Environmental Science and Ecotechnology 17 (2024) 100305

Contents lists available at ScienceDirect

Environmental Science and Ecotechnology

journal homepage: www.journals.elsevier.com/environmental-science-and-ecotechnology/

Original Research

Establishment risk of invasive golden mussel in a water diversion project: An assessment framework

Yao Yang [a], Mengzhen Xu [a, *], Xingyu Chen [a], Jiahao Zhang [a], Shulei Wang [b], Jianying Zhu [c], Xudong Fu [a]

[a] State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing, 100084, China
[b] China South-to-north Water Diversion Corporation Limited, China
[c] Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Article history:
Received 17 January 2023
Received in revised form
18 July 2023
Accepted 22 July 2023

Keywords:
Golden mussel
Environmental variables
Multiclass classification
Logistic regression
Quantitative risk assessment

ABSTRACT

Inter-basin water diversion projects have led to accelerated colonization of aquatic organisms, including the freshwater golden mussel (*Limnoperna fortunei*), exacerbating global biofouling concerns. While the influence of environmental factors on the mussel's invasion and biofouling impact has been studied, quantitative correlations and underlying mechanisms remain unclear, particularly in large-scale inter-basin water diversion projects with diverse hydrodynamic and environmental conditions. Here, we examine the comprehensive impact of environmental variables on the establishment risk of the golden mussel in China's 1432-km-long Middle Route of the South-to-North Water Diversion Project. Logistic regression and multiclass classification models were used to investigate the environmental influence on the occurrence probability and reproductive density of the golden mussel. Total nitrogen, ammonia nitrogen, water temperature, pH, and velocity were identified as crucial environmental variables affecting the biofouling risk in the project. Logistic regression analysis revealed a negative correlation between the occurrence probability of all larval stages and levels of total nitrogen and ammonia nitrogen. The multiclass classification model showed that elevated levels of total nitrogen hindered mussel reproduction, while optimal water temperature enhanced their reproductive capacity. Appropriate velocity and pH levels were crucial in maintaining moderate larval density. This research presents a quantitative analytical framework for assessing establishment risks associated with invasive mussels, and the framework is expected to enhance invasion management and mitigate biofouling issues in water diversion projects worldwide.

1. Introduction

The natural dispersal of species into new environments typically occurs within limited ranges due to geographical barriers. However, human activity has significantly expedited species dispersal both spatially and temporally [1,2]. The invasive freshwater bivalve *Limnoperna fortunei*, commonly known as the golden mussel, originally native to southeast Asia, has rapidly spread to South America since the early 1990s, primarily through ballast water transport. The golden mussel possesses several traits that contribute to its invasive potential, including a small body size, rapid development, high reproductive rate or fecundity, and short lifespan [3,4]. Its life cycle comprises two stages: planktonic larval stages and adult stages attached to substrates. The larvae, measuring less than 2 mm, possess passive dispersal capabilities, allowing for long-distance transportation. Dense attachment of adult mussels on natural or artificial substrates leads to significant biofouling issues, resulting in substantial socio-economic losses and ecological risks. Considered one of the most invasive freshwater species, the golden mussel demonstrates a remarkable ability to invade and thrive in freshwater ecosystems [5–9].

Moreover, human activities significantly contribute to the accelerated spread of the golden mussel. Water diversion projects, aimed at addressing the uneven distribution of water resources, have unintentionally facilitated the invasion of this species. Across different regions, these water diversion projects have played a

* Corresponding author.
E-mail address: mzxu@mail.tsinghua.edu.cn (M. Xu).

https://doi.org/10.1016/j.ese.2023.100305

prominent role in promoting the widespread distribution of the golden mussel within inland water bodies [10–12]. By providing concrete surfaces, these projects create favorable conditions for the settlement of golden mussel larvae [8]. Once established, the golden mussel assumes the role of an “ecological engineer,” leading to biofouling issues in water supply systems [13].

The Middle Route of the South-to-North Water Diversion Project (MRSNWDP), which spans a total length of 1432 km, is recognized as the largest water transfer project globally. It originates from the Danjiangkou Reservoir and extends to Beijing and Tianjin [14]. This project serves as a significant “invasion highway” that connects distinct biogeographical regions, enabling the spread of the golden mussel [10,11]. Shortly after the commencement of the MRSNWDP, a substantial number of golden mussels, encompassing larvae and adults, were discovered along the water transfer channels [10,12]. This occurrence is attributed to the presence of the golden mussels within the Danjiangkou Reservoir, which acts as the source for their invasion from southern to northern regions of China [10,15]. Notably, the rapid invasion of mussels in water transfer projects extends beyond the MRSNWDP and encompasses other endeavors such as the Dongjiang Water Source Diversion Project in China [16]. The biofouling caused by the golden mussel exerts detrimental effects by corroding concrete walls, increasing flow resistance, impairing water conveyance capability, and damaging structures [16,17]. Moreover, golden mussel biofouling leads to water quality contamination through respiration and postmortem decay processes [4,18].

The process through which species invade new habitats with human assistance typically involves transport, introduction, establishment (including survival and reproduction), and spread (including dispersal and environmental adaptation) [19]. In the case of the MRSNWDP, the introduction of golden mussels is facilitated by the breakdown of geographic barriers caused by water diversion [10,11]. Assessing the risks associated with invasion and biofouling requires a focus on the species' ability to survive and reproduce under the environmental conditions present in the water diversion project. Local abiotic factors such as water temperature, flow velocity, total nitrogen, and dissolved oxygen concentrations may impact the attachment density and growth of golden mussels [13,20–23]. However, current knowledge on these environmental limits primarily relies on observations of adult mussels. It is important to note that larvae may exhibit greater sensitivity to certain variables compared to adults, and there is limited information on how environmental factors directly influence larval stages, which in turn affects adult density and invasion risk. Furthermore, most previous studies have concentrated on natural rivers, with limited research on mussel invasion dynamics in water diversion projects [24]. Therefore, for effective invasive species management in the MRSNWDP, it is crucial to gather reliable environmental and mussel-specific data within the project's unique context and develop a comprehensive scientific understanding of the golden mussels' complete life cycle [25].

Our study aims to quantify the relationship between environmental factors and the establishment risk of the golden mussel, specifically focusing on barriers related to survival and reproduction during the establishment stage. To achieve these objectives, we have outlined the following approach: (1) Simultaneous Data Collection: We will collect environmental and biological data concurrently from the MRSNWDP, allowing us to examine the correlation between key environmental factors and indicators of survival conditions at different larval stages, as well as the overall larval density category indicating reproduction conditions. (2) Statistical Analysis: We will utilize logistic regression and multiclass classification models to analyze the relationship between the identified key environmental factors and the probability of occurrence, reflecting survival conditions at different larval stages. Additionally, we will examine the relationship between these environmental factors and the overall larval density category, which serves as an indicator of reproduction conditions. (3) Biofouling Risk Assessment: Our assessment will focus on evaluating the risk of golden mussel biofouling within the MRSNWDP. We will particularly investigate the relationship between environmental variables and the establishment stage. This assessment will consider two critical aspects of the establishment stage: survival conditions (indicated by the occurrence probability) and reproduction conditions (indicated by overall larval density). By employing this quantitative risk assessment analytical framework, our study has the potential to enhance the management of mussel invasion and biofouling not only in the MRSNWDP but also in other inter-basin water diversion projects worldwide. With over 200 such projects globally, our findings can be applied internationally to improve invasive mussel management strategies.

2. Methods and materials

2.1. Sampling

The Middle Route of the South-to-North Water Diversion Project (MRSNWDP) is a significant infrastructure that effectively allocates water resources in China, catering to a population of 85 million in major northern cities such as Beijing, Tianjin, and Shijiazhuang. However, the presence of the golden mussel in the MRSNWDP water source, the Danjiangkou Reservoir, poses a high risk of biofouling. To assess this risk, 16 sampling stations were established along the 1432-km-long MRSNWDP, and environmental variables were measured. Golden mussel larvae were collected during their breeding season, which spanned from July 2017 to October 2018 (Fig. S1).

The collection of golden mussel larvae involved filtering 100 L of water through a 64-μm mesh plankton net. The filtered materials were preserved in 4% formaldehyde solution and left for 48 h before sorting under an optical microscope in the laboratory. Larvae were identified based on their developmental stages, including D-shaped, umbonated, pediveliger, or plantigrade stages [26,27] (Fig. S2).

Environmental variables were measured on a monthly basis from July 2017 to July 2021. These data were provided by the Construction and Administration Bureau (CAB) of the Middle Route of the South-to-North Water Diversion Project in China (Table S1). The environmental data collected before October 2018 were used for modeling in conjunction with larval density, while the remaining data were used for risk assessment at representative stations. Detailed information regarding both larval density and environmental variables from July 2017 to October 2018 is presented (Fig. 1).

2.2. Data processing

The data-processing methodology is illustrated (Fig. 2). A total of 208 samples were collected, encompassing both larval density and environmental variable datasets. The aim of the analysis was to investigate the relationship between environmental variables and the establishment of the golden mussel. In the evaluation of establishment risk, two key factors were considered: the occurrence probability, which indicates survival conditions, and the overall larval density, which reflects reproductive density.

To address discrepancies arising from variations in units and ranges among the input data, the independent variables were normalized using Z-score normalization. This normalization process ensured that all data fell within a specific interval, with a mean (μ) of 0 and a standard deviation (σ) of 1. Moreover, to facilitate the

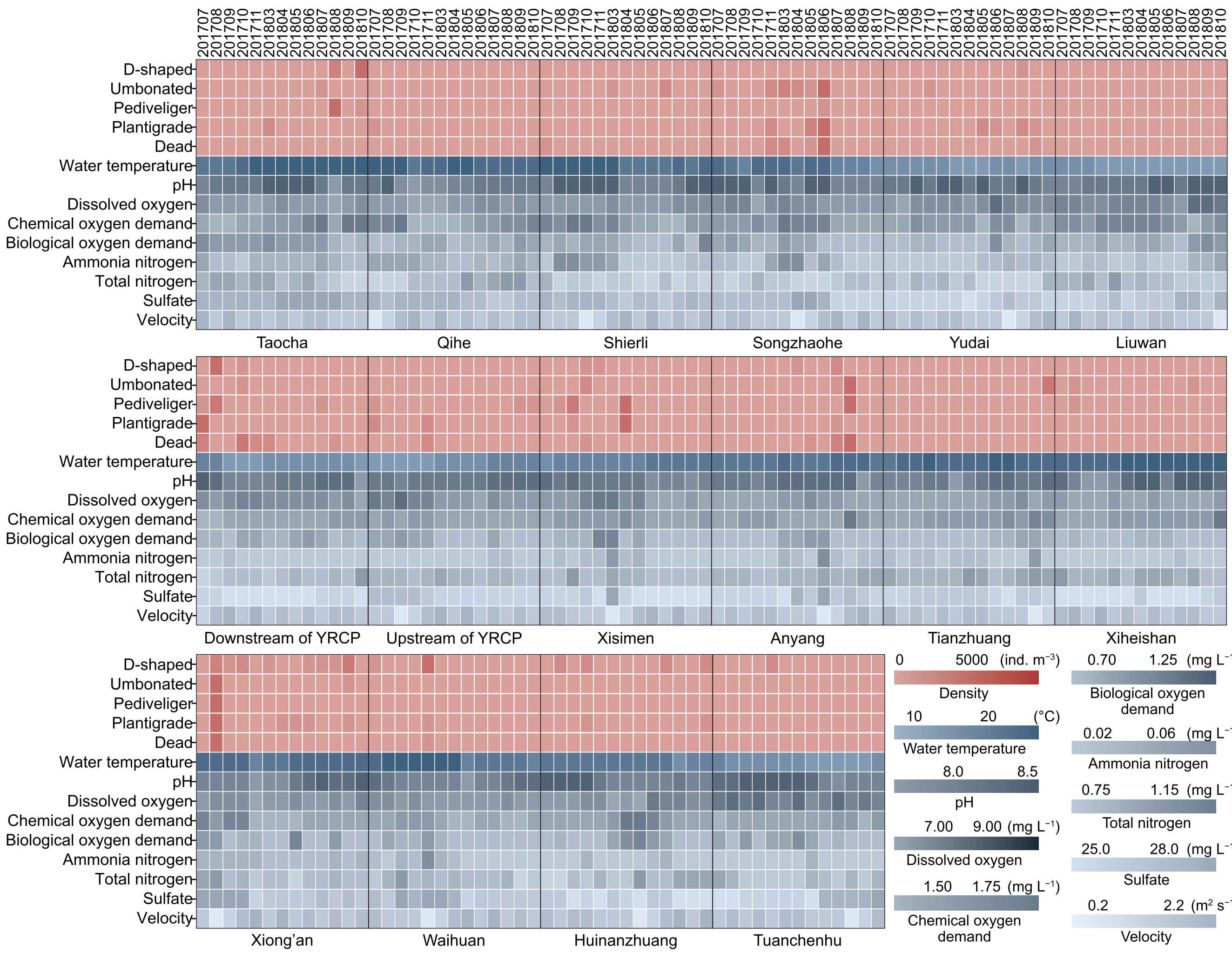

Fig. 1. Larval density of *Limnoperna fortunei* and environmental variables collected from sampling stations.

use of environmental variables in logistic regression and multiclass classification models, these independent factors were transformed into new variables (Table 1).

2.2.1. Global Moran's index

The spatial autocorrelation analysis method, specifically the global Moran's index statistic, is an effective approach for examining the spatial autocorrelation of the same variable across different observation objects [28]. The global Moran's index ranges between −1 and 1 and provides insights into the nature of spatial autocorrelation. A value of 0 indicates no spatial autocorrelation, while values between 0 and 1 suggest positive spatial autocorrelation, and values between −1 and 0 indicate negative spatial autocorrelation. The closer the absolute value of global Moran's index is to 1, the stronger the spatial autocorrelation. The calculation of global Moran's index is performed as follows:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(x_i - x)(x_j - x)}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij} \times \sum_{i=1}^{n}(x_i - x)^2} \tag{1}$$

In equation (1), n represents the number of subjects being analyzed. The variables x_i and x_j denote the spatial properties of objects in regions i and j, respectively. The variable x represents the mean value of all samples. The spatial weight matrix element, W_{ij}, captures the spatial relationship between different spatial objects. The weight matrix is typically calculated based on a distance matrix, where the diagonal elements are set to 0, and the cell elements are defined as the reciprocal of the distance. In this study, the Euclidean distance is utilized to compute the weight matrix.

After calculating the Global Moran's index, it is important to verify its statistical significance using the Z-test [29]. The Z-test helps determine whether the observed Moran's I value is statistically different from what would be expected under the null hypothesis of no spatial autocorrelation. At a significance level of 0.05, a Moran's I index is considered significant if the absolute value of Z is greater than 1.96 and the corresponding p-value is less than 0.05. If both conditions are met, it indicates rejection of the null hypothesis and suggests the presence of significant spatial autocorrelation in the data. On the other hand, if the absolute value of Z is less than or equal to 1.96 or the p-value is greater than or equal to 0.05, the null hypothesis cannot be rejected, suggesting no significant spatial autocorrelation.

2.2.2. Logistic regression

In this study, we employed binary logistic regression [30] to examine the relationships between environmental factors and the occurrence probability of different larval stages, which served as an indicator of larval survival. The calculation of the occurrence

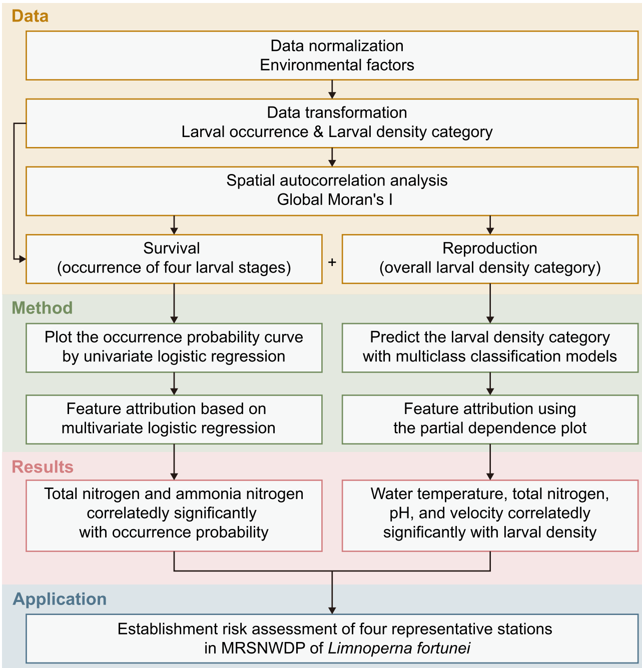

Fig. 2. Methodology flowchart.

probability for each larval stage involved a two-step process. First, univariate logistic regression was conducted to identify relevant environmental variables, while excluding potentially irrelevant ones. Subsequently, a collinearity test was performed to select key environmental factors for subsequent multivariate logistic regression. The collinearity test utilized two criteria: tolerance and the variance inflation factor (VIF) [31]. Generally, collinearity is considered absent when the tolerance (Tol) is greater than 0.1 or the VIF is less than 10. These tests ensured that the selected environmental variables were not highly correlated, allowing for a reliable analysis of their individual contributions. In order to identify the key environmental variables, multivariate logistic regression was conducted, considering the effects of multiple factors simultaneously. The significance of the regression models was assessed using a combination of the omnibus test [32] and the Hosmer-Lemeshow (HL) test [33]. These tests provided measures of the overall model fit and goodness-of-fit, respectively, enhancing the reliability of the logistic regression analysis.

2.2.3. Multiclass classification models

In this study, the relationships between environmental variables and the total larval density category, which serves as an indicator of reproduction condition, were analyzed using multiclass classification models. To select important features for these models, recursive feature elimination was employed. This process helps identify the most relevant environmental variables by iteratively eliminating less important features. To determine the optimal number of input features for achieving the best model performance, the number of selected environmental variables ranged from 1 to 9. Various multiclass classification models were utilized (Table S2), including K-Nearest Neighbor (KNN) [34], Random Forest (RF) [35], Naïve Bayes [36], Gradient Boosting Decision Tree (GBDT) [37], eXtreme Gradient Boosting (XGBoost) [38], and Light Gradient Boosting Machine (LightGBM) [39]. The hyperparameters of these models were tuned using a five-fold cross-validation approach to optimize their performance.

To evaluate the performance of the multiclass classification models, four established error metrics were utilized: recall, accuracy, precision, and F1-score. These metrics provide comprehensive measures of model performance, considering aspects such as the correct identification of positive cases (recall), overall accuracy, precision in predicting positive cases, and the harmonic mean of precision and recall (F1-score). The calculation formulas for the evaluation indicators used in the multiclass classification models are as follows:

$$\mathrm{Recall} = \sum\nolimits_{i=0}^{2} \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FN}_i} \times \frac{\mathrm{N}_i}{\sum_{i=0}^{2} \mathrm{N}_i} \tag{2}$$

$$\mathrm{Precision} = \sum\nolimits_{i=0}^{2} \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FP}_i} \times \frac{\mathrm{N}_i}{\sum_{i=0}^{2} \mathrm{N}_i} \tag{3}$$

$$\mathrm{Accuracy} = \frac{\sum_{i=0}^{2}\mathrm{TP}_i + \sum_{i=0}^{2}\mathrm{TN}_i}{\sum_{i=0}^{2}\mathrm{TP}_i + \sum_{i=0}^{2}\mathrm{FP}_i + \sum_{i=0}^{2}\mathrm{TN}_i + \sum_{i=0}^{2}FN_i} \tag{4}$$

$$\mathrm{F1} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{5}$$

where TP_i is the number of positive samples predicted as positive; TN_i is the number of negative samples predicted as negative; FP_i is the number of negative samples predicted as positive; FN_i is the number of positive samples predicted as negative; N_i is the number of samples in different classes. These error metrics allow for a comprehensive evaluation of the performance of the multiclass classification models. The evaluation process involves comparing the values of these metrics to optimize the models. In the evaluation, the first priority is given to recall. This is because it is important to minimize false negative predictions, ensuring that positive samples are correctly identified as positive. By prioritizing recall, the models aim to avoid missing any positive instances. The multiclass classification model that best fits the relationship between the environmental variables and larval density can be determined by evaluating these error metrics. By analyzing and comparing the values of recall, accuracy, precision, and F1-score, the model with the highest overall performance can be identified.

2.2.4. Partial dependence plot

The partial dependence plot illustrates the marginal effect that specific features have on the predicted outcome of a machine learning model. It provides insights into the relationship between a target variable and a particular feature, revealing whether the

Table 1
Transformations applied to density data.

Occurrence of larvae at each stage		Overall larval density category	
Logistic	Range	Multi-categories	Range (ind. m^{-3})
0	Larval absent	0: Low	<100
1	Larvae present	1: Medium	100–1000
		2: High	≥1000

relationship is linear, monotonic, or more intricate in nature. In regression analysis, the partial dependence function is defined as follows:

$$\widehat{f}_S(x_S) = E_{X_C}[\widehat{f}(x_S, X_C)] = \int \widehat{f}(x_S, X_C) d\mathbb{P}(X_C) \quad (6)$$

In the partial dependence function, x_S represents the features for which we want to plot the partial dependence, while X_C represents the other features used in the multiclassification model, which are treated as random variables. The set S includes the features of interest for which we want to understand their effect on the prediction. The combined feature vectors x_S and X_C form the total feature space. The partial dependence approach works by averaging the multiclassification model's output over the distribution of features in set C. This allows us to examine the relationship between the features in set S and the predicted results. By marginalizing the other features, we can obtain functions that solely depend on the features in set S, including any interactions with other features. To estimate the partial dependence function, the Monte Carlo method is often employed [40]. This involves generating random samples from the distribution of features in set C and calculating the average model output for each value of the features in set S. By repeating this process multiple times, an estimate of the partial dependence function can be obtained. By utilizing partial dependence plots and estimating the partial dependence function, we can gain insights into the relationships between specific features and the predicted outcomes of the multiclassification model, even when considering the effects of other features in the model.

2.2.5. Biofouling risk assessment

According to the unified framework for biological invasions [19], the invasive golden mussel risk assessment framework was established for water diversion projects, focusing on the MRSNWDP. The historical sampled data have provided evidence that invasive mussels have overcome geographical and introduction barriers in the MRSNWDP, highlighting the importance of evaluating the establishment stage [10–12], By focusing on larval survival and reproduction, the risk assessment framework addresses key establishment stage of the invasion process (Fig. 3).

To characterize the survival situation of larvae, a binary variable representing the existing condition of larvae was used. Although it is challenging to collect dead larvae settled at the bottom during larval sampling, the presence or absence of larvae can provide insights into their survival [41,42]. Additionally, the total larval density was used as an indicator of larval reproductive condition [43,44]. To establish a stable relationship between larval density and environmental conditions, the total larval density was classified into different categories, namely low, medium, and high. This classification allowed for the evaluation of larval reproduction in relation to environmental factors. Logistic regression was employed to fit the relationship between the larval occurrence probability and environmental factors, providing insights into larval survival. Multiclass classification models, including K-Nearest Neighbor (KNN), Random Forest (RF), Naïve Bayes, Gradient Boosting Decision Tree (GBDT), eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), were utilized to quantify the relationship between different categories of total larval density and environmental factors, enabling the evaluation of larval reproduction.

The risk assessment framework considered areas where the larval occurrence probability exceeded 50% and the total larval density exceeded 100 ind. m^{-3} as high-risk sites for mussel invasion. These areas were deemed highly suitable for the complete development and sustained reproduction of the mussels, indicating a high likelihood of establishing a new habitat. Representative sites were selected along the MRSNWDP route, including Taocha Station in Henan Province, Xiheishan Station in Hebei Province, Waihuan Station in Tianjin, and Huinanzhuang Station in Beijing. These sites provided comprehensive coverage across different provinces and municipalities involved in the water diversion project. By applying this comprehensive invasive golden mussel risk assessment framework, it becomes possible to effectively evaluate the invasion risk and guide management strategies in water diversion projects, such as the MRSNWDP.

3. Results

3.1. Spatial-temporal pattern of the sampling results

The statistical characteristics of the collected data on larvae and environmental factors were analyzed (Table 2).

The spatiotemporal analysis of environmental factors (Fig. 4a and b) reveals distinctive trends. Specifically, water temperature and dissolved oxygen exhibit a seasonal pattern, while chemical oxygen demand demonstrates a slight upward trend over time. In contrast, sulfate shows an overall decreasing trend over time. Moreover, downstream dissolved oxygen values are marginally higher compared to upstream values. Additionally, the pH value shows a slight increase from upstream to downstream, while the total nitrogen value slightly decreases spatially from upstream to downstream. The water diversion project operates in accordance with the prescribed flow velocity as per its design. Consequently, the design flow rate of each sample point in the MRSNWDP was utilized, as it remains constant without any temporal fluctuations. The spatial-temporal velocity trend for the MRSNWDP sample points has been excluded from the analysis.

The survival condition of larvae at four developmental stages varied over time (Fig. 4c) and space (Fig. 4d). During the sampling period in 2017, the boxplot of total larval density reveals that the average overall larval density was high in July, September, and October, with relatively high outliers during these months. Moreover, at each station, Taocha Station exhibited only a low average larval density category, while downstream stations, including Tianzhuang Station and Xiheishan Station in Hebei, Xiong'an Station and Waihuan Station in Tianjin, and Huinanzhuang Station in Beijing, displayed higher average overall larval density. Notably, stations with high reproduction levels had exceptionally high outliers of larval density.

Spatial autocorrelation analysis was conducted to investigate whether larval survival and reproduction conditions were influenced by upstream larvae. The Global Moran's index was calculated for this purpose (Fig. 5). The results of the significance test for the Global Moran's index indicated a failure, indicating that there is no significant spatial autocorrelation between the survival and reproduction conditions of larvae at different sampling stations. This suggests that the establishment of larvae is primarily influenced by local environmental conditions.

3.2. Environmental variables influencing occurrence probability

3.2.1. Key environmental variables selected by univariate logistic regression

Based on the results of the omnibus and HL tests of univariate logistic regression (Fig. 6a), several environmental variables, namely water temperature, velocity, chemical oxygen demand, total nitrogen, ammonia nitrogen, biological oxygen demand, and sulfate, showed significant correlations with larval occurrence at various stages.

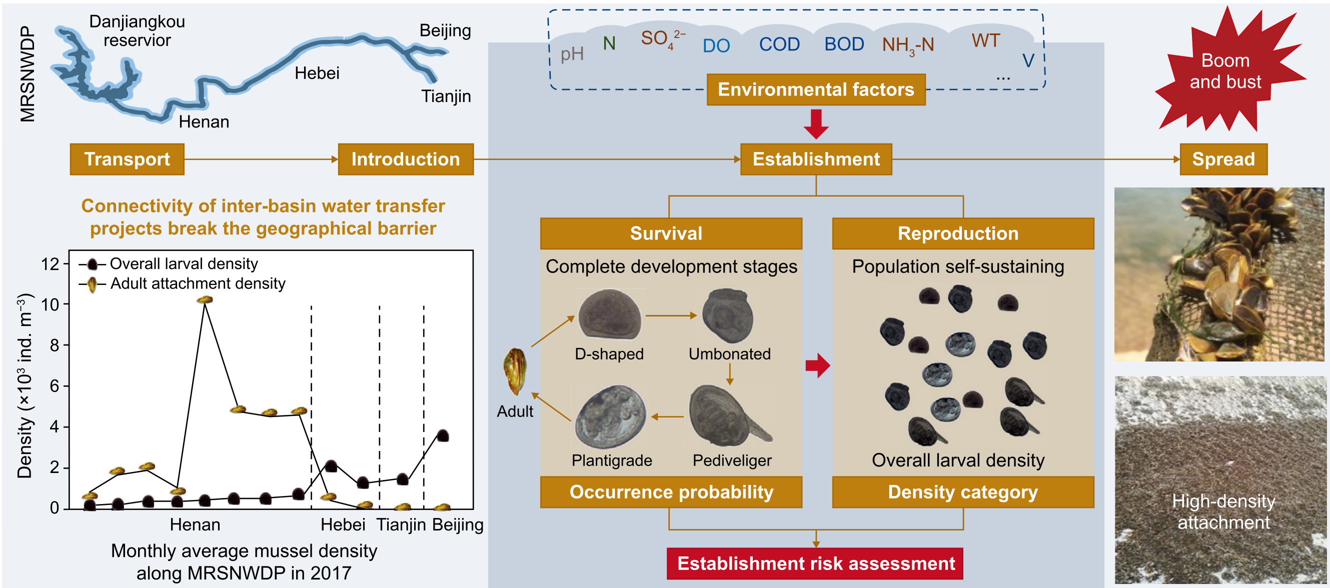

Fig. 3. *Limnoperna fortunei* risk of establishment in the MRSNWDP.

Table 2
Statistical characteristics of sampling results.

Factors	Mean	Median	Range	Factors	Mean	Median	Range
D-shaped larvae density (ind. m^{-3})	202	4	0–16453	Chemical oxygen demand (mg L^{-1})	1.90	1.90	1.40–2.80
Umbonated density (ind. m^{-3})	677	33	0–50277	Biological oxygen demand (mg L^{-1})	0.85	0.70	0.20–2.60
Pediveliger density (ind. m^{-3})	155	17	0–4205	Dissolved oxygen (mg L^{-1})	9.18	9.00	6.40–14.90
Plantigrade density (ind. m^{-3})	158	8	0–6051	Total nitrogen (mg L^{-1})	1.11	1.12	0.65–1.62
Dead veliger density (ind. m^{-3})	293	25	0–16094	Ammonia nitrogen (mg L^{-1})	0.04	0.04	0.03–0.17
Overall larval density (ind. m^{-3})	1486	265	0–60966	Sulfate (mg L^{-1})	28.3	28.4	23.6–34.0
Water Temperature (°C)	22.44	23.60	5.20–32.40	Velocity (m s^{-1})	1.52	1.51	0.28–2.21
				pH	8.24	8.2	7.7–8.7

Further analysis using univariate logistic regression (Fig. 6b) indicated that high water temperature positively influenced the occurrence probability of D-shaped larvae, with a probability exceeding 50% when the water temperature exceeded 20 °C. Conversely, the survival of umbonated and pediveliger larval stages displayed a negative correlation with chemical oxygen demand concentrations, with the probability decreasing to 50% at concentrations above 2.6 mg L−1 and increasing to 80% at concentrations around 1.8 mg L−1. Additionally, biological oxygen demand and sulfate were found to inhibit the occurrence probability of umbonated larvae, with the probability falling below 50% at biological oxygen demand levels above 1.3 mg L−1 and sulfate concentrations exceeding 30 mg L−1. Moreover, the occurrence probability of all four larval stages displayed a negative correlation with ammonia nitrogen and total nitrogen. When ammonia nitrogen levels exceeded 0.4 mg L−1 or total nitrogen levels exceeded 1 mg L−1, the larval occurrence probability could decrease to 50%.

3.2.2. *Attribution of environment variables to the occurrence probability*

For each larval stage, the collinearity test results for factors that passed the univariate logistic regression are shown (Table 3).

The univariate logistic regression identified environmental factors for each larval stage, and subsequent tests for collinearity (tolerance and variance inflation factor) indicated that these factors were not collinear. Therefore, a multivariate logistic regression model was performed, considering independent variables with significant tests. The analysis revealed that total nitrogen significantly influenced the occurrence probability of all four larval stages (Fig. 7). Additionally, ammonia nitrogen was found to have a significant negative effect on the occurrence probability of D-shaped, umbonated, and plantigrade larvae.

The relationship between key environmental factors and the occurrence probability for each larval stage is summarized (Table 4), based on the multivariable logistic regression and significant tests.

3.3. *Environmental variables influencing density category*

3.3.1. *Density category predicted from multiclass classification models*

The selection of effective variables is vital in optimizing the performance of a multiclass classification model, as the impact of each variable may vary. Among the models trained using different numbers of environmental variables selected through recursive feature elimination, the Random Forest multiclassification model utilizing four key independent variables—water temperature, velocity, pH, and total nitrogen—yielded the highest performance in identifying relationships between larval density category and the environmental factors (Fig. 8). The model achieved a recall of 81%, an F1-score of 80%, an accuracy of 81%, and a precision of 81%.

3.3.2. *Attribution of environment variables to density category*

The confusion matrix (Fig. 9a) and the Receiver Operating

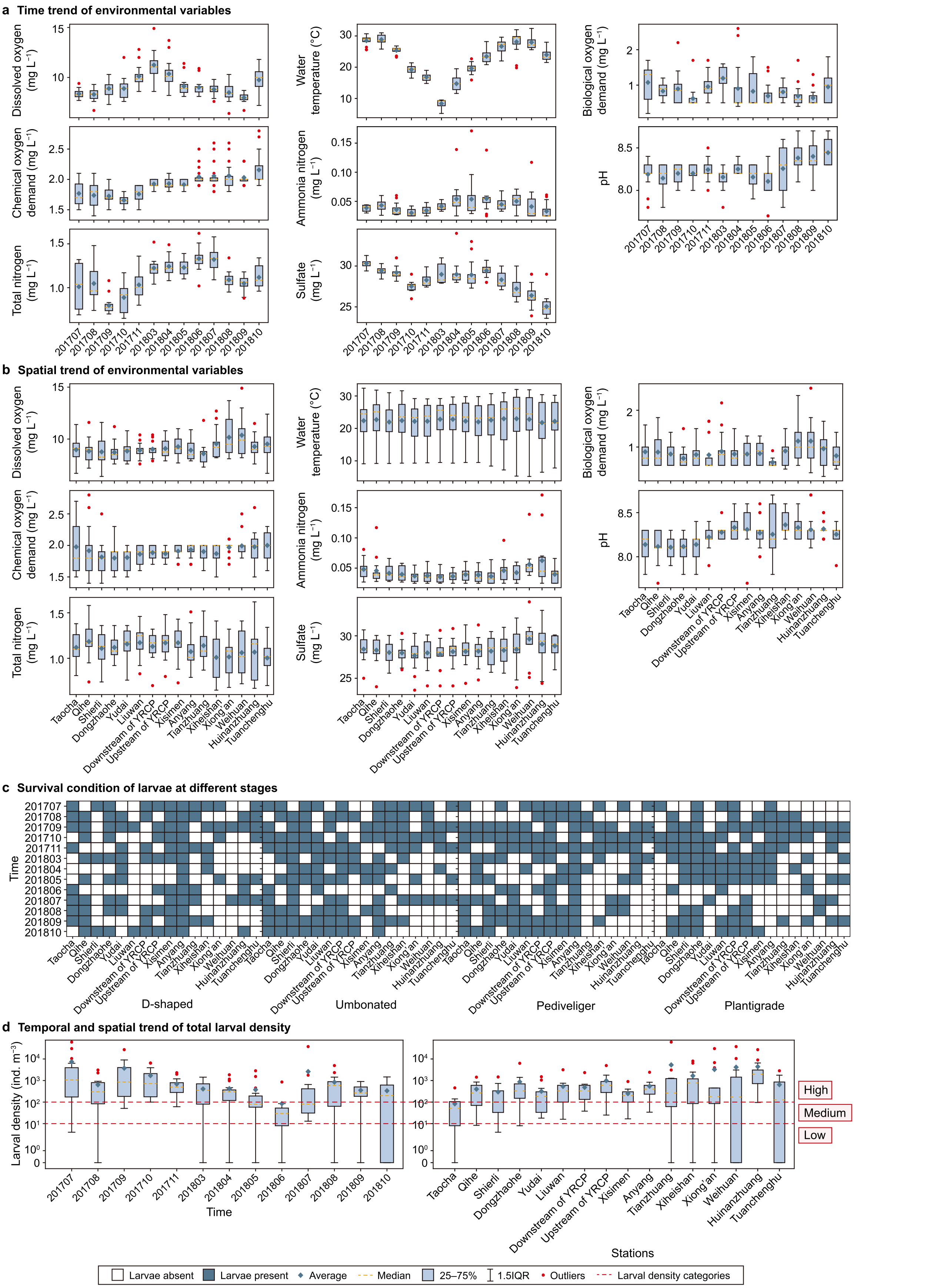

Fig. 4. Environmental factors and larval conditions sampled along the MRSNWDP. **a**, Time trend of environmental variables. **b**, Spatial trend of environmental variables. **c**, Survival condition of larvae at different stages. **d**, Temporal and spatial trend of total larval density.

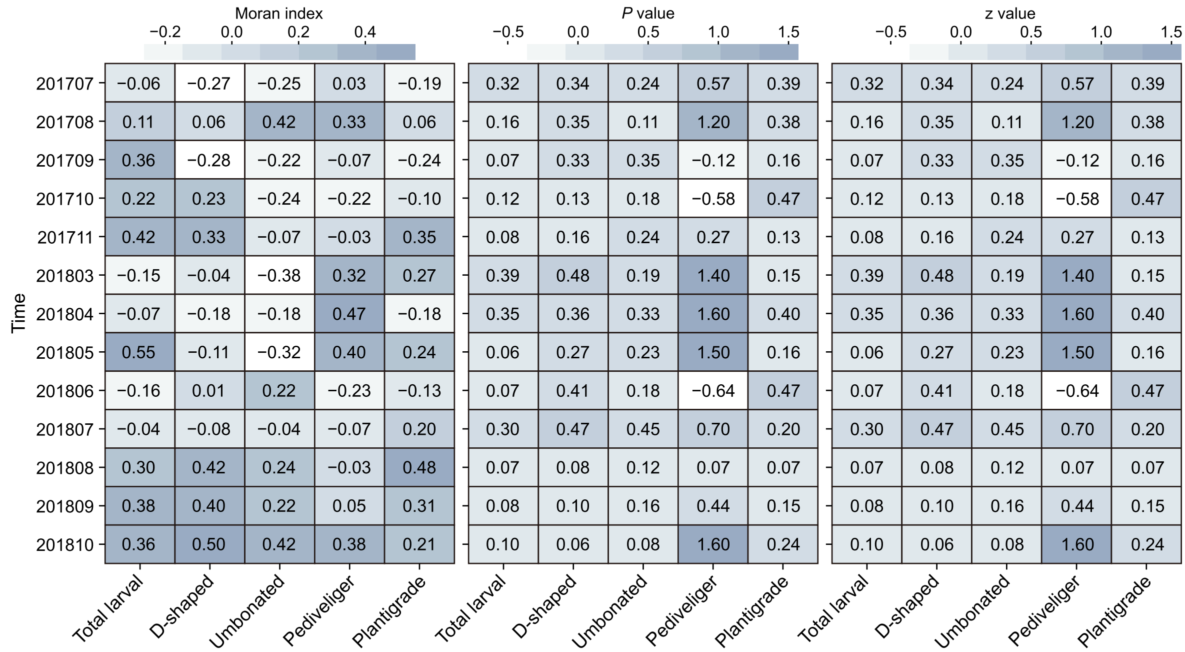

Fig. 5. Spatial and temporal Mann-Kendall test of all collected variables.

Characteristic (ROC) curve (Fig. 9b) illustrate that the Random Forest multiclassification model exhibited better predictions for the medium and low larval density categories. Specifically, the accuracy of predicting the low category in the best-performing model reached 91%. The accuracy for predicting the medium larval density category was 80%, while the accuracy for predicting the high larval density category was 73%. Additionally, the area under the curve (AUC) values for each predicted category were all above 80%.

Upon calculating the feature importance for the best Random Forest model (Fig. 9c), it was determined that total nitrogen emerged as the most influential factor in the model. Water temperature and pH were also found to significantly contribute to the prediction of larval density categories. In contrast, the impact of velocity on the prediction was relatively insignificant when compared to the other factors.

The partial dependence plots of the best Random Forest model (Fig. 9d) depicted the relationships between the four key environmental factors and larval density. Notably, total nitrogen exhibited a positive correlation with low larval density and a negative correlation with high density. Conversely, water temperature displayed a positive correlation with the high larval density category, while pH showed a positive correlation with the medium density category. However, the negative correlation between pH and medium larval density was found to be more pronounced. Furthermore, velocity demonstrated a negative correlation with both low and high larval density, while it exhibited a positive correlation with medium larval density. Higher velocity in the MRSNWDP was observed to maintain larval density at a medium level.

3.4. Assessing establishment risk

Environmental datasets obtained from CAB were utilized for the four representative stations spanning from November 2018 to July 2021 (Fig. S3). Both trained logistic regression and Random Forest model were employed to predict the probability of larval occurrence and the corresponding density category (Fig. 10).

The occurrence probability for the four different larval stages at all four stations generally exceeds 50% based on the predictions. However, the density category at Taocha station, Xiheishan station, and Waihuan station remains predominantly low throughout the predicted period. In contrast, the density category at Huinanzhuang station in Beijing is projected to be consistently high, with only a few months of low larval density during the winter months from November 2018 to July 2021. Furthermore, as anticipated, the occurrence probability for all four larval stages at Huinanzhuang station exceeds 60%. These predictions suggest that the risk of establishment at Huinanzhuang station in Beijing is significantly higher compared to the other three stations.

It has been observed that the levels of total nitrogen, ammonia nitrogen, and velocity at Huinanzhuang station are lower compared to the other stations. Based on the equations of the multivariate logistic regression, lower levels of total nitrogen and ammonia nitrogen contribute to an increased probability of larval survival. Additionally, according to the partial dependence plots of the Random Forest model, lower levels of total nitrogen have the potential to promote a higher larval density category.

4. Discussion

4.1. Key environmental variables influencing larval establishment

Previous studies on *Limnoperna fortunei* [41,45–47] have indicated that this invasive freshwater species exhibits a broad tolerance for chemical and physical variables. The current study builds upon this knowledge by providing a more extensive analysis of the quantitative relationship between the invasion risk of the golden mussel and key environmental factors within inter-basin water diversion projects. By employing a combination of logistic regression to calculate larval survival probability and multiclass classification models to predict larval density categories, a comprehensive risk assessment can be conducted.

In ecological studies, it can be challenging to pinpoint a single variable as the most significant controlling factor due to the temporal and spatial variability of potential stressors. These stressors

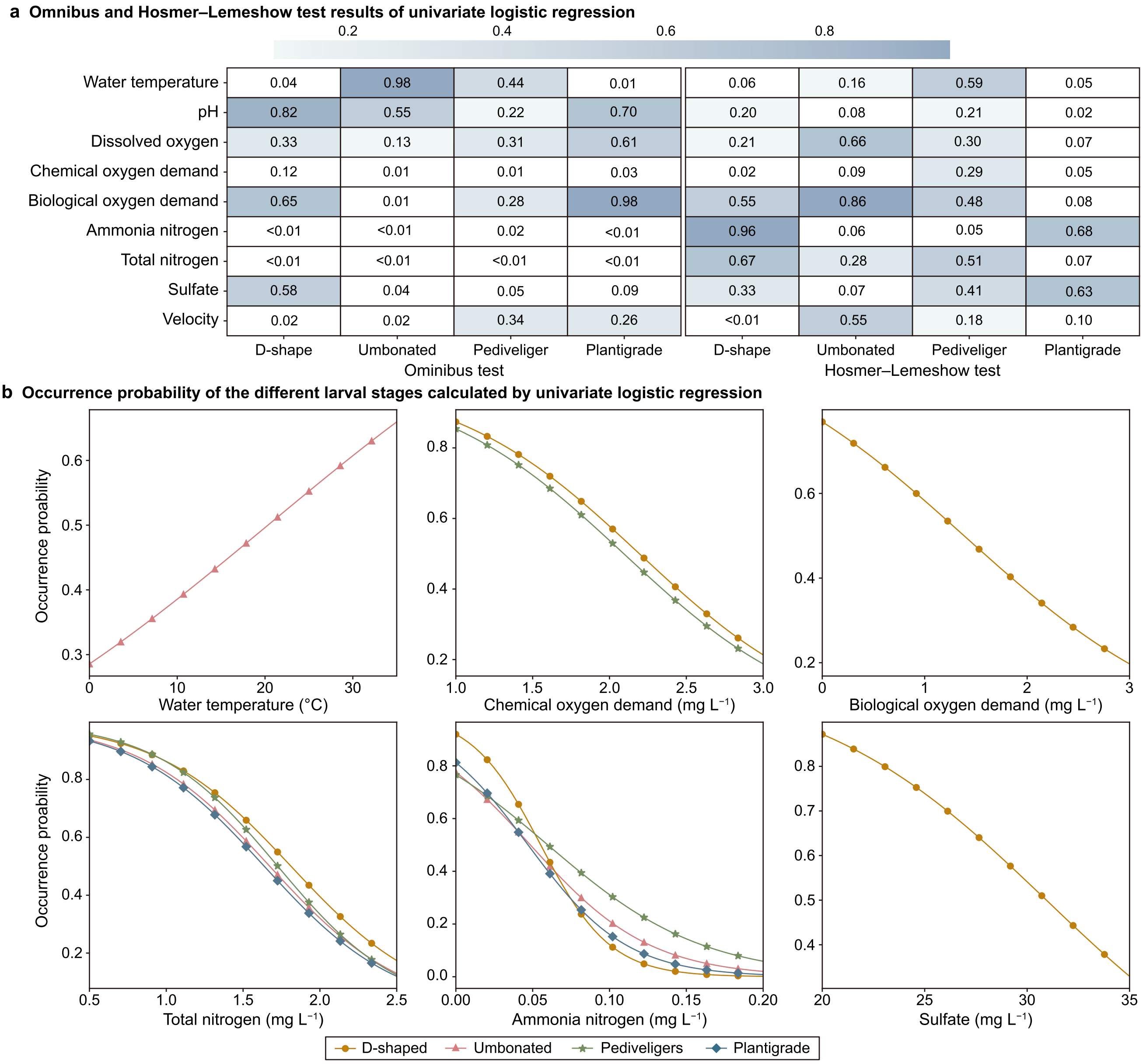

Fig. 6. Significant tests and occurrence probability of univariate logistic regression. **a,** Omnibus and Hosmer–Lemeshow test results of univariate logistic regression. **b**, Occurrence probability of the different larval stages calculated by univariate logistic regression.

Table 3
Collinearity test for the factors selected by univariate logistic regression.

Larval stage	Environmental factors	Tol	VIF	Larval stage	Environmental factors	Tol	VIF
D-shaped	Ammonia nitrogen	0.90	1.11	Umbonated	Ammonia nitrogen	0.66	1.51
	Total nitrogen	0.88	1.14		Total nitrogen	0.87	1.16
	Water temperature	0.95	1.05		Biological oxygen demand	0.87	1.15
Pediveliger	Ammonia nitrogen	0.88	1.13		Chemical oxygen demand	0.81	1.23
	Total nitrogen	0.87	1.16		Sulfate	0.75	1.34
	Chemical oxygen demand	0.89	1.13	Plantigrade	Ammonia nitrogen	0.91	1.09
					Total nitrogen	0.91	1.09

Tol: tolerance; VIF: Variance inflation factor. Tol <0.1 or VIF >10 indicates the existence of collinearity.

often overlap or coincide over time [41,48]. Furthermore, populations being studied differ in their establishment time and preconditions. Nevertheless, by examining records of survival, growth, and reproduction under potential stress conditions, it is possible to identify the range of tolerance for key limiting factors that enable the species to survive and reproduce. This information contributes to the assessment of invasion risk. The findings discussed in this study reveal the relationships between environmental variables

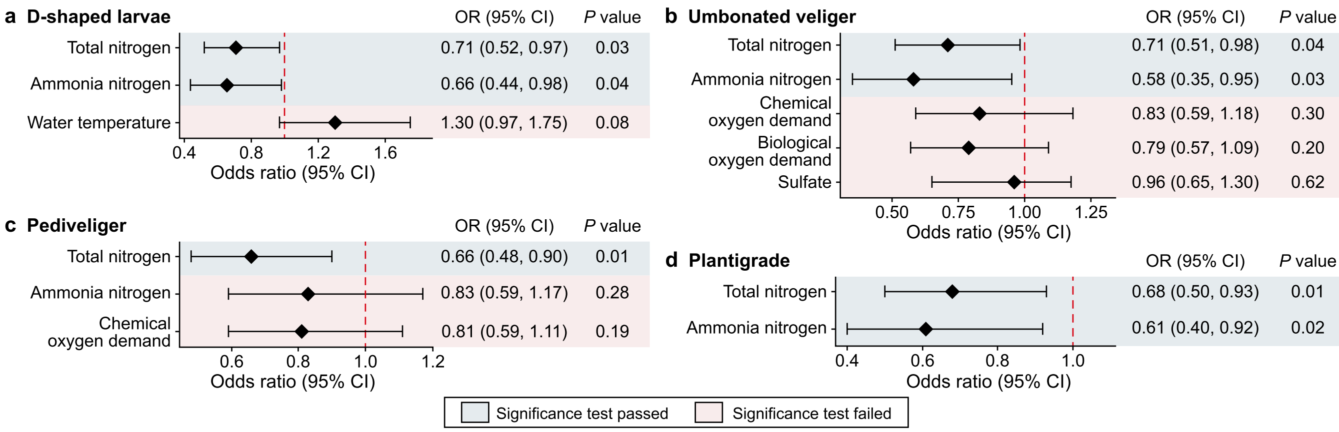

Fig. 7. Multivariable logistic regression between the occurrence probability and key environmental variables. **a,** D-shaped larvae. **b,** Umbonated veliger. **c,** Pediveliger. **d,** Plantigrade.

Table 4
Multivariable logistic regression for occurrence probability of different larval stages.

Larvae	Equation	*P* (omnibus)	*P* (HL)
D-shaped	Logit P = 0.132 − 0.38 × NH_3−N − 0.396 × TN	<0.001	0.349
Umbonated	Logit P = 0.548 − 0.745 × NH_3−N − 0.321 × TN	<0.001	0.223
Pediveliger	Logit P = 0.302 − 0.524 × TN	<0.001	0.514
Plantigrade	Logit P = 0.141 − 0.502× NH_3−N − 0.382 × TN	<0.001	0.963

NH_3−N: ammonia nitrogen; TN: total nitrogen.

and the establishment of the golden mussel, indicating that invasion risk in water transfer projects could be evaluated based solely on these variables. However, it is crucial to select appropriate environmental variables for an accurate risk assessment. In the case of the MRSNDWP, ammonia nitrogen and total nitrogen emerge as two critical variables for assessing the survival of golden mussel larval stages. Additionally, water temperature, velocity, total nitrogen, and pH are identified as the four key variables for predicting larval density.

4.1.1. Total nitrogen

There was a significant negative correlation observed between total nitrogen and both larval occurrence probability and larval density category across all larval stages. These findings indicate that higher trophic conditions may indirectly hinder the reproduction of *Limnoperna fortunei*. This aligns with previous research that has linked elevated nutrient levels to the suppression of golden mussel reproduction due to the proliferation of competing species [49]. Winter data collected by M. A. Pessotto in the La Plata Basin demonstrated a negative correlation between total nitrogen and larval abundance, while a similar negative relationship between larval density and nitrogen was observed in the rivers Paraguay and Miranda [50]. Furthermore, other researchers have attributed the disruption of the species' reproduction in the Salto Grande reservoir (Argentina/Uruguay) during dry summers to the excessive growth of Microcystis, a well-known genus of cyanobacteria that produces toxic substances [6].

4.1.2. Ammonia nitrogen

Previous research has indicated that freshwater mussels exhibit greater sensitivity to ammonia nitrogen compared to other macroinvertebrates and fish [51–54]. However, there is limited information available regarding the toxicity of ammonia in the early life stages of freshwater mussels. In this study, we specifically examined the impact of ammonia nitrogen on the survival of golden mussel larvae. Our findings reveal that exposure to ammonia nitrogen significantly reduces the survival rate of golden mussel larvae. Ammonium ions are known to play a critical role in the larval metamorphosis process under natural conditions [55–57]. Unconjugated ammonia, in particular, is highly toxic to aquatic animals and can result in various behavioral changes, including reduced larval motility, growth, digestion, and reproduction. Even low concentrations of ammonia nitrogen, as low as 0.02 mg L^{-1}, have been reported to induce such effects [58,59].

4.1.3. pH

The findings from this study indicate that there is a relationship between pH and the larval density category, suggesting that an optimal pH range promotes mussel reproduction. While no significant relationship was observed between pH and larval survival probability in our study, other studies, such as the field study conducted in the Upper Paraná River floodplain, have reported positive correlations between pH and larval survival probability [21,48]. Previous research has identified pH as a limiting factor for mussel valve formation [60–62], with pH levels below 5 considered to restrict the survival of *Limnoperna fortunei* adults [63,64]. However, it has been reported that *Limnoperna fortunei* can tolerate a wide range of pH conditions, ranging from 6.2 to 7.4 [7]. Additionally, it is worth noting that golden mussels have been found to survive and reproduce in pH levels ranging from 5 to 10 in South America and Eastern Asia [65]. The relatively stable pH levels observed in the MRSNWDP, which range from 7.7 to 8.7, appear to be suitable for the survival and reproduction of golden mussels.

4.1.4. Velocity

The results of this study demonstrate that an optimal velocity within the MRSNWDP promotes the occurrence probability of the umbonated larval stage. The logistic regression analysis conducted in this study reveals that the survival probability of umbonated larvae exceeds 50% when the flow velocity is below 0.9 m s^{-1}.

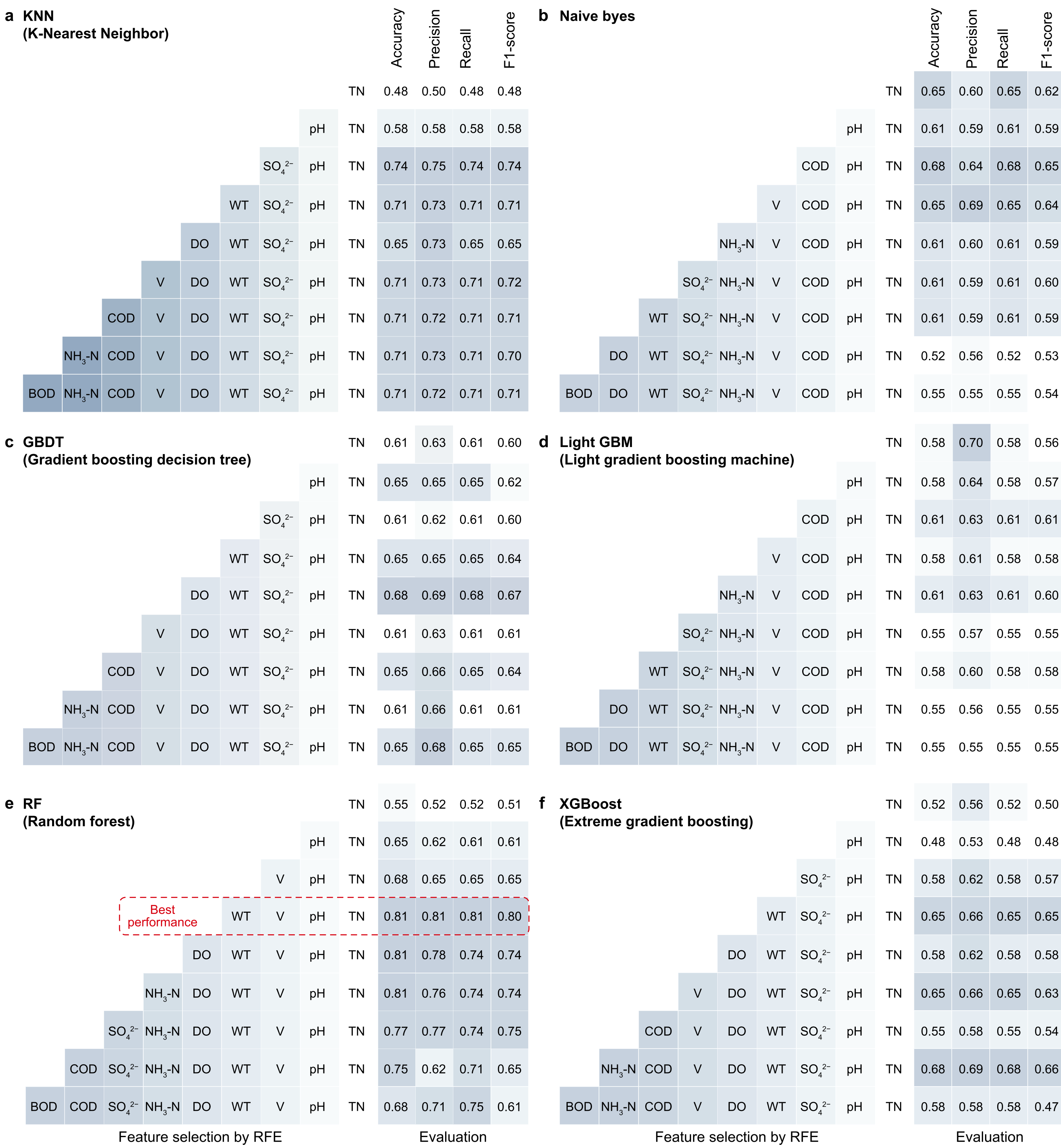

Fig. 8. Evaluating multiclass classification models using variables selected by recursive feature elimination. **a,** KNN. **b**, Naive byes. **c,** GBDT. **d,** LightGBM. **e,** RF. **f,** XGBoost: TN: total nitrogen; V: velocity; BOD: biological oxygen demand; COD: chemical oxygen demand; DO: dissolved oxygen; NH_3—N: ammonia nitrogen; SO_4^{2-}: sulfate; WT: water temperature.

Previous analyses on the attachment of golden mussels to cross-section flow velocity in inter-basin water transfer projects [16] have shown that the mean attachment density of golden mussels was low when the cross-sectional average flow velocity was too low (<0.3 m s^{-1}) or too high (>1.4 m s^{-1}). Conversely, the golden mussel exhibited a higher mean attachment density when the flow velocity ranged from 0.4 to 0.9 m s^{-1}. When the flow velocity is too low, it becomes difficult to ensure a sufficient food supply, which limits the growth and development of the golden mussel. On the other hand, when the flow velocity is too high, a significant number of golden mussels fail to attach to the wall surface before being carried away by the swift water flow. Even if they manage to attach, their byssal thread strength may not withstand the erosive effects of the high-speed water flow. A low attachment rate of adult

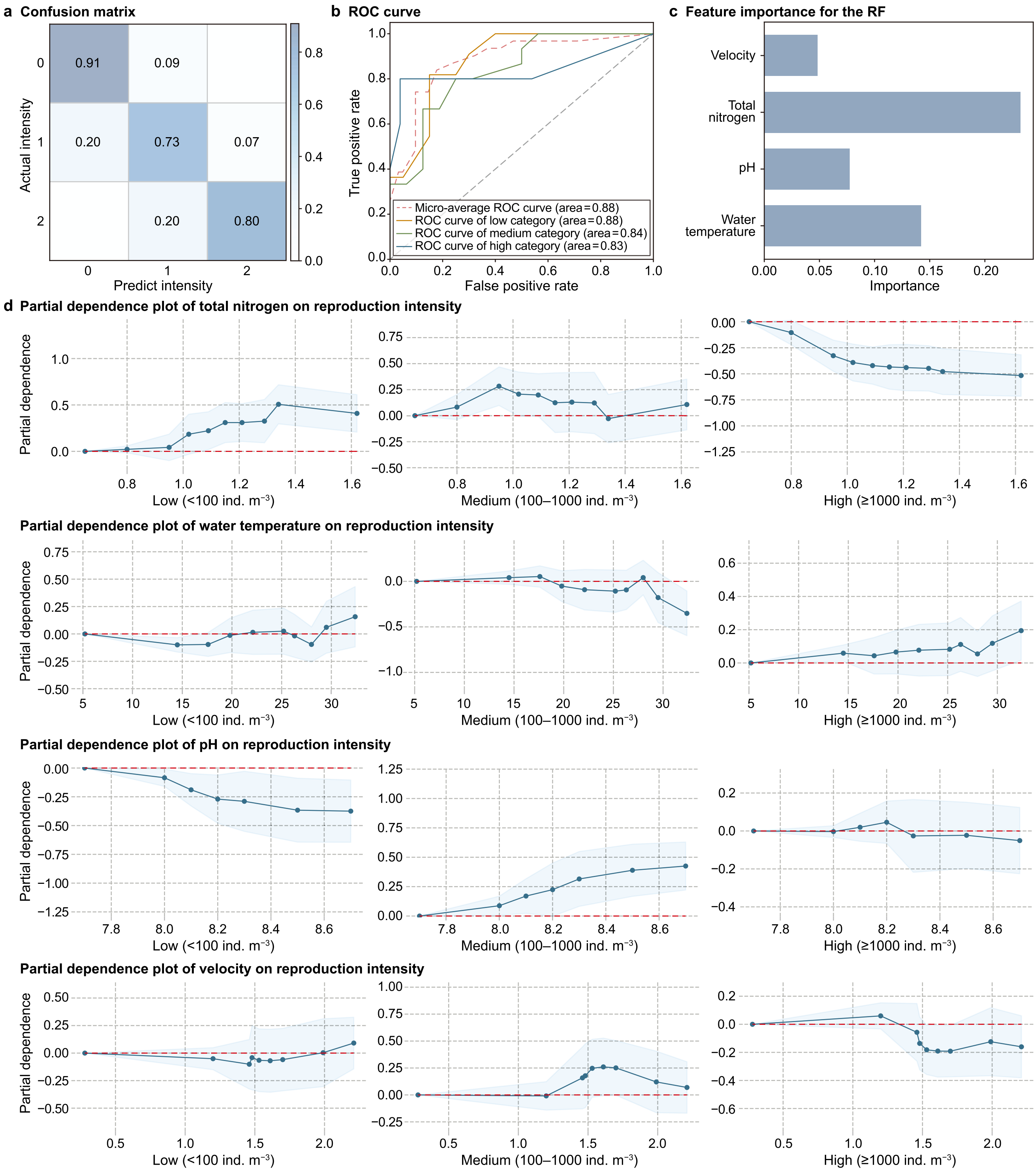

Fig. 9. Evaluation and partial dependence plots of the best-performing Random Forest model. **a**, confusion matrix. **b**, ROC curve. **c**, Feature importance for the RF. **d**, Partial dependence plots.

mussels is unfavorable for the establishment of a complete developmental stage and the continuous reproduction of the golden mussel.

Moreover, it is noteworthy that the upper limit of flow velocity suitable for mussel survival in inter-basin water transfer projects is higher than that observed in natural environments. Other researchers have found that a velocity higher than 1.1 m s^{-1} hindered larval settlement onto artificial substrates in the Miranda River [41].

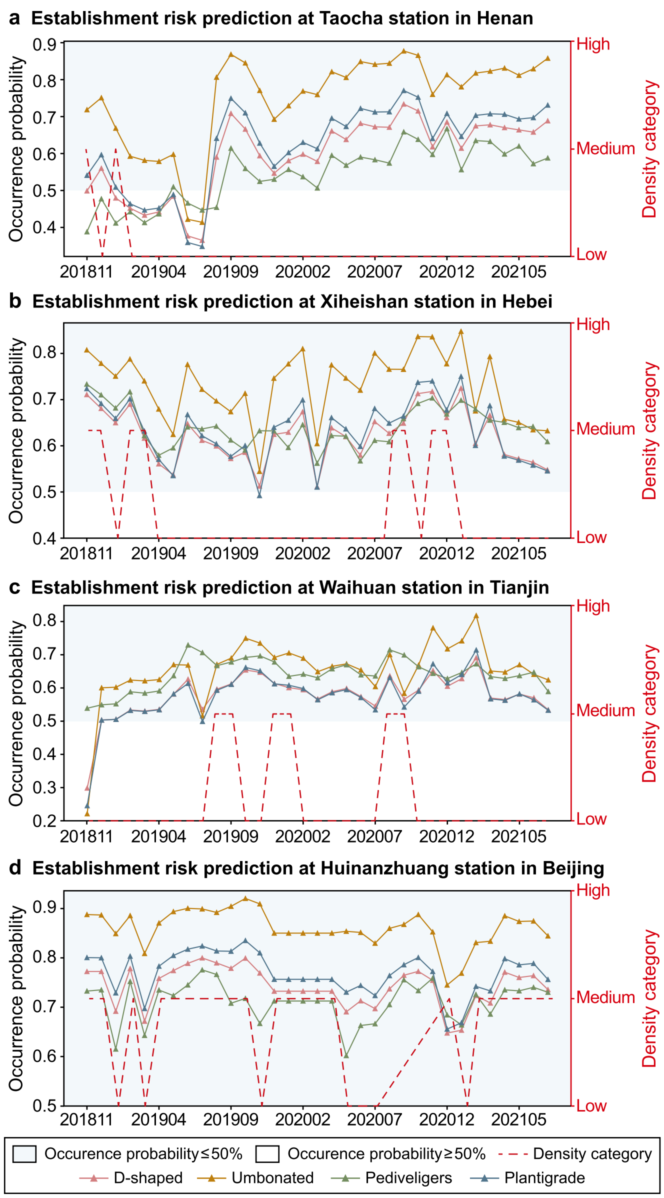

Fig. 10. Risk assessment of *Limnoperna fortunei* at four representative stations along the MRSNDWP. **a,** Establishment risk prediction at Taocha station in Henan. **b**, Establishment risk prediction at Xiheishan station in Hebei. **c**, Establishment risk prediction at Waihuan station in Tianjin. **d,** Establishment risk prediction at Huinanzhuang station in Beijing.

This suggests, to some extent, that the relatively stable water quality and hydrodynamic environment in inter-basin water transfer projects create more favorable conditions for mussel invasion.

4.1.5. Water temperature

The occurrence probability of D-shaped larvae exhibited a positive correlation with water temperature. The feature importance analysis of the best Random Forest multiclass classification model further confirmed the significant relationship between water temperature and larval density category. The partial dependence plot also highlighted the role of temperature in promoting the high larval density category. Temperature is a well-known factor that influences the reproductive rate of organisms [20,66]. In the case of golden mussels, water temperature can affect various physiological processes such as filtration rates, reproductive activity, mortality, and species distribution [6]. Previous studies have reported that increased water temperature can stimulate physiological activities like reproduction and larval metabolism [67]. It has been suggested that a temperature range of 16–17 °C is necessary for larval production to occur [68–70]. Additionally, a minimum temperature of at least 5 °C is required for mussel reproduction and long-term survival, and higher water temperatures contribute to higher reproductive rates [6]. The upper-temperature limit for golden mussels has been reported as 35 °C [41,71].

Investigations on the overwintering survival of caged golden mussels in a reservoir situated at the northern invasion front in northern China revealed that the lowest water temperature for mussel filtration to occur in the laboratory was 5.5 °C [15]. In this study, the water temperature range during the sampling period ranged from 5.2 to 32.4 °C. Notably, during March and April, which corresponded to the coldest sampling period, the water temperature exhibited typical fluctuations within the range of 5.2–16.7 °C. This temperature range has the potential to stimulate the reproductive processes of golden mussels. Therefore, the water temperature conditions within the MRSNWDP provide a favorable environment for the survival and reproduction of golden mussels. It is important to note that the tolerance of golden mussels to low temperatures may differ between natural rivers and water diversion projects. The threshold temperature triggering mussel breeding due to warming water temperatures may vary. However, this hypothesis requires further validation through comparative field studies.

4.2. Establishment risk assessment in water diversion projects

4.2.1. Risk assessment framework

Identifying the environmental factors that exhibit a strong correlation with mussel growth and applying an appropriate statistical analysis to characterize the relationship between these factors and mussel development can facilitate the establishment of a quantitative relationship between mussel invasion and the surrounding environment in water diversion projects.

In this study, we aimed to investigate the relationship between environmental factors and the occurrence probability and larval density category of invasive mussels. The selection of these environmental factors was based on previous research that has established their close association with mussel growth and development. Previous studies have consistently highlighted the significance of water temperature and dissolved oxygen in influencing the growth and reproduction of golden mussels [7,41]. Similarly, the inhibitory effect of high total nitrogen concentrations on larval reproduction has been well-documented in the literature [51–54]. Furthermore, freshwater mussels have been found to be more sensitive to ammonia nitrogen compared to other macroinvertebrates and fish, as recognized by the Environmental Protection Agency [54]. The role of pH as a limiting factor for mussel valve formation has also been widely reported [21,63,74,75]. Additionally, the relationship between velocity and mussel growth attachment has been shown to impact mussel blooms in water diversion projects [41]. Furthermore, sulfate has been observed to reduce the occurrence probability of golden mussel larvae [72]. Chemical oxygen demand has a negative correlation with larval density, possibly due to its influence on water parameters, orthophosphate, and total suspended solids, similar to the effects observed in the zebra mussel [73].

We employed logistic regression and multiclass classification models to analyze the relationship between these environmental factors and mussel occurrence probability and larval density category. The models exhibited good fits, as confirmed by the

Hosmer–Lemeshow test, and achieved high precision and accuracy, exceeding 80% for the best random forest model. These results validate the rationality of selecting these environmental factors for investigation. Furthermore, comparing the critical factors for golden mussel establishment in natural environments and the MRSNWDP offers valuable insights. It has been reported that golden mussels in the MRSNWDP exhibit higher tolerance to cold temperatures than those in other natural environments [11]. Our study revealed that factors such as total nitrogen, ammonia nitrogen, pH, and velocity are particularly influential for the survival and sustained reproduction of mussels in the MRSNWDP, whereas water temperature and dissolved oxygen typically play a vital role in mussel growth and development in other natural environmental conditions [7,41,76]. To gain a more comprehensive understanding of mussel filter-feeding behavior, it is important to explore additional factors that may have a pronounced influence, such as the concentration of sediment particles in the water column and the presence of specific algal species. Furthermore, environmental factors like light, which have been demonstrated to be intricately linked with adult mussel attachment behavior, warrant further examination in future research. Overall, further research and data are needed to validate the differences in key environmental factors and their relationship with invasive mussels in water diversion projects compared to natural conditions. These investigations will contribute to a better understanding of the complex dynamics between invasive mussels and their environment.

Previous studies investigating the relationship between invasive mussel survival and environmental factors have employed various statistical techniques. Some researchers have used Spearman rank correlation to directly assess the correlation between environmental factors and larval density [50,65]. Others have utilized dimensionality reduction techniques like Principal Component Analysis (PCA) or Redundancy Analysis (RDA) to analyze the relationship between environmental characteristics and larval reproduction [23,48,50]. However, the principal components obtained through PCA or CCA may not accurately represent the original variables, leading to the loss of environmental factor information [77]. And the correlation coefficient obtained through Spearman rank can only characterize linear relationships between independent and dependent variables [78]. Logistic regression has been employed by other researchers, using key environmental variables as independent variables to predict the occurrence probability of each larval stage [41,42,74]. In this study, we adopted logistic regression, a classic generalized mixed-effects model with the logit link function, to investigate the relationship between key environmental factors and the probability of larval occurrence at each stage. However, the small sample size and sparsity of the data presented challenges in fully utilizing the advantages of generalized mixed-effects models in data processing [79].

As direct regression analysis did not yield a clear quantitative relationship between larval density and environmental factors, we categorized the total larval density into low, medium, and high density category based on magnitude [80,81]. We then employed multiclass classification models to explore the quantitative relationship between different density levels and the surrounding environmental conditions. To determine the most suitable multiclass classification model, we trained several models, including Naive Bayes, KNN, GBDT, LightGBM, XGBoost, and RF, which have been increasingly used for multiclass classification analysis. Evaluation of these models indicated that the RF model achieved the highest accuracy on the sampled data in this study. RF is an ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy and reduce overfitting. It excels at capturing complex relationships between features and the target variable, particularly in high-dimensional datasets [82]. Moreover, RF randomly selects subsets of features and observations to construct each decision tree, reducing bias and enhancing the model's generalization performance. These characteristics are beneficial for analyzing noisy datasets, as observed in our study. It is important to note that the performance of any algorithm depends on the dataset characteristics, data quality, and the selection of hyperparameters for the model. As our dataset expands, the most suitable model for analyzing the relationship between larval density and environmental factors may change. Therefore, further field investigations are necessary to compare these classification models and identify the most appropriate model for predicting larval density using key environmental factors as independent variables.

4.2.2. Risk assessment at representative stations

The prediction suggests that the mussel density category at the Huinanzhuang Station in Beijing is expected to be high, indicating a greater risk of mussel establishment compared to three other representative stations. The logistic regression analysis revealed that lower levels of total nitrogen and ammonia nitrogen were associated with an increased probability of mussel occurrence. This finding was further supported by the partial dependence plot generated from the best-performing multiclassification random forest model, which showed that lower total nitrogen levels were linked to higher larval density.

It is worth noting that the Huinanzhuang Station holds a unique position among the sampling stations in the MRSNWDP. It is the only large-scale pressurized pumping station and exhibits distinct hydrodynamic and water quality conditions compared to the other predicted stations. This distinction arises from the settling of algae and silt in the forebay and inlet due to the reduced upstream water flow upon entry. Environmental data collected during the forecast period indicated that the Huinanzhuang site had lower levels of total nitrogen, ammonia nitrogen, and flow rate compared to the other stations. Considering the high-risk nature of the Huinanzhuang Pumping Station for golden mussel establishment, it is crucial to implement enhanced monitoring measures in the future. These measures should specifically target the environmental factors identified in this study, such as total nitrogen, ammonia nitrogen, and flow rate, to effectively manage and mitigate the potential spread and establishment of the invasive mussels at the station.

Compared to natural rivers, the occurrence probability of invasive golden mussel larvae in the MRSNWDP is higher due to two primary reasons. Firstly, the sampling method employed for data collection may contribute to this disparity. It is challenging to sample dead larvae settled at the bottom of water diversion channels, resulting in a higher likelihood of sampling living larvae. This biased sampling approach increases the probability of detecting live larvae in the MRSNWDP. Secondly, the water diversion projects create an artificial environment that is more conducive to larval survival and development. Previous research has indicated that freshwater bivalve larvae experience high mortality rates under adverse environmental conditions, such as limited food availability, unsuitable settlement substrates, and extreme temperatures [66,68]. However, the environment in the MRSNWDP is characterized by good water quality, nutrient-rich conditions [83], and suitable concrete structures for mussel attachment [10]. These favorable conditions enhance the survival and reproduction of golden mussel larvae. The breeding cycle of the golden mussel is triggered by warming water temperatures, and a scarcity of food can lead to a halt in breeding [84,85]. As filter feeders, phytoplankton serves as the primary food source for golden mussel larvae during their developmental stage [4,6]. The average water temperature during the sampling period in the MRSNWDP remained relatively stable and sufficiently low to avoid mass larval mortality [26].

Moreover, previous studies have reported the ability of golden mussels to survive the winter in Beijing. Observations revealed that these mussels could filter water at temperatures as low as 5.5 °C and endure temperatures as low as 1 °C for up to 6 days, below 2 °C for 41 days, and below 5 °C for 108 days [15]. The average water temperature during the sampling period in the MRSNWDP was within a range that allowed the presence of larvae in various stages, indicating that the low temperature was not severe enough to cause mass mortality of mussels.

In summary, the MRSNWDP provides a favorable environment for the invasive golden mussel, resulting in a higher occurrence probability of all larval stages compared to natural rivers. However, it is important to note that various complex factors can still influence the survival and reproduction of mussel larvae in different environments. Further research is needed to elucidate the specific reasons for the observed differences in survival probability between water diversion projects and natural rivers.

5. Conclusions

In this study, we utilized the MRSNWDP, the largest water diversion project globally, as a case study to examine the relationship between environmental variables and the risk of golden mussel establishment. We focused on two aspects: larval survival indicated by the occurrence probability and larval reproduction indicated by overall larval density category. Our analysis revealed that certain environmental factors were closely associated with these indicators. Total nitrogen and ammonia nitrogen were found to be the most significant environmental factors influencing the occurrence probability of golden mussel larvae, indicating their impact on larval survival. Additionally, water temperature, velocity, total nitrogen, and pH were identified as key environmental factors significantly correlated with the overall larval density category, which reflects larval reproduction. These findings highlight the importance of monitoring these factors to assess the risk of golden mussel biofouling in the MRSNWDP.

To quantitatively assess the invasion and biofouling risks posed by golden mussels, we proposed an analytical framework that combines logistic regression to calculate the larval occurrence probability and multiclass classification models to predict the overall larval density category. This approach enables the assessment of ongoing risks associated with golden mussel biofouling in water diversion projects. By monitoring easily measurable environmental variables, such as total nitrogen, water temperature, velocity, and pH, it is possible to conduct quantitative risk assessments. This approach offers a rapid and cost-efficient tool for evaluating invasion risks and enhancing the prevention and management of golden mussel biofouling. This study highlights the importance of monitoring key environmental factors in water diversion projects, such as the MRSNWDP, to assess the risk of golden mussel biofouling. By utilizing logistic regression and multiclass classification models, we can quantitatively evaluate invasion risks and improve the effectiveness of prevention and management strategies.

Credit authorship contribution statement

Yang Yao: Conceptualization, Methodology, Software, Writing - Original Draft. **Mengzhen Xu:** Methodology, Writing - Review & Editing. **Xingyu Chen:** Visualization, Writing - Review & Editing. **Jiahao Zhang:** Data Curation, Writing - Review & Editing. **Shulei Wang:** Data Curation. **Jianying Zhu:** Visualization. **Xudong Fu:** Supervision.

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 study was financially supported by the National Key Research and Development Program of China (grant no. 2021YFC3200902 and 2021YFC3200905), and the National Natural Science Foundation of China (grant no. U2243222). Thanks to Liu Wei, Zhao Na, and Han Lujie for their help in mussel sampling. We also thank Steve O'Shea, Ph.D., from Edanz (https://www.edanz.com/ac), for editing a draft of this manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ese.2023.100305.

References

[1] D. Simberloff, J.-L. Martin, P. Genovesi, V. Maris, D.A. Wardle, J. Aronson, F. Courchamp, B. Galil, E. García-Berthou, M. Pascal, Impacts of biological invasions: what's what and the way forward, Trends Ecol. Evol. 28 (2013) 58–66.
[2] J. R Garnas, M.-A. Auger-Rozenberg, A. Roques, C. Bertelsmeier, M.J. Wingfield, D.L. Saccaggi, H.E. Roy, B. Slippers, Complex patterns of global spread in invasive insects: eco-evolutionary and management consequences, Biol. Invasions 18 (2016) 935–952.
[3] P.E. Penchaszadeh, G. Darrigran, C. Angulo, A. Averbuj, M. Brogger, A. Dogliotti, N. Pirez, Predation of the invasive freshwater mussel limnoperna fortunei (Dunker, 1857)(Mytilidae) by the fish Leporinus obtusidens Valenciennes, 1846 (Anostomidae) in the Rio de la Plata, Argentina, J. Shellfish Res. 19 (2000) 229–232.
[4] G. Darrigran, Potential impact of filter-feeding invaders on temperate inland freshwater environments, Biol. Invasions 4 (2002) 145–156, https://doi.org/10.1023/A:1020521811416.
[5] N.P.U. Barbosa, J.A. Ferreira, C.A.R. Nascimento, F.A. Silva, V.A. Carvalho, E.R.S. Xavier, L. Ramon, A.C. Almeida, M.D. Carvalho, A.V. Cardoso, Prediction of future risk of invasion by Limnoperna fortunei (Dunker, 1857) (Mollusca, Bivalvia, Mytilidae) in Brazil with cellular automata, Ecol. Indicat. 92 (2018) 30–39, https://doi.org/10.1016/j.ecolind.2018.01.005.
[6] D. Boltovskoy, Limnoperna Fortunei: the Ecology, Distribution and Control of a Swiftly Spreading Invasive Fouling Mussel, 2015.
[7] G. Darrigran, C. Damborenea, E.C. Drago, I. Ezcurra De Drago, A. Paira, Environmental factors restrict the invasion process of Limnoperna fortunei (Mytilidae) in the Neotropical region: a case study from the Andean tributaries, Ann. Limnol. 47 (2011) 221–229, https://doi.org/10.1051/limn/2011025.
[8] D. Nakano, T. Baba, N. Endo, S. Nagayama, A. Fujinaga, A. Uchida, A. Shiragane, M. Urabe, T. Kobayashi, Invasion, dispersion, population persistence and ecological impacts of a freshwater mussel (Limnoperna fortunei) in the Honshu Island of Japan, Biol. Invasions 17 (2015) 743–759, https://doi.org/10.1007/s10530-014-0765-3.
[9] P. Cordeiro, P. Paixão, W. Lens, M. Lacante, K. Luyckx, The Portuguese validation of the basic psychological need satisfaction and frustration scale: concurrent and longitudinal relations to well-being and ill-being, Psychol. Belg. 56 (2016) 193.
[10] A. Zhan, L. Zhang, Z. Xia, P. Ni, W. Xiong, Y. Chen, G. Douglas Haffner, H.J. MacIsaac, Water diversions facilitate spread of non-native species, Biol. Invasions 17 (2015) 3073–3080, https://doi.org/10.1007/s10530-015-0940-1.
[11] H. Wang, Z. Xia, S. Li, H.J. MacIsaac, A. Zhan, What's coming eventually comes: a follow-up on an invader's spread by the world's largest water diversion in China, Biol. Invasions 25 (2023) 1–5.
[12] N. Zhao, M. Xu, K. Blanckaert, C. Qiao, H. Zhou, X. Niu, Study of factors influencing the invasion of Golden Mussels (Limnoperna fortunei) in water transfer projects, Aquat. Ecosys. Health Manag. 22 (2019) 385–395, https://doi.org/10.1080/14634988.2019.1698860.
[13] G. Darrigran, C. Damborenea, Ecosystem engineering impact of Limnoperna fortunei in South America, Zoolog. Sci. 28 (2011) 1–7, https://doi.org/10.2108/zsj.28.1.
[14] Y. Li, W. Xiong, W. Zhang, C. Wang, P. Wang, Life cycle assessment of water supply alternatives in water-receiving areas of the South-to-North Water Diversion Project in China, 89, https://doi.org/10.1016/j.watres.2015.11.030, 2016.
[15] Z. Xia, J.R. Barker, A. Zhan, G.D. Haffner, H.J. MacIsaac, Golden mussel

(Limnoperna fortunei) survival during winter at the northern invasion front implies a potential high-latitude distribution, Divers. Distrib. 27 (2021) 1422–1434, https://doi.org/10.1111/ddi.13289.

[16] M. Xu, Z. Wang, N. Zhao, B. Pan, Growth, reproduction, and attachment of the golden mussel (Limnoperna fortunei) in water diversion projects, Acta Ecol. Sin. 35 (2015) 70–75, https://doi.org/10.1016/j.chnaes.2015.06.006.

[17] Z. Xia, Quantifying early risks of species invasions: factors regulating south to north bivalve colonization of novel habitats. https://scholar.uwindsor.ca/etd/7745, 2019.

[18] J. Zhang, M. Xu, L. Sun, D. Reible, X. Fu, Impact of golden mussel (Limnoperna fortunei) colonization on bacterial communities and potential risk to water quality, Ecol. Indicat. 144 (2022) 109499.

[19] T.M. Blackburn, P. Pyšek, S. Bacher, J.T. Carlton, R.P. Duncan, V. Jarošík, J.R.U. Wilson, D.M. Richardson, A proposed unified framework for biological invasions, Trends Ecol. Evol. 26 (2011) 333–339.

[20] D. Cataldo, D. Boltovskoy, J.L. Hermosa, C. Canzi, Temperature-dependent larval development rates of Limnoperna fortunei (Mollusca, Bivalvia), J. Molluscan Stud. 71 (2005) 41–46.

[21] J. Ernandes-Silva, G.D. Pinha, R.P. Mormul, Environmental variables driving the larval distribution of Limnoperna fortunei in the upper Paraná River floodplain, Brazil, Acta Limnol. Bras. 29 (2017), https://doi.org/10.1590/s2179-975x7716.

[22] M.C.D. Mansur, C.P. dos Santos, G. Darrigran, I. Heydrich, C.T. Callil, F.R. Cardoso, Primeiros dados quali-quantitativos do mexilhão-dourado, Limnoperna fortunei (Dunker), no Delta do Jacuí, no Lago Guaíba e na Laguna dos Patos, Rio Grande do Sul, Brasil e alguns aspectos de sua invasão no novo ambiente, Rev. Bras. Zool. 20 (2003) 75–84, https://doi.org/10.1590/s0101-81752003000100009.

[23] W. Liu, M. Xu, J. Zhang, T. Zhang, Survival and attachment of biofouling freshwater mussel (Limnoperna fortunei) to environmental conditions: potential implications in its invasion, infection and biofouling control, Limnology 21 (2020) 245–255, https://doi.org/10.1007/s10201-020-00607-1.

[24] D. Boltovskoy, E. Paolucci, H.J. MacIsaac, A. Zhan, Z. Xia, N. Correa, What we know and don't know about the invasive golden mussel Limnoperna fortunei, Hydrobiologia (2022) 1–48.

[25] A.T. Peterson, D.A. Vieglais, Predicting Species Invasions Using Ecological Niche Modeling: new Approaches from Bioinformatics Attack a Pressing Problem: a new approach to ecological niche modeling, based on new tools drawn from biodiversity informatics, is applied to the challenge of, Bioscience 51 (2001) 363–371.

[26] W. Liu, M. Xu, Z. Wang, D. Yu, X. Zhou, Experimental study of attachment characteristics of golden mussel larvae, Sheng Tai Xue Bao 37 (2017) 2779–2787, https://doi.org/10.5846/stxb201601030010.

[27] M. Xu, Z. Wang, C.C. Lin, B. Pan, N. Zhao, Experimental study of invasion and biofouling of freshwater mussel <i>Limnoperna fortunei</i&gt, Int. J. Geosci. 4 (2013) 1–7, https://doi.org/10.4236/ijg.2013.45b001.

[28] P.A.P. Moran, The interpretation of statistical maps, J. R. Stat. Soc. Ser. B. 10 (1948) 243–251.

[29] D.N. Lawley, A generalization of Fisher's z test, Biometrika 30 (1938) 180–187.

[30] C.R. Boyd, M.A. Tolson, W.S. Copes, Evaluating trauma care: the TRISS method. Trauma score and the injury severity score, J. Trauma 27 (1987) 370–378. http://europepmc.org/abstract/MED/3106646.

[31] N. Senaviratna, T. Cooray, Diagnosing multicollinearity of logistic regression model, Asian J. Probab. Stat. 5 (2019) 1–9.

[32] R.B. d'Agostino, An omnibus test of normality for moderate and large size samples, Biometrika 58 (1971) 341–348.

[33] D.W. Hosmer, T. Hosmer, S. Le Cessie, S. Lemeshow, A comparison of goodness-of-fit tests for the logistic regression model, Stat. Med. 16 (1997) 965–980.

[34] E. Fix, J.L. Hodges, Discriminatory analysis. Nonparametric Discrimination: consistency properties, Int. Stat. Rev./Rev. Int. Stat. 57 (1989) 238, https://doi.org/10.2307/1403797.

[35] T.K. Ho, Random decision forests, Proc. Int. Conf. Doc. Anal. Recognition, ICDAR. 1 (1995) 278–282, https://doi.org/10.1109/ICDAR.1995.598994.

[36] I. Kononenko, Semi-naive Bayesian classifier, in: Eur. Work. Sess. Learn., Springer, 1991, pp. 206–219.

[37] L. Mason, J. Baxter, P. Bartlett, M. Frean, Boosting algorithms as gradient descent in Function space, Nips (1999), https://doi.org/10.1109/5.58323.

[38] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, Xgboost: extreme gradient boosting, R Packag. Version (2015) 1–4, 0.4-2. 1.

[39] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, Lightgbm: a highly efficient gradient boosting decision tree, Adv. Neural Inf. Process. Syst. 30 (2017).

[40] M.C. Jackson, L. Huang, Q. Xie, R.C. Tiwari, A modified version of Moran's I, Int. J. Health Geogr. 9 (2010) 1–10.

[41] M.D. Oliveira, D.F. Calheiros, C.M. Jacobi, S.K. Hamilton, Abiotic factors controlling the establishment and abundance of the invasive golden mussel Limnoperna fortunei, Biol. Invasions 13 (2011) 717–729, https://doi.org/10.1007/s10530-010-9862-0.

[42] A.S. Melo, P. Alegre, P. Alegre, Modelo preditivo de sobrevivência do Mexilhão Dourado (Limnoperna fortunei) em relação a variações de salinidade na Laguna dos Patos , RS , Brasil Introdução Material e Métodos, 9 (n.d.).

[43] D.C. Depew, E. Krutzelmann, K.E. Watchorn, A. Caskenette, E.C. Enders, The distribution, density, and biomass of the zebra mussel (Dreissena polymorpha) on natural substrates in Lake Winnipeg 2017–2019, J. Great Lake. Res. 47 (2021) 556–566, https://doi.org/10.1016/j.jglr.2020.12.005.

[44] D. Pestana, M.R. Pie, A. Ostrensky, W.A. Boeger, C. Andreoli, F. Franceschi, P. Lagos, Seasonal variation in larval density of limnoperna fortunei (Bivalvia, Mytilidae) in the iguaçu and Paraná rivers, in the region of foz do iguaçu, Paraná, southern Brazil, Braz. Arch. Biol. Technol. 51 (2008) 607–612, https://doi.org/10.1590/S1516-89132008000300023.

[45] D. Boltovskoy, N. Correa, D. Cataldo, F. Sylvester, Dispersion and ecological impact of the invasive freshwater bivalve Limnoperna fortunei in the Río de la Plata watershed and beyond, Biol. Invasions 8 (2006) 947–963.

[46] B. Morton, The biology and anatomy of Limnoperna fortunei, a significant freshwater bioinvader: blueprints for success, in: Limnoperna Fortunei Ecol. Distrib. Control a Swiftly Spreading Invasive Fouling Mussel, 2015, pp. 3–41.

[47] A. Ricciardi, R.J. Neves, J.B. Rasmussen, Impending extinctions of North American freshwater mussels (Unionoida) following the zebra mussel (Dreissena polymorpha) invasion, J. Anim. Ecol. 67 (1998) 613–619.

[48] J. Ernandes-Silva, F.H. Ragonha, S. Jati, A.M. Takeda, Limnoperna fortunei Dunker, 1857 Larvae in different environments of a Neotropical floodplain: relationships of abiotic variables and phytoplankton with different stages of development, Braz. J. Biol. 76 (2016) 154–161, https://doi.org/10.1590/1519-6984.15514.

[49] M.A. Pessotto, M.G. Nogueira, More than two decades after the introduction of limnoperna fortunei (Dunker 1857) in La Plata Basin, Braz. J. Biol. 78 (2018) 773–784, https://doi.org/10.1590/1519-6984.180789.

[50] V. Eilers, M.D. de Oliveira, K.F. Roche, Density and body size of the larval stages of the invasive golden mussel (Limnoperna fortunei) in two neotropical rivers, Acta Limnol. Bras. 23 (2012) 282–292, https://doi.org/10.1590/s2179-975x2012005000006.

[51] J. Miller-Kleinhenz, X. Guo, W. Qian, H. Zhou, E.N. Bozeman, L. Zhu, X. Ji, Y.A. Wang, T. Styblo, R. O'Regan, Dual-targeting Wnt and uPA receptors using peptide conjugated ultra-small nanoparticle drug carriers inhibited cancer stem-cell phenotype in chemo-resistant breast cancer, Biomaterials 152 (2018) 47–62.

[52] T. Augspurger, A.E. Keller, M.C. Black, W.G. Cope, F.J. Dwyer, Water quality guidance for protection of freshwater mussels (Unionidae) from ammonia exposure, Environ. Toxicol. Chem. An Int. J. 22 (2003) 2569–2575.

[53] K. Deiner, J.C. Walser, E. Mächler, F. Altermatt, Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA, Biol. Conserv. 183 (2015) 53–63, https://doi.org/10.1016/j.biocon.2014.11.018.

[54] EPA, EPA's 2008 Report on the environment, Environ. Protect. (2008). EPA/600/R-07/045F (NTIS PB2008-112484), http://www.epa.gov/roe. EPA/600/R-07/045F (NTIS PB2008-112484).

[55] S. Berking, K. Herrmann, Dicapryloylglycerol and ammonium ions induce metamorphosis of ascidian larvae, Roux's Arch. Dev. Biol. 198 (1990) 430–432.

[56] S. Berking, Ammonia, tetraethylammonium, barium and amiloride induce metamorphosis in the marine hydroid Hydractinia, Roux's Arch. Dev. Biol. 197 (1988) 1–9.

[57] F. Wang, C. Li, W. Zhao, M.D.H. Rahman, Y. Wang, F. Yang, Z. Huo, L. Fang, X. Yan, Toxicity of ammonia to larvae and juveniles of the geoduck clam Panopea japonica, Fish. Sci. 88 (2022) 311–317.

[58] J.A. Camargo, A. Alonso, A. Salamanca, Nitrate toxicity to aquatic animals: a review with new data for freshwater invertebrates, Chemosphere 58 (2005) 1255–1267.

[59] J.A. Camargo, Á. Alonso, Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: a global assessment, Environ. Int. 32 (2006) 831–849.

[60] C.W. Ramcharan, D.K. Padilla, S.I. Dodson, A multivariate model for predicting population fluctuations of Dreissena polymorpha in North American lakes, Can. J. Fish. Aquat. Sci. 49 (1992) 150–158.

[61] G.L. Mackie, L.A. Flippance, Relationships between buffering capacity of water and the size and calcium content of freshwater mollusks, Freshw. Invertebr. Biol. 2 (1983) 48–55.

[62] A.G. Checa, F.J. Esteban-Delgado, A.B. Rodríguez-Navarro, Crystallographic structure of the foliated calcite of bivalves, J. Struct. Biol. 157 (2007) 393–402.

[63] L. Montalto, M. Marchese, Limnoperna fortunei (Dunker, 1857)(Bivalvia: Mytilidae) tolerance to temperature and pH in experimental conditions, Neotropica. 49 (2003) 26–34.

[64] A.Y. Karatayev, D. Boltovskoy, D.K. Padilla, L.E. Burlakova, The invasive bivalves Dreissena polymorpha and Limnoperna fortunei: parallels, contrasts, potential spread and invasion impacts, J. Shellfish Res. 26 (2007) 205–213.

[65] G.L. Mackie, J.K. Brinsmead, A risk assessment of the golden mussel, Limnoperna fortunei (Dunker, 1857) for Ontario, Canada, Manag. Biol. Invasions. 8 (2017) 383–402, https://doi.org/10.3391/mbi.2017.8.3.12.

[66] D. Boltovskoy, F. Sylvester, A. Otaegui, V. Leites, D.H. Cataldo, Environmental modulation of reproductive activity of the invasive mussel Limnoperna fortunei: implications for antifouling strategies, Austral Ecol. 34 (2009) 719–730.

[67] G.C. Favero, T.P. Boaventura, A.L. Ferreira, A.C.F. Silva, L.A. Porto, R.K. Luz, Fasting/re-feeding and water temperature promote the mobilization of body reserves in juvenile freshwater carnivorous catfish Lophiosilurus alexandri, Aquaculture 511 (2019) 734223, https://doi.org/10.1016/j.aquaculture.2019.734223.

[68] D. Boltovskoy, D.H. Cataldo, Population dynamics of Limnoperna fortunei, an invasive fouling mollusc, in the lower Paraná River (Argentina), Biofouling 14 (1999) 255–263.

[69] A.G. Checa, C. Jiménez-López, A. Rodríguez-Navarro, J.P. Machado, Precipitation of aragonite by calcitic bivalves in Mg-enriched marine waters, Mar. Biol. 150 (2007) 819–827.
[70] D.H. Cataldo, D. Boltovskoy, Yearly reproductive activity of Limnoperna fortunei (Bivalvia) as inferred from the occurrence of its larvae in the plankton of the lower Paraná river and the Río de la Plata estuary (Argentina), Aquat. Ecol. 34 (2000) 307–317.
[71] A.Y. Karatayev, L.E. Burlakova, D.K. Padilla, Zebra versus quagga mussels: a review of their spread, population dynamics, and ecosystem impacts, Hydrobiologia 746 (2015) 97–112, https://doi.org/10.1007/s10750-014-1901-x.
[72] J. Salerno, P.L. Gillis, H. Khan, E. Burton, L.E. Deeth, C.J. Bennett, P.K. Sibley, R.S. Prosser, Sensitivity of larval and juvenile freshwater mussels (unionidae) to ammonia, chloride, copper, potassium, and selected binary chemical mixtures, Environ. Pollut. 256 (2020) 113398.
[73] D. Glaser, J.R. Rhea, D.R. Opdyke, K.T. Russell, C.K. Ziegler, W. Ku, L. Zheng, J. Mastriano, Model of zebra mussel growth and water quality impacts in the Seneca River, New York, Lake Reserv, Manag 25 (2009) 49–72.
[74] J. Ernandes-Silva, F.H. Ragonha, L.C. Rodrigues, R.P. Mormul, Freshwater invasibility level depends on the population age structure of the invading mussel species, Biol. Invasions 18 (2016) 1421–1430, https://doi.org/10.1007/s10530-016-1091-8.
[75] A.Y. Karatayev, D.K. Padilla, D. Minchin, D. Boltovskoy, L.E. Burlakova, Changes in global economies and trade: the potential spread of exotic freshwater bivalves, Biol. Invasions 9 (2007) 161–180, https://doi.org/10.1007/s10530-006-9013-9.
[76] M.I. O'Connor, J.F. Bruno, S.D. Gaines, B.S. Halpern, S.E. Lester, B.P. Kinlan, J.M. Weiss, Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation, Proc. Natl. Acad. Sci. USA 104 (2007) 1266–1271.
[77] S. Karamizadeh, S.M. Abdullah, A.A. Manaf, M. Zamani, A. Hooman, An overview of principal component analysis, J. Signal Inf. Process. 4 (2013) 173.
[78] T.D. Gauthier, Detecting trends using Spearman's rank correlation coefficient, Environ. Forensics 2 (2001) 359–362.
[79] S. Saeed, E.E.M. Moodie, E.C. Strumpf, M.B. Klein, Segmented generalized mixed effect models to evaluate health outcomes, Int. J. Public Health. 63 (2018) 547–551.
[80] P.D. Johnson, K.M. Brown, The importance of microhabitat factors and habitat stability to the threatened Louisiana pearl shell, Margaritifera hembeli (Conrad), Can. J. Zool. 78 (2000) 271–277.
[81] W.R. Haag, J. Warren Melvin L, Role of ecological factors and reproductive strategies in structuring freshwater mussel communities, Can, J. Fish. Aquat. Sci. 55 (1998) 297–306.
[82] J. Gall, N. Razavi, L. Van Gool, An introduction to random forests for multi-class object detection, in: Outdoor Large-Scale Real-World Scene Anal. 15th Int. Work. Theor. Found. Comput. Vision, Dagstuhl Castle, Ger. June 26-July 1, 2011. Revis. Sel. Pap., Springer, 2012, pp. 243–263.
[83] X. Nong, D. Shao, H. Zhong, J. Liang, Evaluation of water quality in the South-to-North Water Diversion Project of China using the water quality index (WQI) method, Water Res. 178 (2020) 115781.
[84] J. Widdows, The effects of temperature on the metabolism and activity of Mytilus edulis, Netherlands J. Sea Res. 7 (1973) 387–398.
[85] E.A. Newell, Direct and delayed costs of reproduction in Aesculus californica, J. Ecol. (1991) 365–378.