available or when samples are too small. In opposition to fuzzy logic, grey systems can deal with small samples (Delgado and Romero, 2016) and focus on objects with clear extension and unclear intension (Liu and Lin, 2010), as further explained by Delgado and Romero (2016).

The grey systems theory was developed by Deng (1985) to deal with situations where the available information is poor or the samples used are small (Liu and Lin, 2010). This theory works with uncertain systems in which only partial or low quality data are available (Gong and Forrest, 2014), allowing the decision maker to excavate and extract useful information and to reach an accurate conclusion. A few authors have investigated the implementation of grey theory in water quality analysis (Qi et al., 2008; Wen and Wei, 2006; Zhu and Liu, 2009), which is a useful technique when the system is only partially known. Grey clustering is one of the most useful contents of the grey systems theory, which allows the classification of objects into definable classes (Delgado and Romero, 2016). The grey clustering method based on whitening weight functions and the entropy weighting method to determine the clustering weights in grey clustering: the Grey Nitrogen Management Priority index (GNMP index). This index was applied to the evaluation of nitrogen pollution based on spatiotemporal variations of a single pollutant: total nitrogen. Then, we developed a second index with grey clustering which evaluated the nitrogen pollution pressures based on land use: Grey Land Use Pressure index (GLUP index). The results of both indexes were compared to determine the accuracy of the methodology. This method was applied to eight estuaries with mangroves and other wetlands of the Southern Gulf of Mexico where nitrogen pollution is a threat to the ecosystems but where very little information is available.

2. Materials and methods

Water pollution management decisions often needs to be taken out of limited data, i.e. samples which are too small for statistical analysis or where only information from one pollutant is available. With the aim of classifying nitrogen pollution and its management requirements in several estuaries for which only limited information is available, we developed two indexes based on grey clustering:

1. Grey Nitrogen Management Priority index: GNMP index
2. Grey Land Use Pressure index: GLUP index

Since water pollution should be evaluated in relation to the anthropogenic pressures (Ninčević-Gladan et al., 2015), both indexes were confronted in order to determine the relationship between nitrogen pollution and land use pressures. As such, nitrogen pollution management priorities can be established, and remediation plans can be proposed based on the GLUP index. The method developed is schematized in Fig. 1 and explained with detail in the next sections.

2.1. Study area

The Mexican legislation does not consider nutrient pollution in natural water bodies and the accurate monitoring of nitrogen in coastal systems is not regulated. As a consequence, the lack of data prevents stakeholders from implementing the existing assessment methods and proposing recovery measures. Phytoplankton growth is generally nitrogen limited in the Gulf of Mexico (Turner and Rabalais, 2013) and the control of nitrogen pollution is a must for environmental protection.

The Mexican state of Veracruz is located in the Southwestern Gulf of Mexico and covers 745 km of coastline. Approximately 27% of the state population lives within 20 km of the coast (Macauley et al., 2007). The estuaries in Veracruz have been affected by nutrient enrichment for decades (Macauley et al., 2007; Temino-Boes et al., 2019), but agriculture, urbanization and other economic activities such as tourism along the coast are still a grown source of water pollution (Adame et al., 2018; Delgado and Romero (2016) proposed the incorporation of the entropy weighting method to determine the clustering weights in grey clustering analysis for environmental conflict analysis. Since then, other researchers have evaluated the suitability of integrating grey clustering and the entropy weighting method in other applications such as power systems security risk assessment (Peng et al., 2017), green transportation planning (Ma et al., 2017) or power quality assessment (Sacuș et al., 2018). The entropy weighting method has also been used for the development water quality indexes in combination with other tools which deal with uncertainty such as fuzzy systems theory (Chen et al., 2019).
Rivera-Guzmán et al., 2014). The Mexican legislation does not regulate nitrogen emissions to natural water bodies, and no data exist on direct inputs from urban or industrial sources. Nonetheless, nitrogen pollution has dramatic consequences for mangroves in estuaries, and it allows the massive intrusion of water hyacinths into beaches and mangroves (Temino-Boes et al., 2019). Clearly, management decisions need to be taken as soon as possible to set remediation plans and avoid further deterioration.

We evaluated 8 monitoring sites located within estuaries with mangroves of the Central Gulf hydrological region of the state of Veracruz in Mexico (Fig. 2). The local government provided the nitrogen concentrations used in this study, who measured nitrogen concentrations in several monitoring sites located along the coast. Total nitrogen was measured according to the Mexican standards NMX-AA-026-SCFI-2010 and NMX-AA-079-SCFI-2001. The data used for this study includes four annual measurements from 2013 to 2016: two measures correspond to the dry season (16 October to 15 May) and two correspond to the wet season (16 May to 15 October). However, the campaigns were not always equally spaced, and some uncertainties exist related to the exact date of the sampling. To deal with the inaccuracies and the uncertainty in the methods used during the campaigns (locations of the monitoring site, time of sample collection, etc.), we considered grey systems theory to be a reliable tool. The land use associated to mangroves from 2010 and 2015 was downloaded from the National Commission for the Knowledge and Use of Biodiversity website (”CONABIO”, 2016). The main objective was to prioritize nitrogen pollution management within the estuaries due to the consequences it may have for the surrounding ecosystems such as mangroves.

2.2. Spatiotemporal criteria for Grey Nitrogen Management Priority (GNMP) index

Water pollution evaluation has always relied on spatiotemporal analysis (Ali et al., 2016; Hajigholizadeh and Melesse, 2017; Temino-Boes et al., 2019). As such, a good water pollution evaluation should consider both spatial and temporal variations of water quality, which should be reflected when developing a classification method (Li et al., 2016). The detection of upward temporal trends in environmental parameters is necessary to not only rank concentrations but also identify those areas with increasing pollution (Chaudhuri and Dutta, 2014), especially in developing countries where water quality problems are often rapidly increasing (Li et al., 2016). The evaluation of differences in spatial mean concentrations of pollutants gives an idea of which site is more polluted now but does not provide information on the future trends. Additionally, an increasing trend clearly indicates a growing source of pollution. Considering the above-mentioned, and with the aim of developing a method which evaluates a single pollutant based on its spatiotemporal variations, we selected two criteria as inputs for the GNMP index: spatial nitrogen differences among sites and temporal trends in each site.

2.2.1. Temporal criterion

The first parameter evaluates the temporal trend of total nitrogen concentrations in each monitoring site. The usual method to determine the trend is based on least squares regression. However, this method requires a linear trend, assumptions on the normal distribution of the data and is very sensitive to outliers. Hence, we used a non-parametric estimation of the trend called Sen’s slope (Sen, 1968). Non-parametric methods do not require assumptions on the normal distribution of the data and are not distorted by outliers or missing data (Kitsiou and Karydis, 2011). These methods show useful results for the evaluation of incomplete environmental monitoring data (Scannapieco et al., 2012). Sen’s slope has been used in many applications for water quality analysis to detect trends in pollutants (Koh et al., 2017; Machiwal et al., 2019; Tabari et al., 2011a, 2011b) as it represents an absolute measure of change (Miró et al., 2018). As such, we used the Sen’s slope to derive an indicator which represents the temporal trend in nitrogen concentrations. We calculated the Seasonal Sen’s slope for total N concentrations in each monitoring site with the package “trend” in R version 3.5.1. The Seasonal Sen’s slope takes into account the seasonality of the data. In our case study, the seasonality was four, as four samples were collected each year. The scores are first computed for each season...
separately and finally the corrected Z-statistics for the entire series is calculated (Pohlert, 2020). In order to generate a normalized criterion ranging from zero to one which would allow us to compare pollution trends in each site, we used the following equation:

\[ C_{\text{temporal}_i} = \frac{s_i - s_{\text{min}}}{s_{\text{max}} - s_{\text{min}}} \]  

where \( C_{\text{temporal}_i} \) is the temporal variation criterion in site \( i \), \( s_i \) is Sen’s slope in \( i \), \( s_{\text{min}} \) is the minimum Sen’s slope between all sites and \( s_{\text{max}} \) is the maximum Sen’s slope between all sites.

### 2.2.2. Spatial criterion

To evaluate the spatial differences of nitrogen concentrations among the monitoring sites we used the median of total N of each site. The median is a robust measure of central tendency, which is not skewed by outliers. As the sampling size is not large and data reliability is not clear, the median was selected as a better measure than the mean. We determined a normalized criterion with values going from 0 to 1 which allows the comparison of nitrogen concentrations among sites:

\[ C_{\text{spatial}_i} = \frac{M_i - M_{\text{min}}}{M_{\text{max}} - M_{\text{min}}} \]  

where, \( C_{\text{spatial}_i} \) is the spatial variation criterion in site \( i \), \( M_i \) is the median of total nitrogen in site \( i \), \( M_{\text{min}} \) is the minimum median value between all study sites and \( M_{\text{max}} \) is the maximum median value between all sites.

### 2.3. Pressure criteria for the Grey Land Use Pressure (GLUP) index

For the development of the GLUP index, we defined the criteria by evaluating the anthropogenic pressures on N pollution based on land use. We downloaded maps of land use from the National Commission for the Knowledge and Use of Biodiversity website (“CONABIO”, 2016), which used SPOT images to create them. The images which more closely correspond to nitrogen data were images from 2010 and 2015. Although this timeframe does not correspond fully with nitrogen monitoring years (2013–2016), it represents the trend of land use changes. Nonetheless, the inaccuracy in the dates of land use maps is addressed by grey clustering which can deal with incomplete information. After carefully reviewing the scientific literature we identified three main sources of nitrogen pollution associated to land use in the study area: agriculture, livestock and urban expansion along the coast. We defined one criterion for each source.

#### 2.3.1. Agriculture

Veracruz is the second state in Mexico with the highest amount of land used for agriculture. The agriculture production is mainly composed of cereals (40.1% of the cultivated land), industrial crops (31.5%) and fruit trees (21.6%) (SAGARPA, 2009). The application of fertilizers to increase agricultural productivity is encouraged by governmental policies, while farmers are not provided with appropriate training (Anguiano-Cuevas et al., 2015). Consequently, natural water bodies are being increasingly impacted by diffuse nutrient pollution generated from such practices. Besides, the conversion of forest and grassland to crop agriculture also may contribute significantly to nitrogen loading to coastal systems (López-Portillo et al., 2017). As such, the first pressure criterion \( C_{\text{agriculture}} \) used as input for the GLUP index calculation was the percentage of the watershed used for agriculture.

#### 2.3.2. Livestock

The livestock subsector in Mexico is very diverse and widespread throughout the territory. Livestock farming in Veracruz consists of both farms that are managed with modern and competitive systems and others which use the most traditional practices. It is also characterized by its extensive management and seasonal production. Most cattle feed is based on grazing, with pastures managed in a free grazing system. Livestock occupies about 51% of the total area of the state with 3.7 million hectares. The bovine cattle stands out for its importance in production, which is used both for meat and milk production (SAGARPA, 2009). As a consequence, land clearing for cattle ranching
is also a predominant source of pollution throughout the studied watersheds (González-Marín et al., 2017; López-Portillo et al., 2017; Riveras-Guzmán et al., 2014; Rodríguez-Romero et al., 2018; Vázquez-González et al., 2015). Therefore, the second pressure criterion used for the calculation of the GLUP index is the percentage of the watershed used for livestock production, Clivestock.

2.3.3. Urban

Most of the urban areas in Veracruz are located along the coast and consequently urban development expands over beaches, dunes and mangroves, parallel to the coastline (Martínez et al., 2014). In fact, many researchers identified the rapid urbanization over beaches and mangroves throughout the coast as one of the main source of water pollution to our study sites (Marín-Muñiz et al., 2016; Martínez et al., 2014; Martínez et al., 2017; Mendoza-González et al., 2012; Rodríguez-Romero et al., 2018). The accelerated urban development does not allow the implementation of the required wastewater treatment facilities (Rodríguez-Romero et al., 2018) or the adequate coastal ecosystem management (Martínez et al., 2014). Moreover, direct urban pollution affects the lower course of the watersheds, adding to the fact that self-purification is not as efficient as in the upper course (Rodríguez-Romero et al., 2018). As such, for the urban pressure criterion we considered more adequate to focus on the urban development along the coast. The aim of this criterion was to represent exclusively the urban expansion over the coastal ecosystems such as beaches, dunes or mangroves. This expansion occurs generally at less than 1 km from the coast, especially in touristic areas (Mendoza-González et al., 2012). As such, the selection of a buffer of 1 km around the sampling points was considered adequate. Therefore, we calculated the urban expansion within 1 km of the sampling points observed between 2010 and 2015, which was the third input criterion C_Urban for the GLUP index.

2.4. Grey clustering

In grey systems theory, a system with totally unknown information is called a black system, while a system with fully known information is a white system. In between we find grey systems which have partially known information (Tseng, 2009), with small samples and poor information (Liu and Lin, 2010). Similarly, a grey number is a number whose value lies within an interval, but whose exact value is unknown. In this context, whitenization weight functions are used to determine the preference a grey number has over the interval of values it might take by describing what is known (Liu and Lin, 2010). Grey clustering is a method developed to classify observation objects into classes using either grey incidence or whitenization weight functions (Liu and Lin, 2010). The second method is mainly used to check whether objects belong to predefined classes (Liu and Lin, 2010).

The point of a grey class with a maximum degree of greyness is known as the center λ. (Liu and Lin, 2010). The center-point triangular whitenization weight function relies on the center of the interval, where the cognitive certainty of the object belonging to a defined class is higher and therefore is often considered more reliable and scientific (Delgado and Romero, 2016). For grey clustering in s classes, the left and right endpoints are extended horizontally from λ_l to zero and from λ_u to the highest possible value of the criterion (Ye et al., 2018). As such, an object whose criterion is lower than λ_l totally belongs to the first grey class, while an object whose criterion is greater than λ_u totally belongs to the highest class. Center-point triangular whitenization weight functions were used in other environmental applications, demonstrating the usefulness of this method to solve such problems (Delgado et al., 2018; Delgado and Romero, 2016). The steps followed for the grey clustering are described below (Liu and Lin, 2010):

**Step 1.** Define n criteria (j = 1, 2, ..., n), m objects (i = 1, 2, ..., m) to be classified in s classes (k = 1, 2, ..., s), and the observed data values x_{ij}.

**Step 2.** Divide the values field of each criterion into s equal grey intervals ([a_{0k}, a_1], [a_1, a_2], ..., [a_{s-1}, a_s]) and define the center-points λ_k of each interval (λ_{1k}, λ_{2k}, ..., λ_{sk}).

**Step 3.** Determine the whitenization weight function f_k(x_{ij}) for each kth class of each jth criterion with the next equations:

For k = 1

\[
 f_j^1 = \begin{cases} 
 1 & x < \lambda_1 \\
 \frac{\lambda_2 - x}{\lambda_2 - \lambda_1} & \lambda_1 \leq x < \lambda_2 \\
 0 & x \geq \lambda_2 
\end{cases}
\]

(3)

For 1 < k < s

\[
 f_j^k = \begin{cases} 
 0 & x \notin [\lambda_{k-1}, \lambda_{k+1}] \\
 \frac{x - \lambda_{k-1}}{\lambda_k - \lambda_{k-1}} & x \in [\lambda_{k-1}, \lambda_k] \\
 \frac{\lambda_{k+1} - x}{\lambda_{k+1} - \lambda_k} & x \in [\lambda_k, \lambda_{k+1}] 
\end{cases}
\]

(4)

For k = s

\[
 f_j^s = \begin{cases} 
 0 & x < \lambda_{s-1} \\
 \frac{x - \lambda_{s-1}}{\lambda_s - \lambda_{s-1}} & x \in [\lambda_{s-1}, \lambda_s] \\
 1 & x \geq \lambda_s 
\end{cases}
\]

(5)

**Step 3.** Select a clustering weight η_j for each criterion j.

**Step 4.** Calculate the clustering coefficients for each criterion j and each class k as:

\[
\sigma^k = \sum_{j=1}^{m} \eta_j f_j^k(x_{ij}) \quad i = 1, 2, ..., m \quad j = 1, 2, ..., n \quad k = 1, 2, ..., s
\]

(6)

**Step 5.** If \( \max_{1 \leq k \leq s} \{\sigma^k\} = \sigma^k \) then object i belongs to grey class k.

For each criterion, we defined three classes corresponding to “High”, “Medium” and “Low” priorities, and thus, the whitenization weight functions are as represented in Fig. 3. The clustering objects correspond to the monitoring sites. The clustering criteria used were the spatial and temporal criteria defined in Section 2.2 for the GNMP index and the agriculture, livestock and urban criteria defined in Section 2.3 for the GLUP index (Table 1). The clustering weights are defined based on the entropy weighting method explained in Section 2.5.

![Fig. 3. Center-point whitenization weight functions used for each criterion and with three grey classes (high, medium and low).](image-url)
2.5. Entropy weighting method

The clustering weights are calculated with the Shannon entropy, which measures the uncertainty in the information provided by each criterion (Delgado and Romero, 2016):

**Step 1.** Normalize each criterion:

\[ P_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} K_{ij}} \]  \hspace{1cm} (7)

**Step 2.** Calculate the entropy \( H_j \) of each criterion:

\[ H_j = -\frac{1}{\ln(m)} \sum_{i=1}^{m} p_{ij} \ln \left( p_{ij} \right) \]  \hspace{1cm} (8)

**Step 3.** Calculate the degree of divergence \( \text{div}_j \) of the average intrinsic information provided by each criterion:

\[ \text{div}_j = 1 - H_j \]  \hspace{1cm} (9)

**Step 4.** Calculate the clustering weight \( \eta_j \) of each criterion:

\[ \eta_j = \frac{\text{div}_j}{\sum_{j=1}^{n} \text{div}_j} \]  \hspace{1cm} (10)

3. Results

3.1. Total nitrogen concentrations

Total nitrogen concentrations in each sampling site are presented in Fig. 4. The median was greater in the Northern regions with a maximum value in Nautla. The lowest total nitrogen concentrations were found in Papaloapan and Coatzacoalcos. In Fig. 5, the time variations of total nitrogen between 2013 and 2016 are represented with four measurements per year. An upward tendency was observed in Tuxpan, Tecolutla and Nautla.

3.2. Grey Nitrogen Management Priority (GNMP) index

Sen’s slope and the median of total nitrogen are shown in Table 2, together with temporal and spatial criteria and their clustering weights. The spatial criterion, which compares the current total nitrogen concentrations in each site, had a higher clustering weight (0.60) derived from the entropy-weighting method, compared to the temporal indicator which measures the trend in total nitrogen concentrations in each site (0.40).

Three grey classes were obtained corresponding to low, medium and high N pollution management priority. The clustering coefficients for each class and each criterion, together with the global clustering coefficients and the derived GNMP index are shown in Table 3. Two sites, Tuxpan and Nautla, were classified with a high nitrogen management priority, while two sites, Papaloapan and Coatzacoalcos, were classified with low priority. The other sites had a medium priority.

3.3. Grey Land Use Pressure (GLUP) index

We evaluated the extent of the agricultural area within each watershed, together with the livestock area (Fig. 6). Additionally, we calculated the increase in urban areas around the study sites. As an example, the three sites with a greater increase in the urban areas are shown in Fig. 7.

The three criteria for land use evaluation and their clustering weights are presented in Table 4, while the clustering coefficients and the derived GLUP index are presented in Table 5. The classification agrees with the results obtained with the GNMP index (see Table 3). The urban criterion had the highest clustering weight (0.67), while the agricultural criterion weight (0.22) and the livestock weight (0.11) were lower.

4. Discussion

The GNMP index, based on the global clustering coefficients presented in Table 3, was high for Tuxpan and Nautla. In terms of the temporal criteria both sites were classified as high while the spatial criterion indicated high priority for Nautla and medium priority for Tuxpan. Similarly, Tecolutla was high according to the temporal criterion indicating rapidly increasing nitrogen concentrations but was classified medium based on the spatial criterion; its overall classification was medium. The remaining monitoring sites could be evaluated similarly, indicating that the integration of both temporal and spatial differences among sites derives in a more accurate evaluation of the nitrogen management requirements. The inclusion of temporal trends into the index developed is a new proposal which was not used in most of the water quality assessment indexes developed to date (da Costa Lobato et al., 2015; Gharibi et al., 2012; Islam et al., 2013; Ocampo-Duque et al., 2007; Shooshtarian et al., 2018). And yet, the detection of temporal trends is
especially important in areas where water pollution increases rapidly due to the lack of water pollution management (Li et al., 2016), which are the areas aimed by the newly developed method. Environmental managers could determine the restoration practices required based on whether the nitrogen pollution is increasing over time or on whether the current concentrations are high.

The comparison of the classification of the GNMP index with the GLUP index is necessary in order to validate the nitrogen pollution evaluation. The linkage between pollution levels and pressures allows the evaluation of the anthropogenic influence on nutrient concentrations. It also allows to check whether the selected land use pressures have a real impact on nutrient pollution and to guide management plans (Romero et al., 2013). All monitoring sites were placed in the same class by the GNMP and the GLUP indexes, indicating that land use pressures were detected accurately. The results indicate that Nautla and Tuxpan estuaries have the highest urgency for N pollution management which agrees with previous studies which indicated the existence of N pollution in these estuaries (González-Marín et al., 2017; Marín-Muñiz et al., 2016; Rivera-Guzmán et al., 2014; Rodríguez-Romero et al., 2018; Temino-Boes et al., 2019). Studies in Nautla river indicated that human settlements are a major source of N pollution to the river at its lower course (Rodríguez-Romero et al., 2018). Casitas, the town located at Nautla estuary, has undergone severe changes in physicochemical characteristics of water in the last years (Rivera-Guzmán et al., 2014), and has lost most of its mangroves losing its filtering capacity (Marín-Muñiz et al., 2016; Rivera-Guzmán et al., 2014).

The entropy weighting method allowed the detection of the most divergent criteria for nitrogen management. As such, in the studied area the spatial criterion had a higher clustering weight (0.60) than the temporal criterion (0.40), indicating that the divergence in the values of the median of total nitrogen are higher than the divergence in the Sen's slope. The entropy weighting method allows a more flexible determination of the importance of each criterion depending on the characteristics of the area under study by evaluating the useful information provided by each criterion. On the other hand, the clustering weights assigned to the GLUP index indicate which land use criterion has a higher divergence and thus which land use activity may have a higher influence in nitrogen pollution. The highest clustering weight for the GLUP index was assigned to the urban criterion, indicating that the urban development along the coast should be the first pollution source to be addressed for nitrogen management. Tecolutla and Nautla are located within a popular touristic area named Costa Esmeralda, where tourism expansion has led to urban development over beaches and mangroves (Martínez et al., 2014). Tecolutla river basin experienced an increase of 67% of the urban area between 1994 and 2010 (Karen Osuna-Osuna et al., 2015) and in Tuxpan the human population almost doubled in 20 years (Rivera-Guzmán et al., 2014). At large, urban expansion in Veracruz has taken place over mangroves, grasslands, beaches and croplands (Martínez et al., 2014), and tourism has increased along the coast (Mendoza-González et al., 2012), reducing the coastal resilience of Veracruz while population grows (Martínez et al., 2017).

While conventional clustering methods assign a fixed class to each object, grey clustering does not provide a deterministic solution, but rather allows the partial assignment of an object to a class by means of grey numbers. For example, both Cazones and Tecolutla were classified in the medium class. However, the GNMP index of Cazones is somewhere between low (0.40 clustering coefficient) and medium (0.59 clustering coefficient), while Tecolutla’s GNMP is localized between medium and high (Table 3). The same applies for the GLUP index (Table 5). This flexibility in the classification allows the pollution managers to make decisions which are aligned with the ecosystems’ requirements, the site restoration strategies or the economic resources available, allowing a scientific understanding of nitrogen pollution when only small sampled are available. The exploration of the information provided by the available data enables a comprehensive evaluation of nitrogen pollution.

The limitations of the developed methodology should be clear before its application. This approach evaluates one single pollutant while no additional data is available, but the development of more frequent and exhaustive campaigns is necessary for the management of the coastal water quality. When economic resources are limited, governments and scientists could start by managing nutrient pollution in the areas.

Table 2

<table>
<thead>
<tr>
<th>Site</th>
<th>Sen’s slope (mgN.L−1)</th>
<th>Median TN (mgN.L−1)</th>
<th>Ctemporal</th>
<th>Cspatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuxpan</td>
<td>0.34</td>
<td>0.99</td>
<td>0.78</td>
<td>0.64</td>
</tr>
<tr>
<td>Cazones</td>
<td>−0.13</td>
<td>0.90</td>
<td>0.07</td>
<td>0.51</td>
</tr>
<tr>
<td>Tecolutla</td>
<td>0.36</td>
<td>0.89</td>
<td>0.81</td>
<td>0.49</td>
</tr>
<tr>
<td>Nautla</td>
<td>0.49</td>
<td>1.24</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Actopan</td>
<td>−0.18</td>
<td>0.90</td>
<td>0.00</td>
<td>0.51</td>
</tr>
<tr>
<td>Jamapa</td>
<td>0.07</td>
<td>0.86</td>
<td>0.37</td>
<td>0.45</td>
</tr>
<tr>
<td>Papaloapan</td>
<td>0.06</td>
<td>0.55</td>
<td>0.36</td>
<td>0.00</td>
</tr>
<tr>
<td>Coatzacoalcos</td>
<td>0.00</td>
<td>0.60</td>
<td>0.27</td>
<td>0.07</td>
</tr>
<tr>
<td>Clustering weight</td>
<td>–</td>
<td>–</td>
<td>0.40</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Fig. 5. Total nitrogen concentrations in all monitoring sites, between 2013 and 2016 with four values per year.
indicated by the method developed. But once more data become available, the use of integrative methods to evaluate the overall water quality would be a necessary step forward. Additionally, it is important to point out that as no pristine site exists, it is not possible to indicate whether all monitored sites are polluted. Although the priority should be put in those areas with a greater pollution, the sites with a lower priority should not be completely left aside. On the other hand, it is also important to consider that the studied estuaries belong to watersheds of different sizes. As such, the cleaning capacity of rivers via dilution differs among sites. The risk of water pollution in Coatzacoalcos and Papaloapan is therefore reduced due to their dilution capacity, compared to Actopan river for example. The aim of our approach is to evaluate the pollution within the estuary to estimate the potential consequences of nitrogen pollution to the surrounding ecosystems such as beaches or mangroves. For example, the consequences of nitrogen pollution in Nautla estuary include the degradation of touristic beaches with consequences for the local economy, as well as the deterioration of mangroves (Temino-Boes et al., 2019). But the impact of high nitrogen concentrations in Papaloapan river in the eutrophication of the Gulf of Mexico as a whole would be much worse than N pollution in Actopan river. The large marine ecosystem perspective is not addressed by this study.

Nitrogen pollution in estuaries, mangroves and coastal waters has severe environmental consequences. In Veracruz, 75% of the estuaries rated poor for water quality a decade ago (Macauley et al., 2007), and the urbanization along the coast has since then increased nutrient pollution (Temino-Boes et al., 2019). Massive mats of water hyacinths, driven by nutrient pollution, were observed in Nautla and Tecolutla estuaries, extending over mangroves, altering their nutrient cycles and blocking sunlight (Temino-Boes et al., 2019). Nitrogen pollution also has big economic impacts, as many of the small villages located in the coastal regions of Veracruz depend on tourism and fishing, activities which are directly affected by nitrogen pollution (González-Marín et al., 2017). Ultimately, water pollution also affects wildlife populations (González-Marín et al., 2017) and eutrophic conditions in coastal systems along the Gulf of Mexico may drive harmful algal blooms (Ulloa et al., 2017). Despite the consequences mentioned, the lack of effective water monitoring and evaluation programs prevents stakeholders from developing management plans. The methodology developed herein allows the detection of high N concentrations in estuaries as

Table 3

<table>
<thead>
<tr>
<th>Site</th>
<th>$C_{temporal}$</th>
<th>$C_{spatial}$</th>
<th>Global</th>
<th>GNMP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Tuxpan</td>
<td>0.00</td>
<td>0.17</td>
<td>0.83</td>
<td>0.00</td>
</tr>
<tr>
<td>Cazones</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Tecolutla</td>
<td>0.00</td>
<td>0.08</td>
<td>0.92</td>
<td>0.02</td>
</tr>
<tr>
<td>Nautla</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Actopan</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Jamapa</td>
<td>0.38</td>
<td>0.62</td>
<td>0.00</td>
<td>0.15</td>
</tr>
<tr>
<td>Papaloapan</td>
<td>0.42</td>
<td>0.58</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Coatzacoalcos</td>
<td>0.69</td>
<td>0.31</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Fig. 6. Agricultural and livestock production areas within the studied watersheds.
well as those estuaries with increasing trends and links the pollution to land uses. As such, based on our results, Nautla and Tuxpan estuaries have a high priority for N pollution management which should be approached mainly through the sustainable management of urban development. Both Tuxpan and Nautla are touristic destinations, which has enhanced urban growth. Tecolutla which is also a touristic destination was classified as medium priority, but the high temporal criterion indicates that N pollution increased over time and could reach higher N concentrations in a near future. Therefore, the sustainable management of tourism growth could lead to a reduction of the coastal pollution which in turn would allow a conservation of the natural heritage.

In regions where the monitoring of coastal waters is not regulated, simple methods which allow the evaluation of water pollution with limited data are very useful. It is necessary to recognize the scarcity factor to allow the distribution of the available resources efficiently (Sánchez-Hernández et al., 2017). As such, grey clustering allows the detection of areas with a high urgency for N management and allows the planning of the available economic and human resources. When data sources are limited and inaccurate, the grey evaluation developed can help with the establishment priority areas to allow decision makers to identify potential threats and propose recovery measures. This method could be used to evaluate other pollutants, and could be applied in other countries with limited data such as most Latin American countries which present

![Fig. 7. Urban area development between 2010 and 2015 in a 1 km buffer around the study sites Nautla, Tuxpan and Tecolutla.](image)
similar limitations with water pollution assessment and management (Gómez et al., 2012; Kathuria, 2006). Future research should focus on how to efficiently combine the existing tools to deal with uncertainty, such as fuzzy logic, grey systems or rough sets theory, whose employment can deal with real-world problems especially in countries with limited economic resources.

5. Conclusion

Many countries lack proper coastal water monitoring programs and consequently the data available for pollution assessment is limited. As such, the application of grey clustering together with entropy weighting significantly contributes to the accurate prioritization of N pollution management. The integration of spatial and temporal variations in a unique index, i.e. the GNMP index, evaluated both current and future trends of N pollution. On the other hand, the analysis of land use changes through the GLUP index and the application of the entropy weighting method identified the main sources of N pollution based on land use. For the study area we found urban development around the sampling site to be the main driver of N pollution. This allows the establishment of N pollution management strategies, such as the control of the urban expansion over beaches and mangroves, allowing a sustainable development while conserving the natural heritage. When economic resources are limited, the establishment of priority areas is necessary in order to allow a scientifically sound assignment of the scarce economic resources.

It is nonetheless crucial to understand and consider the limitations of the methodology. Grey clustering provides useful information derived from small and inaccurate samples, which can be extremely useful when the situation from which we departed is a completely lack of pollution evaluation. From this perspective, the information provided by the grey clustering analysis is with no doubt an important step forward. However, this is not an ideal situation which provides a thorough and unique diagnosis of the pollution levels, its pressures and impacts. The implementation of more stringent coastal monitoring programs and the development of strict environmental policies to protect water resources is necessary for the correct management of coastal pollution. But this situation is far from realistic in many developing countries. Meanwhile, research should focus on how to deal with the lack of data by combining and implementing tools such as fuzzy logic, grey systems or rough sets theory.

CRediT authorship contribution statement

Regina Temino-Boes: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization, Project administration, Funding acquisition. Rabindranath Romero-Lopez: Investigation, Validation, Supervision, Writing - review & editing. Sara Patricia Ibarra-Zavaleta: Conceptualization, Investigation, Writing - review & editing. Inmaculada Romero: Validation, Supervision, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Erasmus Mundus - MAYANET Grant Agreement Number 2014-0872/001-001, funded with support of the European Commission, and an Excellence Scholarship awarded by the Mexican Government through the Mexican Agency for International Development Cooperation (AMEXCID). The Mexican National Water Commission provided the field data.

References


