Analysis of the eutrophication in a wetland using a data-driven model

Rahmat Zarkami [ID] · **Ali Abedini** ·
Roghayeh Sadeghi Pasvisheh

Abstract Eutrophication is a major problem in the international Anzali wetland (northern Iran). The present research initially aimed to determine the trophic state index (TSI) in ten sampling sites in the main parts of the Anzali wetland (western, eastern, central, and Siahkeshim parts). After determining the TSI in the wetland, a data-driven method (classification tree model with a J48 algorithm) was implemented to predict the trophic condition in the wetland based on a set of water quality and physical-structural variables. One hundred twenty samples related to chlorophyll-*a* (the model's output) and environmental variables (the model's inputs) were measured monthly during 1-year study period (2017–2018). Based on the TSI calculation, the western, Siahkeshim, eastern, and central parts of the wetland are classified as eutrophic, super-eutrophic, hyper-eutrophic, and hyper-eutrophic, respectively. When all environmental variables were introduced to the model (with five-time randomization effort, pruning confidence factor=0.01, and seven-fold cross-validation), eight variables (bicarbonate, pH, water temperature, electric conductivity, dissolved oxygen, total phosphorus, water depth, and water turbidity) were predicted by the model. The model predicted that an increase in total phosphate, water turbidity, and electric conductivity concentration may contribute to the hyper-eutrophic state of the wetland. In contrast, the hyper-eutrophic of the wetland is associated with a decrease in water depth, dissolved oxygen, and pH concentration. According to ANOVA test, the trophic condition in the wetland can be affected by spatial and temporal patterns. Anthropogenic pressures such as the influx of chemicals particularly the nutrients (phosphorus and nitrogen) are the main cause of water enrichment (eutrophication problem) in main parts of the Anzali wetland ecosystem.

Keywords Classification tree · Chlorophyll-*a* · Phytoplankton · Trophic state index

R. Zarkami (✉)
Faculty of Natural Resources, Department of Environmental Science, University of Guilan, Sowmeh Sara, P.O. Box 1144, Guilan, Iran
e-mail: rzarkami2002@yahoo.co.uk

A. Abedini
Inland Waters Aquaculture Research Center, Education and Extension Organization, Iranian Fisheries Sciences Research Institute, Agricultural Research, Bandar Anzali, Iran

R. Sadeghi Pasvisheh
Department of Animal Sciences and Aquatic Ecology, Ghent University, 9000 Ghent, Belgium

Introduction

Eutrophication can occur in many types of aquatic ecosystems such as rivers, wetlands, lakes, and coastal areas (Le Moal et al., 2019). It is one of the main causes for destructing water quality in many types of aquatic ecosystems (Bennett et al., 2001; de

Jonge et al., 2002; Le Moal et al., 2019; Smith, 2003; Smith et al., 1999; Vitousek et al., 1997; Zhou et al., 2020). This process also can have various impacts on human health, goods and services, and the economic activities (Le Moal et al., 2019).

The trophic state mainly deals with the primary production such as algae and aquatic plant productivity (Dodds & Cole, 2007). Variations in the nutrient levels (e.g. phosphate) may cause significant changes in algal biomass (chlorophyll-*a*). This can, in turn, cause changes in water clarity or transparency (Deyab et al., 2021). A trophic state index (hereafter TSI) is a suitable method for quantifying this relationship. Carlson (1977) initially developed the trophic classification systems for the temperate lakes (scaling from 0 to 100). The most important parameters that were developed in the given index consisted of Secchi disk transparency, chlorophyll-*a*, and total phosphorus (Bekteshi & Cupi, 2014). A modified index based on the Carlson's trophic state index was proposed by Lamparelli (2004), and almost the same parameters were proposed to examine the TSI in water body. (Xu et al., 2001; Du et al., 2019) used several physical, chemical, and biological indicators consisting of total phosphorus (TP, an important nutrient for water pollution), total nitrogen (TN), chemical oxygen demand (COD), Secchi disk transparency (SD), and chlorophyll-*a* (Chl-*a*) to describe the degree of eutrophication or trophic status of lake ecosystem.

The occurrence of eutrophication in a wetland somewhat depends on the nutrient inputs from its watershed. The most common nutrients causing eutrophication are the excessive amounts of the nutrients (e.g. phosphate and nitrogen) of which phosphate concentration plays a key role for enriching wetlands and lakes. The main source of N pollutants is run-off from agricultural land, whereas most P pollution comes from households and industry including phosphorus-based detergents (Beusen et al., 2016).

Knowing the trophic state of aquatic ecosystems (e.g. particularly wetland ecosystem) can give insight into human impact, biotic integrity, and restoration planning (Dodds & Cole, 2007). Both natural and human factors can enhance eutrophication problem (Liu et al., 2010). Since eutrophication is a complicated phenomenon, many studies need to be directed toward broader-scale analyses of aquatic ecosystems (e.g. wetland, lakes, and reservoirs). To achieve this goal, various issues need to be considered such as measuring the most important physical and chemical parameters, calculating the trophic state indices or their relationship with morphology, land use, and land occupation (Fraterrigo & Downing, 2008; Liu et al., 2010; Noges, 2009; Taranu & Gregory-Eaves, 2008). According to the earlier reports (Armengol & Miracle, 1999; Hamilton et al., 2001; Liu et al., 2010; Taranu & Gregory-Eaves, 2008), eutrophication in water body also may be significantly influenced by the structural parameters such as the depth and volume of the aquatic ecosystems.

International Anzali wetland (one of the most important wetlands in northern Iran) is an enriched freshwater lagoon in the southwest of the Caspian lowlands. This wetland is a nursery habitat for various fish species and a reproduction and wintering habitat for waterbirds from different parts of the world (Nazarhaghighi et al., 2014; Sadeghi et al., 2021) and several mammals (JICA, 2019). The wetland is threatened by a wide variety of anthropogenic pressures consisting of eutrophication, deforestation, erosion, unsustainable hunting, inappropriate tourism, road and urban development (Sadeghi et al., 2012a, b), toxic substances (Mortazavi, 2018), heavy metals (Hassanzadeh et al., 2021), and invasive of exotic species (Zarkami, 2016). Besides, agricultural activities, discharge of industrial wastewater, and sanitary wastewater are the main sources of pollutants in the Anzali wetland (Sadeghi et al., 2013). In the last few decades, this wetland has been threatened by the influx of chemicals particularly phosphorus and nitrogen leading to the eutrophication problem in many parts of this international ecosystem (Sadeghi et al., 2021). As a result of this, eutrophication has become a serious threat in the Anzali wetland ecosystem. Numerous contaminants have negatively affected water quality and thus have degraded the structural nature of the wetland. For the moment, more manure and chemical fertilizers are applied in the wetland watershed to promote soil fertility and increase crop production (particularly in the rice paddy fields) leading to the water enrichment in this valuable wetland. This problem has, in turn, created favourable conditions for the overgrowth of invasive exotic species such as water fern (*Azolla filiculoides*) (Sadeghi et al., 2017) and water hyacinth (*Eichhornia crassipes*) (Zarkami et al., 2021) leading to reduced water quality and the degradation of several habitat characteristics (Sadeghi et al., 2021).

Among different types of pollution in wetlands, an appropriate monitoring campaign needs to be designed for the eutrophication problem. The reason is that wetland and lake eutrophication are becoming a global concern in water pollution. To do so, both biotic (e.g. algal biomass and its relationship with the chlorophyll-*a*) and a set of the most pertinent abiotic variables (total phosphorus, water transparency, etc.) must be considered for monitoring the eutrophication. In addition to this, the appropriate predictive modelling techniques (e.g. Ambelu et al., 2010; Zarkami et al., 2019) need to be implemented for the proper management program of wetlands (Sadeghi et al., 2013). Up to now, many studies have been conducted to examine eutrophication phenomenon in aquatic ecosystems around the world since this process is currently making serious problem for the lakes and wetlands (Hecky & Kilham, 1988).

The present study initially aimed to determine the TSI in the main parts of the international Anzali wetland. After calculating the TSI in the wetland, a data-driven model so-called classification tree or CT model with a J48 algorithm (e.g. Ambelu et al., 2010) was implemented to predict to what extent the environmental variables may contribute to the eutrophication in the wetland. The output of the model was the concentration of chlorophyll-*a*, and the inputs of the model consisted of a combination of the categorial variables, water quality, and physical and structural variables (the variables associated with the shape and structure of the wetland such as water depth) which were measured monthly in main parts of the Anzali wetland.

Materials and methods

Description of study area

As stated earlier, various problems such as anthropogenic impacts (Sadeghi et al., 2017) and invasive of exotic species (Zarkami, 2016) are threatening the Anzali wetland of which the eutrophication is one of the main problems.

In the present research, the main parts of the Anzali wetland consisting of the eastern, western, central, and southern parts of the wetland were monitored to examine the trophic condition in this valuable ecosystem.

The eastern wetland (Sheijan) has the lowest water depth among other parts of the wetland. The growth of aquatic plants (mainly *Phragmites australis*) is high because of large amount of nutrients entering this section of the ecosystem. The spread of common reed (e.g. *Phragmites australis*) and hence losing the habitat's waterfowl are the main concern in the given area. Therefore, many species are threatened, and they require a low level of chemical oxygen demand to have a sustainable life (JICA, 2019). Artificial disturbance is relatively small due to low human activities in this area. Although the eastern area constitutes a small part of the wetland ecosystem, many polluted rivers discharge into this area. Based on this, the eastern wetland is more polluted than other parts of the ecosystem.

The western wetland (Abkenar) covers more than one-third of the Anzali wetland ecosystem, and it is considered a large lagoon area. In contrast to the eastern part, the western wetland constitutes the deepest area of the Anzali wetland ecosystem. There are many adult fish species inhabiting the lagoon. The water quality condition in the western area is relatively better than other parts of the wetland. Moreover, in the western wetland, water is turbulent, and dissolved oxygen is relatively higher than other parts of the wetland (Abedini, 2015). Overall, the western area is less polluted than other sections of the ecosystem. Artificial disturbance is relatively low in the western end. The western end is abundant with marsh in which the emergent plants particularly common reed (e.g. *Phragmites australis*) are plentiful in the vicinity of this part of the wetland. The marsh of the western end has high potential for waterfowl. Fishing and hunting are the most important activities in this area. Many boats enter the area.

The central wetland (Sorkhankol) is considered a wildlife refuge not a legally protected area. This area is located between Pirbazar River and Anzali Port. There are open water areas in this part of the wetland. Floating and emergent plants are abundantly found here. The central wetland is deeper than the eastern part. Many boats pass through this part, and polluted water also flows into the central part. The biodiversity is relatively low in the central wetland.

The southern part (Siahkeshim) is a protected area covering an area of about 4500 ha. Many rivers and streams discharge into the Siahkeshim. Emergent plants like *Phragmites australis* overgrow the entire

area. Water depth is low in the southern part. This area includes many bird species. The southern part of the wetland is in danger of encroachment (JICA, 2005). Figure 1 shows the geographical location of the study areas with illustrating the sampling sites in the main parts of the Anzali wetland.

Sampling campaign

Ten sampling sites were selected in four parts of the Anzali wetland consisting of the eastern, central, western, and Siakeshim protected area (Fig. 1). The sites were monitored monthly from the 15th of April 2016 to the 15th of March 2017 (in total 12 months). The site 1 was selected at the end of shipping channel (wave breaking section). This site had less contribution to the monitoring relative to other parts of the ecosystem. The sites 2 and 3 were chosen in the central ecosystem. Two sampling sites (sites 4 and 5) were monitored in the eastern section of the wetland. Three sampling sites (sites 6, 7, and 8) were chosen in the western area. The western wetland had more contribution to the sampling than other parts. The sites 9 and 10 were selected in the Siakeshim protected area. The eastern, central, and Siahkeshim parts were equally sampled in the wetland.

Various environmental variables were monitored monthly at each sampling location. The sampling effort was repeated three times in the vicinity of each site. The three water samples, thus, were mixed up to obtain an individual sample. Therefore, in total, 120 samples were obtained from the entire sampling locations for 1-year study period (12 months). Overall, more samples were taken in the western wetland (36 samples). In contrast, the shipping channel had less contribution to the sampling (12 samples). Other three parts (Siahkeshim protected area, the eastern and central parts) were equally sampled (24 samples for each part and overall 72 samples for three parts).

The water quality measured on-site monitoring consisted of pH (using a combination electrode and calibrating against the buffer solutions), electric conductivity (measuring with or without temperature compensation and calibrating against a standard solution of potassium chloride), and dissolved oxygen (keeping a hand-held metre in a gentle motion through the water column while reading the device). The physical-structural variables which were monitored in situ field measurement were water and air temperature (with a digital thermometer) and water depth (a yardstick). The sample related to chlorophyll-*a* was
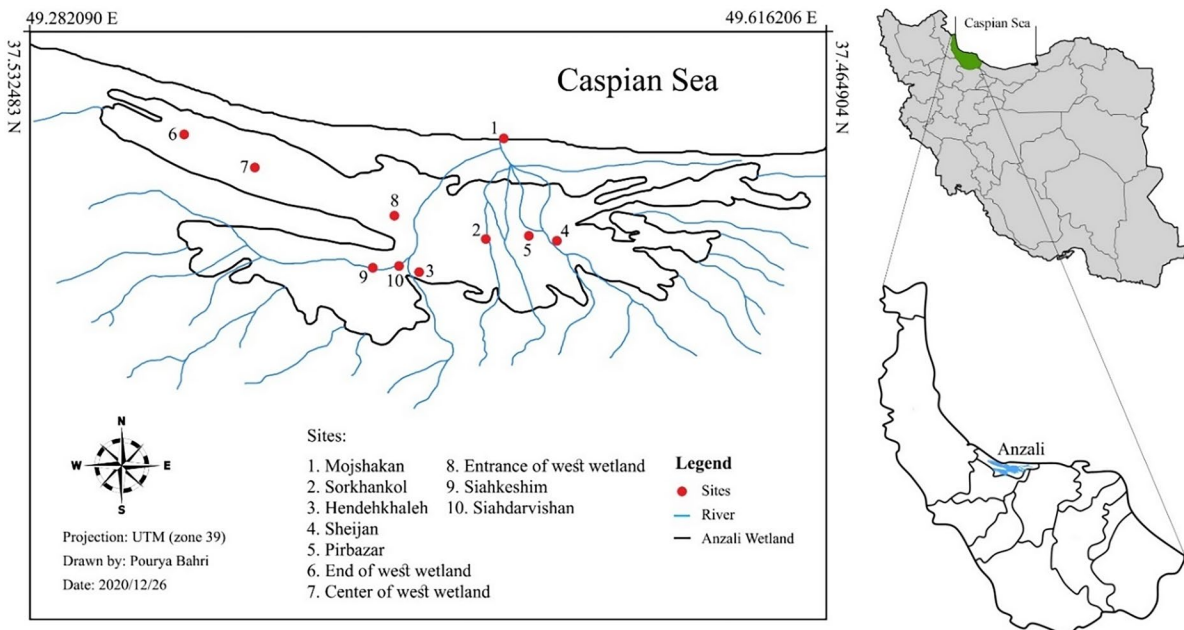

Fig. 1 Map of study areas indicating the sampling sites in the main parts of the Anzali wetland

collected on a new glass fibre filter paper after filtration of a known volume of sample. The pigments were extracted from the paper in the laboratory by spectro-photometric method. All water quality samples were kept in cold storage container (4° C). The samples were instantly sent to the laboratory (within 2–3 h) for analysing with standard method (APHA/AWWA/WEF, 2012). The samples were monitored in the middle of each month.

Trophic state index (TSI)

In the present research, a multi-parameter evaluation method of TSI (e.g. Carlson, 1977, and modified by Lamparelli, 2004) was primarily used to determine the eutrophication condition in the different parts of the wetland (Table 1). In this method, the range of TSI related to the most important parameters like Secchi disk (SD), chlorophyll-a (Chl-*a*), and total phosphate (TP) has been proposed to evaluate the TSI in water body. In the multi-parameter evaluation method of TSI, the equations related to SD (Secchi disk), CHL (chlorophyll-a), TP (total phosphate), and TN (total nitrogen) are assessed from 0 to 100. The concept of the TSI is based on the changes in the nutrient levels (particularly total phosphorus). Total phosphorus can cause changes in algal biomass (chlorophyll-*a*). This can, in turn, lead to changes in water clarity (Secchi disk transparency).

The applied model (the CT model with a J48 algorithm)

The decision trees (Quinlan, 1986) or classification trees (Breiman et al., 1984) which were applied in the

present study are artificial intelligence-based model prediction. The models predict the dependent variable from a set of independent variables, either continuous or discrete (D'heygere et al., 2003).

After determining the TSI for different parts of the Anzali wetland, a CT model with (a J48 algorithm) was implemented to predict the most important factors for assessing the trophic condition in the study areas. Due to the relatively small dataset in the present research, the straight "folds cross-validation" was used instead of other test options which are available in the WEKA toolbox consisting of "supplied test" (the model specifies a separate file containing the test set which is very often used for large datasets), "use training set" (the derived model performs on the training set), and "percentage split" (a certain percentage of the data can be held out for testing) (Witten et al., 2011).

By applying the CT model (after determining the TSI), it was possible to determine to what extent the predicted variables may have contributed to the eutrophication in the wetland.

Various environmental variables were inserted to the model (air temperature was excluded from the model due to high correlation with water temperature). In addition to the inclusion of the environmental variables to the model, all sampling sites and different seasons were also used as categorial variables for the applied model. The sites and seasons (functioning as a seasonality) were classified into 10 sites and 4 seasons to show the possible effect of the spatial (the sites) and temporal (seasonality) patterns on the eutrophication condition in the wetland. Therefore, in total, 12 input variables (10 environmental variables + 2 categorial variables) and one output variable (chlorophyll-*a*) were inserted to the model.

Since chlorophyll-*a* (the only output variable in the present research) functions a photosynthetic pigment being essential for photosynthesis in organisms (e.g. eukaryotes, cyanobacteria, and prochlorophytes), the concentrations of this variable are closely associated with the abundance and biomass of phytoplankton. Therefore, various independent variables (like the input variables recorded in the present study) may influence the amount of chlorophyll-*a* in the wetland. As a result of this, chlorophyll-*a* (an indicator of phytoplankton abundance and biomass in coastal and estuarine waters) was selected as a model's output (a dependent variable). To do so, the

Table 1 Classification of water resources based on the proposed range of trophic state index (TSI) (Carlson, 1977, and modified by Lamparelli, 2004)

Trophic classification	TSI range
Ultra-oligotrophic	$TSI \leq 47$
Oligotrophic	$47 < TSI \leq 52$
Mesotrophic	$52 < TSI \leq 59$
Eutrophic	$59 < TSI \leq 63$
Super-eutrophic	$63 < TSI \leq 67$
Hyper-eutrophic	$TSI > 67$

output of the model was divided in three TSI categorical classes which were obtained from the TSI in the wetland (eutrophic, super-eutrophic, and hyper-eutrophic). As stated already, the environmental variables were employed as the model's inputs.

The reliability of the model was checked with four pruning confidence factors (PCFs = 0.50, 0.25, 0.10, and 0.01) which are commonly used in the literature (Witten et al., 2011). Based on the Eqs. (1) and (2), the accuracy of the J48 algorithm was assessed based on two predictive performance criteria so-called the Cohen's Kappa or simply k (Cohen, 1960) and the percentage of correctly classified instances (CCI%) (Witten et al., 2011). The performance criteria of the predictive CT model for the multiclass classification (eutrophic, super-eutrophic, and hyper-eutrophic) were derived from a confusion matrix presented in Table 2. In the given matrix, the model predictions were tabulated vs real observations. For assessing the multiclass classification, the accuracy of the CT model was calculated based on these two equations, that is, k and CCI% (Sadeghi, 2022). Based on this, the two following performance criteria of the CT model were used for evaluating three TSI classes:

Results

Data pre-processing and assessment of trophic state index before model development

The distribution of data for the recorded variables in the wetland

Table 3 shows the data distribution for the recorded variables which were used for analysing the eutrophication status in the Anzali wetland. The data for the environmental variables (except bicarbonate, $p > 0.05$) had no normal distribution ($p < 0.05$). A significant difference (based on Mann–Whitney test) was found between different TSI classes and total phosphate ($p < 0.01$) and dissolved oxygen ($p < 0.01$) which indicates that there is a significant relationship between various TSI classes and the concentration of the variables.

The obtained results demonstrated that the outliers were observed in several variables (despite, all observations were retained in the model). The mild and extreme outliers were observed in some variables consisting of chlorophyll-a, water turbidity, total phosphate, total nitrogen, electric conductivity, and

$$\text{Correctly Classified Instances (CCI\%)} = (a + e + i)/\text{N} \times 100 \tag{1}$$

$$\text{Cohen's Kappa}(k) = \frac{(\text{I}/\text{N}) - (\text{II}/\text{N}) * (\text{III}/\text{N}) + (\text{IV}/\text{N}) * (\text{V}/\text{N}) + (\text{VI}/\text{N}) * (\text{VII}/\text{N})}{1 - [(\text{II}/\text{N}) * (\text{III}/\text{N}) + (\text{IV}/\text{N}) * (\text{V}/\text{N}) + (\text{VI}/\text{N}) * (\text{VII}/\text{N})]} \tag{2}$$

where N is total number of instances (120 instances); $a + e + i = \text{I}$; $a + d + g = \text{II}$; $a + b + c = \text{III}$; $b + e + h = \text{IV}$; $d + e + f = \text{V}$; $c + f + i = \text{VI}$; $g + h + i = \text{VII}$.

water transparency (Fig. 2). As visualized in the box and whisker plots, the extreme outliers were observed when the concentration of chlorophyll-a exceeded 200 μg/l. The extreme data values were also detected for the nutrients where the concentration of total phosphate and total nitrogen went above 0.60 mg/l and 3 mg/l, respectively. Such outliers for the water transparency and turbidity were observed at the values higher than 1.95 m and 100 FTU, respectively. For the electric conductivity, the amounts of greater than 9000 μS/cm were considered the extreme outliers. The outliers appeared on the total phosphate and total nitrogen are quite normal due to the eutrophication of the wetland with the excessive amount of the nutrients in different sampling sites/seasons, and for the water transparency, the extreme values can be related

Table 2 The confusion matrix for three TSI classes as a basis for evaluating the observed (actual) classes vs predicted ones

		Predicted		
		Eutrophic	Super-eutrophic	Hyper-eutrophic
Actual	Eutrophic	a	b	c
	Super-eutrophic	d	e	f
	Hyper-eutrophic	g	h	i

Table 3 The data distribution of the recorded variables for analysing the eutrophication status in the Anzali wetland ecosystem. *Stand. dev* standard deviation, *EC* electric conductivity, *DO* dissolved oxygen, *TN* total nitrogen, *TP* total phosphate, *WT* water temperature, *AT* air temperature, *Chl-a* chlorophyll-*a*, *Tra* transparency. *: not used for the model development

	*AT (°C)	WT (°C)	EC (µS/cm)	pH	Turbidity (FTU)	Depth (m)
Min	10.00	9.50	254.00	6.87	4.00	0.15
Max	31.00	32.00	14,250.00	10.03	102.00	5.00
Mean	21.23	19.26	2335.98	8.12	25.18	1.78
Stand. dev	7.17	7.43	3042.30	0.60	17.16	1.31
Median	20.20	18.50	1008.00	8.13	21.00	1.50
25 percentile	15.00	11.70	580.00	7.66	12.00	0.70
75 percentile	29.00	27.60	2550.00	8.52	33.00	2.50
	HCO3− (mg/l)	DO (mg/l)	Tra (m)	TN (mg/l)	TP (mg/l)	Chl-a (µg/l)
Min	15.30	1.67	0.10	0.28	0.01	0.20
Max	366.00	14.20	2.00	3.19	0.91	330.39
Mean	187.72	8.72	0.64	1.03	0.14	47.37
Stand. dev	66.64	2.77	0.38	0.55	0.15	65.38
Median	189.10	8.98	0.50	0.92	0.09	20.75
25 percentile	140.30	6.82	0.40	0.65	0.05	5.33
75 percentile	231.80	11.17	0.75	1.26	0.17	60.19

to the variations in the inorganic particles (e.g. sediment from erosion) or organic particles (such as dead algae and phytoplankton) in the wetland in different seasons. The difference between the concentrations of dissolved ions in various sampling sites/seasons may cause such extreme outliers for the electric conductivity. The extreme data points appeared on the variable "chlorophyll-*a*" can be attributed to the variation
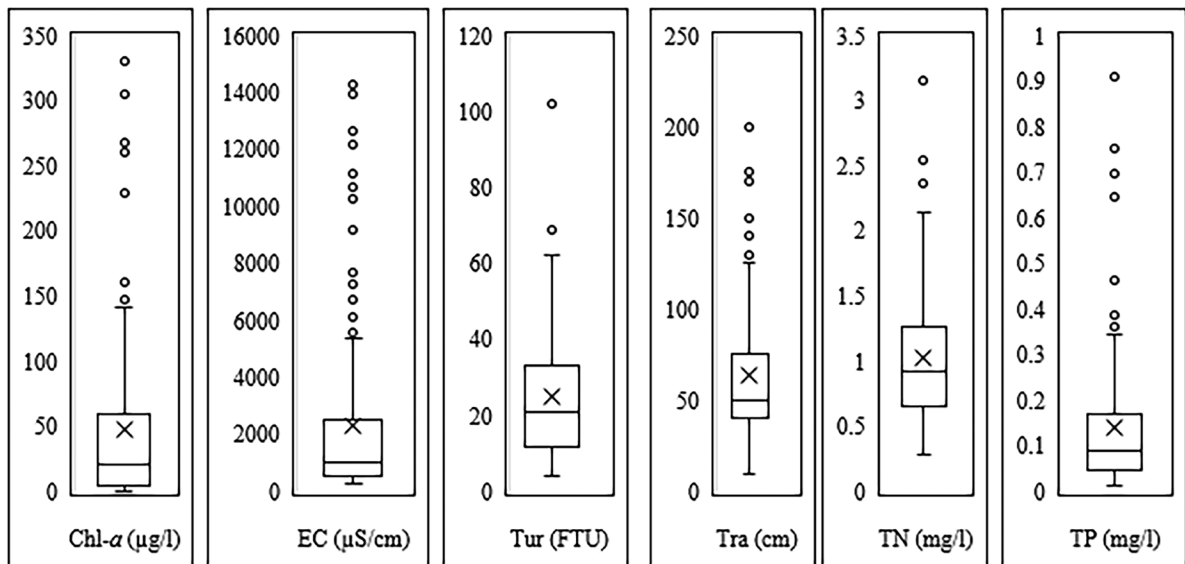

Fig. 2 The box and whisker plots for representing the range of data distribution in the Anzali wetland in different seasons (the number of instances: 120). The outliers are indicated with circles. TP Total phosphate; TN Total Nitrogen, Tur Turbidity, Chl-*a* Chlorophyll-*a*, Tra Transparency (the unit of transparency is presented in terms of "cm" for better visualizing the outliers), *EC* Electric Conductivity

in the photosynthesis activities in different seasons. Significant fluctuation in the water turbidity is as a result of the presence of suspended matter entering the Anzali wetland in different seasons.

The yearly average of the monitored variables in the wetland

The yearly average of the recorded variables in different parts of the wetland is presented in Table 4. As demonstrated in the table, the maximum and minimum of the nutrients (total phosphate and total nitrogen) were mainly observed in the eastern and western parts of the wetland, and for the water transparency, the given ranges were observed in the western and central sections of the wetland, respectively. The maximum and minimum of the chlorophyll-*a* were correspondingly found in the central and western sections of the ecosystem.

TSI

Table 5 shows the results of trophic state index (TSI) for different parts of the Anzali wetland. With looking at Table 5 and comparing it with Table 1, it can be concluded that the western and Siahkeshim parts of the wetland are classified as eutrophic and super-eutrophic, respectively. Other sections of the ecosystem (eastern, central, and shipping channel) are classified as a hyper-eutrophic, and the whole parts of the wetland are also classified as a hyper-eutrophic.

Correlation analysis

According to the correlation analysis, positive and strong correlation (with an absolute value > 0.50) was found between water temperature and air temperature ($p < 0.01$, $r = 0.94$) (as stated already, air temperature was excluded from the CT model due to strong correlation with water temperature). Chlorophyll-*a* was positively correlated with several variables including total phosphate ($p < 0.01$, $r = 0.62$), total nitrogen ($p < 0.01$, $r = 0.49$), water turbidity ($p < 0.01$, $r = 0.56$), water temperature ($p < 0.01$, $r = 0.61$), and air temperature ($p < 0.01$, $r = 0.59$). This indicates that an increase in the amount of chlorophyll-*a* is associated with an increase in the concentration/value of the nutrients (total phosphate and total nitrogen), water turbidity, and water temperature. In contrast,

Table 4 The yearly average of the recorded variables in the different parts of the Anzali wetland. *WT* water temperature, *EC* electric conductivity, *DO* dissolved oxygen, *TN* total nitrogen, *TP* total phosphate, *AT* air temperature, *Tra* transparency. *: not used for the model development

	*AT (°C)	WT (°C)	EC (µS/cm)	pH	Turbidity (FTU)	Depth (m)	DO (mg/l)	HCO_3^- (mg/l)	Tra (m)	TN (mg/l)	TP (mg/l)	Chl-*a* (µg/l)
Siahkeshim	22.75	20.08	1479.08	7.96	27.00	1.94	9.06	200.29	0.65	0.94	0.12	32
Western	21.80	19.40	1830.72	8.62	19.50	1.28	10.16	123.70	0.78	0.94	0.60	30
Eastern	20.92	18.85	1295.58	7.73	27.20	2.12	6.39	248.45	0.63	1.20	0.27	59
Central	20.12	19.33	3150.48	8.08	33.00	0.65	8.42	192.29	0.38	1.03	0.12	73
Shipping channel	19.20	17.88	6085.25	7.83	19.58	4.36	9.00	224.44	0.76	1.18	0.18	58
Total wetland	**21.23**	**19.26**	**2335.98**	**8.12**	**25.18**	**1.78**	**8.72**	**187.72**	**0.64**	**1.03**	**0.14**	**47**

Table 5 Calculation of TSI for different parts of the Anzali wetland

Sampling sites	Trophic state	TSI
Western wetland	Eutrophic	63
Siahkeshim	Super-eutrophic	66
Shipping channel	Hyper-eutrophic	69
Eastern wetland	Hyper-eutrophic	71
Central wetland	Hyper-eutrophic	76
Total wetland	**Hyper-eutrophic**	**69**

a strong and negative correlation (with an absolute value > 0.50) was found between water turbidity and water transparency ($p < 0.01$, $r = -0.57$) and bicarbonate and pH ($p < 0.01$, $r = -0.54$) which demonstrates that an increase in the concentration of water turbidity and bicarbonate may decrease water transparency and pH in the wetland, respectively.

Relationship between the environmental variables and the monitoring sites/seasons

According to the one-way ANOVA (post hoc test), several variables showed a significance difference with various sampling sites and seasons in the wetland consisting of pH (sites: $F = 7.53$, $p < 0.05$; seasons: $F = 19.06$; $p < 0.05$), electric conductivity (sites: $F = 4.00$, $p < 0.05$; seasons: $F = 22.29$; $p < 0.05$), dissolved oxygen (sites: $F = 6.50$, $p < 0.05$; seasons: $F = 5.76$, $p < 0.05$), water transparency (sites: $F = 3.70$, $p < 0.05$, seasons: $F = 9.63$, $p < 0.05$), and total phosphate (sites: $F = 5.54$, $p < 0.05$, seasons: $F = 8.67$; $p < 0.05$). Water turbidity, total nitrogen, chlorophyll-a, and air/water temperature had a significant difference only with the sampling seasons ($p < 0.05$ for all variables), while water depth and bicarbonate showed a significant difference only with the sampling sites ($p < 0.05$ for both variables).

Modelling

When the induced CT model was run with all variables (1 output and 12 input variables), only one variable (the sampling sites) was decided upon by the model, so other variables were not predicted by the model. Therefore, this CT result did not have an added value for the interpretation of eutrophication because no exact tree was constructed by the model.

Consequently, this categorial variable (the sites) had to be excluded from the model to construct a meaningful tree. When other input variables (the environmental variables and seasonality) were retained in the model, a tree (including seasonality) was constructed, but the variable "seasonality" was merely predicted at the lower PCF level (the $PCF = 0.50$). This made the constructed tree very complex (the number of leaves: 14 and size of the tree: 25). The complexity of the constructed tree also had a less added value for the interpretation of eutrophication in the wetland. As a result of this, the variable "seasonality" also was excluded from the model in the next step. With including the rest of input variables (10 variables) to the J48 algorithm, eight input variables were finally decided upon by the model. In this case, a seven-fold cross-validation (albeit after five-time randomization effort) gave the best predictive performances ($CCI = 80\%$, $k = 0.67$) than other fold cross-validation (Fig. 3). This indicates that the model reached its highest reliability to predict the trophic condition in the sampling sites.

The constructed tree (with the highest pruning confidence factor: $PCF = 0.01$) (Fig. 4) showed that several variables consisting of bicarbonate, total phosphate, water turbidity, pH, electric conductivity, dissolved oxygen, water temperature, and water depth may play a key role for predicting the trophic state in the Anzali wetland of which the decision tree initially started growing with "bicarbonate." In this step, the model did not make an ultimate decision for the variable "bicarbonate." However, all predicted variables in this phase depended on the "bicarbonate." According to the rule extracted from the CT model, when pH concentration goes toward the basic condition (> 7.75) and when the amount of water turbidity is high (> 60 FTU), the ecosystem is considered hyper-eutrophic. When the amount of water turbidity falls below 60 FTU, the wetland may be classified as eutrophic. In this case, the rule for model's decision is reasonable since 34.29 out of 120 instances were assigned to the given rule and 5.29 out of 34.29 instances still were categorized as misclassified. In the right side of the tree, the variables "water temperature, dissolved oxygen, electric conductivity, water depth, water turbidity, total phosphate, and pH" were predicted by the model. If water temperature shows a decrease in the Anzali wetland region (≤ 28 °C), the prediction of the model depends on several factors. On

Fig. 3 Checking the reliability of the model with applying different fold cross-validations for analysing the trophic condition in the Anzali wetland (CCI %: percentage of correctly classified instances)
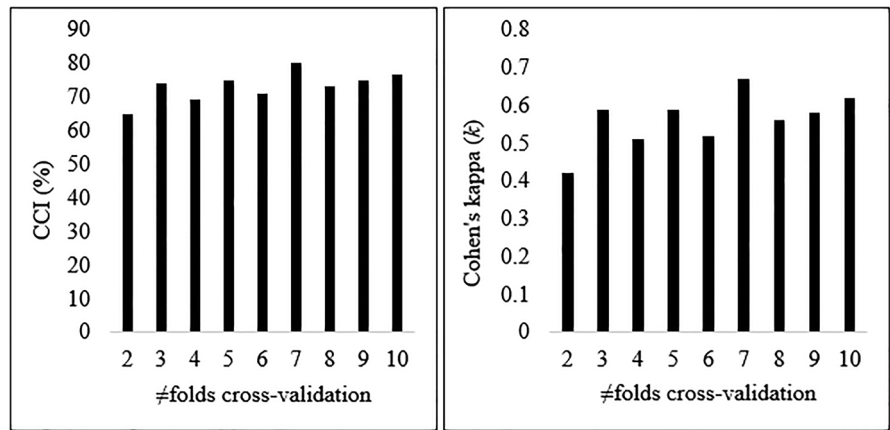
the contrary, if water temperature increases (> 28 °C), the concentration of dissolved oxygen will determine the trophic condition in the wetland. According to the decision made by the model, if the dissolved oxygen drops (≤ 6.93 mg/l), the wetland will become hyper-eutrophic. If electric conductivity is high (> 1008 µS/cm), the wetland will be hyper-eutrophic, while when

the amount of the electric conductivity is decreasing (≤ 1008 µS/cm), the trophic condition in the wetland may depend on several factors consisting of water depth, water turbidity, total phosphate, and pH. The wetland may be classified hyper-eutrophic, when total phosphate is high (> 0.13 mg/l) or depth of the wetland is low (≤ 0.8 m).
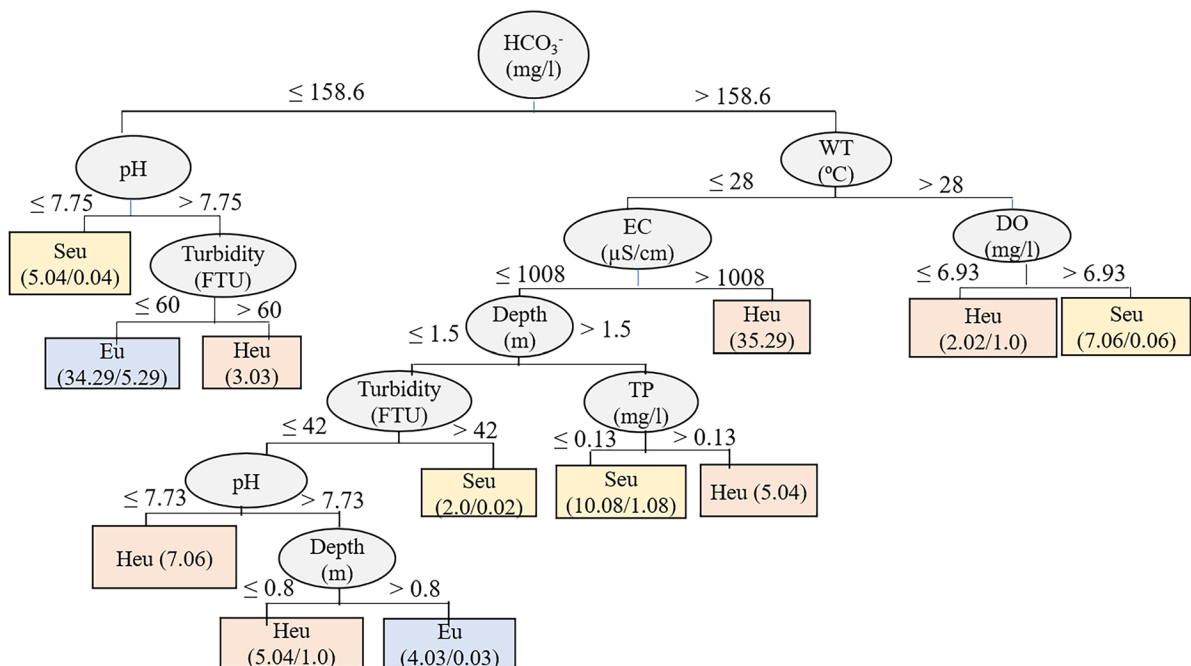

Fig. 4 The induced tree for predicting the trophic state in the Anzali wetland (*PCF* = 0.01; *values between brackets indicate the instances in which rules are true/false*); the number of instances: 120; DO, dissolved oxygen; WT, water temperature; blue rectangle, eutrophic (Eu); orange rectangle, super-eutrophic (Seu); pink rectangle, hyper-eutrophic (Heu); grey oval, the selected attributes by the model

Among 10 input variables inserted to the model (after excluding sampling sites and seasonality from the model in two consecutive stages), total nitrogen and water transparency were never predicted by the induced model. As stated in the correlation analysis, total nitrogen and water transparency were strongly correlated with total phosphate ($r = 0.60$) and water turbidity ($r = -0.57$), respectively. Based on this, the model decided to predict the most important variable(s).

Discussion

Since eutrophication is affected by both anthropogenic and natural factors (Liu et al., 2010), the scientific research for analysing the trophic condition in wetlands is globally important for the management of these valuable ecosystems. Based on this, the present research firstly aimed to examine the TSI in one of the most important wetlands in Iran (the international Anzali wetland). Additionally, a predictive modelling (the CT model with a J48 algorithm) helped identify the most important explanatory variables that may play a key role for analysing the eutrophication condition in the wetland. To do so, it was possible to determine to what extent the predicted variables may contribute to the eutrophication in the wetland.

The maximum amount of chlorophyll-*a* was found in the sampling sites where the nutrients (mainly phosphate and nitrate) had the highest concentration. According to the earlier findings (e.g. Le Moal et al., 2019), the availability of nutrients (particularly the phosphate and nitrate) can stimulate photosynthesis activity leading to a high algal production. This can, in turn, intensify the eutrophication problem in the wetland. Moreover, an elevated amount of chlorophyll-*a* may enhance water turbidity in the wetland.

According to the correlation analysis, the minimum amount of chlorophyll-*a* is linked to an elevated water transparency. This can be justified that the sampling sites characterized with the clear waters (those characterized with high water transparency or less water turbidity) as well as less polluted waters (those characterized with high oxygen concentration) are not favourable places for the phytoplankton production (Le Moal et al., 2019; Silvino & Barbosa, 2015).

With introducing various variables to the J48 algorithm (physical–chemical, structural, and the variables related to the spatial and temporal patterns), the most important explanatory variables that were responsible for predicting the trophic state of the wetland were identified by the model. Such predicted variables can help wetland managers to appropriately restore aquatic ecosystems (Lock & Goethals, 2012) based on the eutrophication problem. In the present study, the highest level of tree pruning ($PCF = 0.01$) was applied to make a simple but more powerful tree (De'ath & Fabricius, 2000). The earlier studies (Witten et al., 2011) showed that tree pruning at high PCF value (e.g. PCF-0.01) may decrease the size and complexity of the model. The optimization of tree with the highest PCF resulted in an easy interpretation of the predicted variables for analysing the eutrophication condition in the Anzali wetland.

Based on the model's outcome, the variable "bicarbonate" was the first explanatory predictor to examine the eutrophication condition in the wetland. This predictor appeared at the beginning of the induced tree. Other predicted factors in the wetland were associated with the bicarbonate. In this step, however, no definitive decision was made by the model. The main reason is that an increase in the bicarbonate concentration may cause more nutrients (particularly total phosphate) that are taken up by the algae in the Anzali wetland. This may, in turn, increase the biomass production of algae and hence intensify eutrophication in this ecosystem (e.g. Bordalo et al., 2001; Cloern, 2001). Water temperature and pH were constructed at the right and left side of tree, respectively. All three predicted variables in the first and second levels of the induced tree (bicarbonate, water temperature, and pH) are intercorrelated with each other. In the sampling sites, pH was strongly and negatively correlated with bicarbonate ($r = -0.54$).

As confirmed by the model, along with water quality variables, the physical-structural variables such as water temperature and water depth may play a key role in predicting the trophic condition in the Anzali wetland. Based on the outcomes of the model, the hyper-eutrophic of the wetland (high phytoplankton production) may be linked to high water temperature together with low dissolved oxygen. The prediction of CT model is logic because high water temperature particularly in hot seasons (with hot spring and summer seasons characterized in the climate of northern Iran) can stimulate algal blooms in the wetland. However, many factors

(e.g. nutrients) may also accelerate algal blooms and hence contribute to the eutrophication problem in the wetland of which water/air temperature can be one of the most important driving predictors to speed up the growth of algae. The overgrowth of algal blooms may, in turn, result in a significant drop of the dissolved oxygen concentration in the wetland. However, it is worth mentioning that the concentration of dissolved oxygen can considerably fluctuate during daytime and nighttime in the eutrophicated wetland. Despite, based on the applied model, the general trend of dissolved oxygen shows a significant drop of this variable over a long period of time (1-year study period) in this international ecosystem. The impact of water/air temperature on the eutrophication of aquatic ecosystems was already confirmed by earlier studies (e.g. Le Moal et al., 2019).

The results of the model also showed that the wetland may be hyper-eutrophic when the water depth is becoming shallower. In contrast, the wetland may be eutrophic when the water depth is becoming deeper. The maximum water depth was measured in the western part of the wetland. This indicates that a gradual increase in the water depth in the Anzali wetland ecosystem may limit the phytoplanktonic blooms. As a result, this would lead to a low chlorophyll-*a* production by phytoplankton. The finding is in line with previous reports (e.g. Le Moal et al., 2019; Liu et al., 2010; Taranu & Gregory-Eaves, 2008) who confirmed that the shallower waters can cause wetlands more eutrophic than the deeper waters. However, based on the outcomes of model's prediction, the wetlands may still be hyper-eutrophic (at the higher water depths) when the nutrients (e.g. phosphate) are found in high concentration.

Based on the outcomes of the predictive model, a hyper-eutrophic in the wetland is associated with an increase in pH concentration together with high water turbidity. The possible reason is that a high primary production by phytoplankton is linked to high pH concentration due to easy access to the nutrients (e.g. phosphate). In other words, the change of pH may result in algal blooms which eliminate carbon dioxide from the water. The minimum (6.9) and maximum of pH (10.03) were recorded in the sites 1 (the site characterized with the hyper-eutrophic condition) and 6 (the site characterized with the eutrophic condition), respectively. One of the main causes of pH variation

in the Anzali wetland can be the availability of carbonate (positive correlation with pH) or bicarbonate (negative correlation with pH).

Another factor that contributed to the prediction of the eutrophication in the wetland is electric conductivity. As already stated in the correlation analysis, electric conductivity was significantly and positively correlated to the amount of chlorophyll-*a* in the wetland ($p < 0.01$, $r = 0.40$). Model predicted that the higher trophic condition (hyper-eutrophic) may be found in the Anzali wetland where electric conductivity reaches the highest amount. The maximum electric conductivity was found in the site 1 (due to the vicinity of this sampling site to the Caspian Sea). It can be argued that due to the low precipitation and rising air/water temperature particularly in the warm seasons in the Anzali wetland's climate, the amount of evaporation increases from the surface water of the wetland. In addition to this, the vicinity of some sampling sites to the Caspian Sea (e.g. sites 1, 2, and 3), the concentration of electric conductivity increases in the water. Therefore, phytoplankton activities and hence chlorophyll-*a* may show an increase in such conditions.

According to the model, an increase in chlorophyll-*a* (displayed as super and hyper-eutrophic condition on the constructed tree) is associated with an increased concentration of water turbidity. The model's prediction is reasonable because the variables "electric conductivity, chlorophyll-*a*, total phosphate, and total nitrogen" have a close relation with the amount of water turbidity in the Anzali wetland. Earlier studies (e.g. Silvino & Barbosa, 2015) showed that an increase in water turbidity (especially at the lower water depths) may considerably contribute to the water enrichment. In the present research, the mean depth recorded in the whole parts of the Anzali wetland was 1.78 m (the minimum and maximum depths of 0.15 m and 5 m, respectively). Also, the yearly average of the recorded depth in the Anzali wetland varied from 0.65 m (in the central part) to 4.36 m (in the shipping channel). This indicates that the entire wetland is becoming more shallower than the past few years so that such a water appearance may become cloudy or murky at the lower parts of the wetland. The shallower depth of the Anzali wetland may, in turn, have high turbidity characterized with several factors consisting of phytoplankton, erosion, urban runoff, wastewater discharge, algae, and

sediment disruption. Therefore, at the lower depths of the wetland, light can be easily available for the photosynthesis activities of phytoplankton or algae.

According to the model's prediction, an adequate light penetration into the lower waters (Le Moal et al., 2019) together with the availability of the nutrients (e.g. phosphate and nitrate) will, in turn, increase the density of phytoplankton due to sufficient photosynthesis activity.

The results of the predictive model showed that an increase in the concentration of the nutrients such as total phosphate will certainly enhance the amount of eutrophication in the Anzali wetland resulting in a hyper-eutrophic condition in this ecosystem. This was also confirmed by the previous studies (e.g. Le Moal et al., 2019; Silvino & Barbosa, 2015) demonstrating that hyper-eutrophic of the wetland particularly in hot seasons is associated with high concentration of nutrients such as total phosphate and total nitrogen. The impact of enriched waters on the growth of algae and excessive amount of nutrients (eutrophicated water) was already confirmed by Le Moal et al. (2019).

Among the variables used to analyse the TSI in the Anzali wetland, total nitrogen and water transparency were never predicted by the model. As already stated, two nutrients (total phosphate and total nitrogen) were significantly and positively correlated with chlorophyll-*a* in the study wetland (they are also strongly intercorrelated with each other). In such a case, the data-driven methods like the induced CTs model would give the top priority to the most important variable(s) so that total phosphate is more important nutrient than total nitrogen to stimulate algal blooms in wetlands/lakes. For that reason, the nitrate concentration may not be very often used in the TSI analysis in lakes and wetlands (e.g. Bekteshi & Cupi, 2014; Carlson, 1977). That is the possible reason that total nitrogen has not been predicted by the model for evaluating the eutrophication condition in the Anzali wetland. Water turbidity and water transparency were positively and negatively correlated with the chlorophyll-*a* in the wetland, respectively ($p < 0.01$, $r = 0.56$ for water turbidity and $p < 0.01$, $r = -0.40$ for water transparency) so that water turbidity was predicted by the model instead of water transparency.

The outcomes of the present research (according to the one-way ANOVA analysis) showed that the eutrophic condition in the Anzali wetland is influenced by significant variation in the spatial and temporal patterns. Some sites (e.g. 4 and 5) situated in the eastern part of the wetland are highly eutrophicated particularly in hot seasons (e.g. summer season). The main cause of the pollution in these sites is related to the excessive amount of nutrients. Besides, there are various human activities in different parts of the Anzali wetland (Sadeghi et al., 2017). For the last few years, the wetland has been adversely affected by numerous pollutants. These contaminants result from the industrial, urban, and agricultural activities taking place in the surrounding areas (Vesali Naseh et al., 2012; Sadeghi, 2022). Based on this, various levels of pollutants unfavourably influence the sampling sites in different seasons causing the Anzali wetland very polluted (Sadeghi et al., 2014). The most important pollution sources to the Anzali wetland are rivers, domestic sewage, industrial and agriculture wastewater, and hospitals. The wastewaters contain all sorts of pollutants such as heavy metals, nutrients, and precipitates (Sadeghi, 2022). Therefore, the main difference in highly eutrophicated sites (from the eutrophic to hyper-eutrophic) may be linked to the extent of pollutants. In contrast, some sampling sites (e.g. 6, 7, and 8) situated in the western part of the wetland are relatively less polluted (eutrophic condition) than other sites particularly in the cold seasons (e.g. winter). This can be justified that the amount of water transparency in the western part of the wetland is relatively higher than other sites due to deep water.

In summary, to achieve a more reliable prediction about the eutrophication in the Anzali wetland, the applied model can be extended by adding more variables (e.g. biological and chemical oxygen demand and total hardness). Moreover, more sampling sites need to be monitored in the entire Anzali wetland. Thus, adding more sampling sites and measuring some important variables will certainly result in a more reliable prediction for the eutrophication in this valuable ecosystem in future monitoring. This can, in turn, help wetland managers and decision-makers to appropriately protect this valuable ecosystem.

Conclusions

In the present research, the eutrophic condition in the Anzali wetland was initially determined by trophic state index (TSI). Besides the assessment of the TSI, a data-driven technique (CTs model) was

implemented to identify the most important predictors (physical-structural, water quality, and the variables related to the spatial and temporal patterns) for analysing the eutrophication in the wetland. Based on the results of the TSI, the whole parts of Anzali wetland are facing with eutrophication problem. Except the western part of the wetland (classified as a eutrophic), other parts of the wetland are classified as a super and hyper-eutrophic. It is concluded that the most important parameter settings need to be applied for the CT model to successfully predict the trophic condition in wetlands. Besides that, an appropriate model calibration (e.g. finding a best pruning optimization and optimal cross-validation of the model) may improve the model performances and hence select the most important explanatory variables for predicting the eutrophication in the Anzali wetland. Based on the outcomes of the CTs model, it is concluded that water quality variables (e.g. bicarbonate, pH, turbidity, electric conductivity, dissolved oxygen, and total phosphate) and physical-structural factors (e.g. water temperature and water depth) may jointly contribute to the prediction of the trophic conditions in the wetland. Consequently, for the future monitoring and the management of this valuable wetland, one needs to select the predicted variables which were decided upon by the model. Another conclusion derived from the current study (according to ANOVA test) is that the temporal and spatial patterns can influence the eutrophic condition in the Anzali wetland.

Author contribution   First author: data collection and manuscript writing and the proposed applied modelling techniques. The second author: data collection and sampling design. The last author: the proposed modelling techniques and manuscript writing.

Data availability   If article and information are requested, the corresponding author will be available to respond the request.

Declarations

Ethics approval   This study was approved by the Faculty of Natural Resources, University of Guilan.

Consent to participate   All authors declared the consent to participate.

Consent for publication   All authors declared the consent for publication.

Competing interests   The authors declare no competing interests.

References

Abedini, A. (2015). The study of trophic index in Anzali lagoon. Agricultural Research and Education Organization, Iranian Fisheries Science Research Institute. Inland water aquaculture research centre, Tehran, 57 p (In Persian).

Ambelu, A., Lock, K., & Goethals, P. L. M. (2010). Comparison of modelling techniques to predict macroinvertebrate community composition in rivers of Ethiopia. *Ecological Informatics, 5*, 147–152.

APHA, Awwa, WEF. (2012). *Standard methods for the examination of water and wastewater* (22nd ed.). American Public Health Association.

Armengol, X., & Miracle, M. R. (1999). Zooplankton communities in doline lakes and pools, in relation to some bathymetric parameters and physical and chemical variables. *Journal of Plankton Research, 21*(12), 2245–2261.

Bekteshi, A., & Cupi, A. (2014). Use of trophic state index (Carlson, 1977) for assessment of trophic status of the Shkodra lake. *Journal of Environmental Protection and Ecology, 15*, 359–365.

Bennett, E. M., Carpenter, S. R., & Caraco, N. F. (2001). Human impact on erodable phosphorus and eutrophication: A global perspective: Increasing accumulation of phosphorus in soil threatens rivers, lakes, and coastal oceans with eutrophication. *AIBS Bulletin., 51*(3), 227–234.

Beusen, A. H., Bouwman, A. F., Van Beek, L. P., Mogollón, J. M. & Middelburg, J. J. (2016). Global riverine N and P transport to ocean increased during the 20th century despite increased retention along the aquatic continuum. *Biogeosciences, 13*(8):2441

Bordalo, A., Nilsumranchit, W., & Chalermwat, K. (2001). Water quality and uses of the Bangpakong River (Eastern Thailand). *Water Research., 35*, 3635–3642.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.

Carlson, R. E. (1977). A trophic state index for lakes. *Limnology and Oceanography., 22*, 361–369.

Cloern, J. E. (2001). Our evolving conceptual model of the coastal eutrophication problem. *Marine Ecology Progress Series., 210*, 223–253.

Cohen, J. (1960). A coefficient of agreement for nominal scale. *Educational and Psychological Measurement, 20*, 37–46.

De'ath, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology, 81*, 3178–3192.

de Jonge, V. N., Elliott, M., & Orive, E. (2002). Causes, historical development, effects and future challenges of a common

environmental problem: Eutrophication. *Hydrobiologia, 475*(1), 1–19.

D'heygere, T., Goethals, P.L.M. & De Pauw, N. (2003). Use of genetic algorithms to select input variables in decision tree models for the prediction of benthic macroinvertebrates. *Ecological Modelling, 160*, 291–300.

Deyab, M., El-Adl, M., Ward, F., & Omar, E. (2021). Trophic status, phytoplankton diversity, and water quality at Kafr El-Shinawy drinking-water treatment plant, Damietta. *Journal of Water Supply: Research and Technology-Aqua., 70*(3), 342–360.

Dodds, W. K., & Cole, J. J. (2007). Expanding the concept of trophic state in aquatic ecosystems: It's not just the autotrophs. *Aquatic Sciences, 69*(4), 427–439.

Du, H., Chen, Z., Mao, G., Chen, L., Crittenden, J., Lie, Y. M., & R. & Chai, L. (2019). Evaluation of eutrophication in freshwater lakes: A new non-equilibrium statistical approach. *Ecological Indicators, 102*, 686–692.

Fraterrigo, J. M., & Downing, J. A. (2008). The influence of land use on lake nutrients varies with watershed transport capacity. *Ecosystems, 11*(7), 1021–1034.

JICA. (2005). The Study on Integrated Management of the Anzali Wetland in the Islamic Republic of Iran-Final Report, p. 182. Available online: https://openjicareport.jica.go.jp/pdf/11784097.pdf (Accessed on 11 May 2021).

JICA. (2019). Mid-term plan for conservation of the Anzali wetland for 2020–2030 (prepared under the Anzali wetland ecological management project-phase 2). 81pp.

Hamilton, P. B., Gajewski, K., Atkinson, D. E., & Lean, D. R. S. (2001). Physical and chemical limnology of 204 lakes from the Canadian Arctic Archipelago. *Hydrobiologia, 457*(1–3), 133–148.

Hassanzadeh, M., Zarkami, R., & Sadeghi, R. (2021). Uptake and accumulation of heavy metals by water body and *Azolla filiculoides* in the Anzali wetland. *Applied Water Science., 11*, 91.

Hecky, R. E., & Kilham, P. (1988). Nutrient limitation of phytoplankton in freshwater and marine environments: A review of recent evidence on the effects of enrichment. *Limnology and Oceanography., 33*, 796–822.

Lamparelli, M. C. (2004). Trophic status in São Paulo State water bodies: Evaluation of monitoring methods. Dissertation, University of São Paulo, São Paulo (Brazil).

Le Moal, M., & Gascuel-Odoux., Ménesguen, A., Souchon, Y., Etrillard, C., Levain, A., Moatar, F., Pannard, A., Souchu, P., Lefevre, A. & Pinay, G. (2019). Eutrophication: A new wine in an old bottle? *Science of the Total Environment., 651*, 1–19.

Liu, W., Zhang, Q., & Liu, G. (2010). Lake eutrophication associated with geographic location, lake morphology and climate in China. *Hydrobiologia, 644*(1), 289–299.

Lock, K., & Goethals, P. (2012). Habitat suitability modelling for mayflies (Ephemeroptera) in Flanders (Belgium). *Ecological Informatic, 17*, 30–35.

Mortazavi, S. (2018). Survey of modified hazard Quotient, potential ecological risk factor and toxicity units of heavy metals in surface sediments of some wetlands of Iran. *Archives of Hygiene Sciences, 7*, 251–263.

Nazarhaghighi, F., Timm, T., Nadoushan, R. M., Shabanipour, N., Fatemi, M. R., & Moradi, A. M. (2014). Oligochaetes (Annelida, Clitellata) in the Anzali International Wetland, north-western Iran. *Estonian Journal of Ecology, 63*, 130–144.

Noges, T. (2009). Relationships between morphometry, geographic location and water quality parameters of European lakes. *Hydrobiologia, 633*(1), 33–43.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*, 81–106.

Sadeghi, R. (2022). Habitat suitability models for control of the invasive fern species *Azolla filiculoides* in the Anzali wetland (Iran) Ghent University 256.

Sadeghi, R., Forio, M. A. E., Ho, L. T., & Goethals, P. L. M. (2021). Evidence-based management of the Anzali wetland system (Northern Iran) based on innovative monitoring and modelling methods. *Sustainability, 13*, 5503.

Sadeghi, R., Zarkami, R., Sabetraftar, K., & Van Damme, P. (2012a). Use of support vector machines (SVMs) to predict distribution of an invasive water fern *Azolla filiculoides* (Lam.) in Anzali wetland, southern Caspian Sea, Iran. *Ecological Modelling, 244*, 117–126.

Sadeghi, R., Zarkami, R., Sabetraftar, K., & Van Damme, P. (2012b). Application of classification trees to model the distribution pattern of a new exotic species *Azolla filiculoides* (Lam.) in Selkeh Wildlife Refuge, Anzali wetland, Iran. *Ecological Modelling, 243*, 8–17.

Sadeghi, R., Zarkami, R., Sabetraftar, K., & Van Damme, P. (2013). Application of genetic algorithm and greedy stepwise to select input variables in classification tree models for the prediction of habitat requirements of *Azolla filiculoides* (Lam.) in Anzali wetland, Iran. *Ecological Modelling, 251*, 44–53.

Sadeghi, R., Zarkami, R., & Van Damme, P. (2014). Modelling habitat preference of an alien aquatic fern, *Azolla filiculoides* (Lam.), in Anzali wetland (Iran) using data-driven methods. *Ecological Modelling, 284*, 1–9.

Sadeghi, R., Zarkami, R., & Van Damme, P. (2017). Analysing the occurrence of an invasive aquatic fern in wetland using data-driven and multivariate techniques. *Wetlands Ecology and Management, 25*, 485–500.

Silvino, R. F., & Barbosa, F. A. R. (2015). Eutrophication potential of lakes: An integrated analysis of trophic state, morphometry, land occupation, and land use. *Brazilian Journal of Biology, 75*(3), 607–615.

Smith, V. H. (2003). Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environmental Science and Pollution Research, 10*(2), 126–139.

Smith, V. H., Tilman, G. D., & Nekola, J. C. (1999). Eutrophication: Impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environmental Pollution, 100*(1), 179–196.

Taranu, Z. E., & Gregory-Eaves, I. (2008). Quantifying relationships among phosphorous, agriculture, and lake depth at an inter-regional scale. *Ecosystems, 11*(5), 715–725.

Vesali Naseh, M. R., Karbassi, A., Ghazaban, F., & Baghvand, A. (2012). Evaluation of heavy metal pollution in Anzali wetland, Guilan. *Iran. Iranian Journal of Toxicology, 5*(15), 565–576.

Vitousek, P. M., Aber, J. D., Howarth, R. W., Likens, G. E., Matson, P. A., Schindler, D. W., Schlesinger, W. H., & Tilman, D. G. (1997). Human alteration of the global nitrogen cycle: Sources and consequences. *Ecological Applications, 7*(3), 737–750.

Witten, I. H., Frank, E. & Mark, A. (2011). 'Data mining: Practical machine learning tools and techniques.' 3rd edn. (Morgan Kaufmann: San Francisco) 629.

Xu, F. L., Tao, S., Dawson, R. W., & Li, B. G. (2001). A GIS-based method of lake eutrophication assessment. *Ecological Modelling, 144*, 231–244.

Zarkami, R. (2016). Invasive aquatic plants. Haghshanas publishing. 184 p. (in Farsi).

Zarkami, R., Darizin, Z., Sadeghi, R., Bani, A., & Ghane, A. (2019). Use of data-driven model to analyse the occurrence patterns of an indicator fish species in river: A case study for *Alburnoides eichwaldii* (De Filippi, 1863) in Shafaroud River, north of Iran. *Ecological Engineering, 133*, 10–19.

Zarkami, R., Esfandi, J., & Sadeghi, R. (2021). Modelling occurrence of invasive water hyacinth (*Eichhornia crassipes*) in wetlands. *Wetlands, 41*, 8.

Zhou, Y., Wang, L., Zhou, Y., & Mao, X. Z. (2020). Eutrophication control strategies for highly anthropogenic influenced coastal waters. *Science of the Total Environment, 705*, 135760.