Using ROC-curves to illustrate the use of GLM-models in environmental activity analysis

K V Zakharov¹, Yu V Somova², D P Sychugov², T V Sviridova², E A Nekerov² and O V Aryukhin³

¹Moscow Region State University, 24 Very Voloshinoy St., Moscow Region, Mytishi, Russian Federation
²Nosov Magnitogorsk State Technical University, 38 Lenin Avenue, Magnitogorsk, Russian Federation
³All-Russian Scientific Research Institute of Fisheries and Oceanography

E-mail: coz.prutkoff@yandex.ru

Abstract. Generalized linear models (GLM-models) are used to assess the effectiveness of environmental protection in Moscow. As the object of research, we selected projects of economic activity in natural and green areas in Moscow. The effectiveness of environmental protection work was considered as a dependent variable, and the nature of economic activity, environmental protection status, the work of social activists and environmental protection structures, as well as biotopic characteristics of the territory were considered as independent predictors. To assess the quality of the obtained models, ROC-curves were constructed. In accordance with the results obtained, the best model includes three predictors - active environmental protection work by the public and officials, and forest vegetation.

1. Introduction
As urbanization intensifies, the problem of protecting wildlife in cities is becoming more and more relevant and attracts the attention of specialists and the general public. The importance of wildlife in the city is shown in the works of many authors [1, 2], moreover, despite geographical, historical and social differences, urban formations, as a rule, have similar ecological problems. At the same time, assessments of ecological balance, geological diversity, environmental protection and practical recommendations are often subjective in nature [3–5].

As an object of research, we have chosen environmental protection work in Moscow, the largest city in Russia, which has a long history of environmental protection movement [6–9].

We note that, in contrast to biological diversity [10], the effectiveness of environmental protection activity is often difficult to quantify [11, 12]. The commonly used method is multiple regression, the application of which requires a normal distribution of the dependent variable and has other statistical constraints. For this reason, we opted for generalized linear models.

2. Materials and methods
To assess the effectiveness of environmental work, we used projects of economic activities affecting natural and green areas in Moscow (n = 89) and associated with obvious negative ecological impact. If such projects could be prevented, the environmental work was considered successful. Our work includes the period of 2000–2016.
A number of environmental laws have been adopted in Moscow, the most notable natural areas have the status of specially protected natural areas (SPNA), rare and endangered species of animals, plants and hornbeams are protected. All nature protection statuses are included in the work, in addition, for each of the cases of economic activity we considered, active environmental protection activities of social activists or state environmental structures were taken into account.

The analysis includes the nature of economic activity (construction, landscaping) and the type of vegetation (forest, open biotopes, etc.). The prevention of environmentally harmful economic activities was considered as a result (dependent variable), and environmental status, environmental activity, the nature of economic activity and biotopic features were considered as predictors (independent variables).

All variables had a binary distribution, for this reason, generalized linear models (GLM-models) were used for the analysis (function “glm”, family = binomial, link = “logit”). For the calculations, the Excel, Statistica and R software packages were used. The mathematical processing algorithm includes the following stages:

– calculation of correlation and removal of correlated predictors;
– modeling using GLM-models;
– use of ROC-curves to evaluate the simulation results.

Since all variables are binary, the tetrachoric coefficient is used to calculate the correlation. After removing statistically insignificant predictors, the models were compared using the ANOVA test (chi-square version) (function “anova”, test = “Chisq”).

Akaike Information Criterion (AIC) was also used when comparing the models: models with different variables were selected using the “stepAIC” function (MASS package).

With the stepwise selection a combined method was used. Excessive variance was detected using the “pchisq” function (p <0.05).

To assess the forecast quality of the obtained models, matrices of inaccuracies with a threshold of $t=0.5$ and ROC-curves (Receiver Operator Characteristics) (functions “ROCRpred” and “ROCRperf” from the ROCR package) were constructed. To estimate the probability, the “predict” function with artificial data sets was used.

3. Results and discussion

After removing correlated predictors and selecting all models, several significant factors that affect the success of environmental protection work in Moscow (predictor “result”) were identified.

First of all, it is the active work of state structures (predictor “officials”) and the society (predictor “society”) (table 1).

Table 1. Statistically significant coefficients for the predictor “result”.

| Coefficients | Estimate | Std. Error | z value | Pr(>|z|) |
|--------------|----------|------------|---------|---------|
| (Intercept)  | -4.305   | 1.047      | -4.111  | 3.94e-05 *** |
| society      | 3.153    | 1.107      | 2.848   | 0.00440 ** |
| officials    | 4.588    | 1.225      | 3.745   | 0.00018 *** |

The ROC-curve expresses the model with the “officials” and “society” predictors in figure 1. The larger the area occupied by the ROC-curve, the higher the forecast quality.
Figure 1. ROC-curve, expressing the probability of preventing environmentally harmful economic activities in the case of activities of the public and government agencies.

ROC-curve, expressing the probability of preventing environmentally harmful economic activities in the case of the society and government agencies work.

The model was improved only by including the “forest” predictor, as a result of which the ROC-curve acquired a slightly different shape, and the forecast quality improved (figure 2).
Figure 2. ROC-curve expressing the probability of preventing environmentally harmful economic activities for a model that includes active environmental work and the variable “forest”.

Comparison of the model with the three indicated predictors (model 1) and the model including all variables (model 2) did not reveal a statistically significant difference, since the chi-square test value is insignificant (table 2) and we can use a simpler model.

Table 2. Results of applying the “chi-square” test for comparing generalized linear models.

| Resid. | Df Resid. | Dev   | Df  | Deviance | Pr(>Chi) |
|--------|-----------|-------|-----|----------|----------|
| 1      | 81        | 31.507| -4  | -0.57098 | 0.9662   |
| 2      | 85        | 32.078|      |          |          |

Using statistically significant predictors alone also degrades the quality of the model. As an example, let's take a ROC-curve built for a model with one “officials” predictor (figure 3).

The area occupied by the ROC curve is smaller and, therefore, the quality of the model and forecast is worse.
Figure 3. ROC-curve, expressing the quality of the forecast for the model, which includes only one predictor (environmental protection work of government agencies).

4. Conclusion
Thus, the best model includes three variables: active environmental protection work of social activists and government agencies and the presence of forest vegetation.

References
[1] Chiesura A 2004 The role of urban parks for the sustainable city Landsc. and Urban Plann. 68 129–38
[2] Hartig T, Mitchell R, de Vries S and Frumkin H 2014 Nature and health Annu. Rev. of Public Health 35 207–28
[3] Zakharov K 2020 The assessment of the efficiency of environmental activities in Moscow Eco. Env. & Cons. 26 (3) 89–94
[4] Ruban D, Zorina S and Rebezov M 2019 Dispersed geoheritage points of the lagonaki highland, SW Russia: contribution to local geoheritage resource Geosciences (Switzerland) 9 367 DOI:10.3390/geosciences9090367
[5] Mikhailenko A, Ruban D, Yashalova N and Rebezov M 2019 The unique granite gorge in mountainous adygeya, Russia: evidence of big and complex geosite disproportions. Geosciences (Switzerland) 9 372 DOI:10.3390/geosciences9090372
[6] Zakharov K 2019 Nature protection in Moscow and its efficiency Ecology of urban areas 1 55–62 DOI: 10.24411/1816-1863-2019-11055
[7] Oldfield J D and Shaw D J B 2002 Revisiting sustainable development: Russian cultural and scientific traditions and the concept of sustainable development Area 34 391–400
[8] Oldfield J D 2002 Russian environmentalism Eur Env 12 117–29
[9] Newell J P and Henry L A 2017 The state of environmental protection in the Russian Federation: a review of the post-Soviet era Eurasian Geography and Economics DOI:
[10] Emelyanova L G 2019 Assessment of the distribution of rare bird species in regional arid biomes of Russia as the basis for their protection *Arid Ecosystems* 9 (1) 59–68

[11] Wendland K J, Baumann M, Lewis D J, Sieber A and Radeloff V C 2015 Protected area effectiveness in European Russia: A post-matching panel data analysis *Land Economics* 91 (1) 149–68

[12] Henry L A and Douhovnikoff V 2008 Environmental issues in Russia *Annu. Rev. Environ. Resour.* 33 437–60