UNIVERSITY OF PÉCS

Biological and Sportbiological Doctoral School

Longitudinal zonation of riverine net-spinning caddisfly larvae (Trichoptera: Hydropsychidae) in Northern Hungary and the Northern Great Plain – reasons, explanation, and the aspects of ecological modelling

PhD Thesis

Márk Ficsór

Supervisor:
Dr. Zoltán Csabai
head of department, associate professor
Department of Hydrobiology

PÉCS, 2024

1. INTRODUCTION

The ecological niche is a multidimensional spcace defined by environmental factors as its axes, within which populations of species are able to exist (Chase & Leibold, 2003; Hutchinson, 1978; MacArthur & Levins, 1967). Identifying and exploring these hypothetic axes are key issues in ecology, that lead us at the same time towards a better understanding of driving forces behind species distribution. Exact values, local differences and long-term changes of these environmental parameters have a direct effect on the presence or absence of certain species, and can also be key facilitators of speciation (i.e. the evolutionary process of new species formation) (Hughes et al., 2009; Mayr, 1973).

Special scenes of such processes are riverine ecosystems where the uni-directional flow creates a continuous gradient of many biotic and abiotic variables, for example that of discharge, mean velocity, mean annual temperature, substrate particle size or riparian inputs, all of which, in addition, change from headwaters towards estuaries in a rather predictable way. These environmental gradients produce longitudinal zonation patterns in several groups of lotic organisms, e. g. fish or benthic invertebrates (Allan, 1995; Gordon et al., 1992; Hynes, 1970; Illies, 1961; Statzner & Dolédec, 2011; Vannote et al., 1980). Such zonation can be observed in the case of net-spinning caddisfly larvae (Trichoptera: Hydropsychidae).

Larvae of Hydropsychidae, especially the members of the eponymic genus *Hydropsyche* are known to exhibit an overlapping downstream sequence of characteristic species, or more commonly a combination of two or three of them along river networks (Roux et al., 1992).

The worldwide similarity of their sequential zonation, as well as the wide range of sensitivities and tolerance of organic pollution, anthropogenic effects or changes in water quality make net-spinning caddisfly larvae ideal bio-indicators and an ideal group of model organisms for studying species distribution in relation of environmental factors (Bonada et al., 2004; Higler & Tolkamp, 1983; Pîrvu et al., 2015). Despite their renowned suitability in applied ecology and in different ecological quality assessment methods based on benthic invertebrates, there is still more to reveal about or model the types and extent of impacts certain environmental factors have in shaping the sequential distribution of *Hydropsyche* species.

Gradient-like longitudinal environmental parameters, such as annual water temperature range (Statzner & Dolédec, 2011), flow velocity (Hildrew & Edington, 1979;

Tachet et al., 1992) or substrate particle size (Alstad, 1980, 1982; Fuller & Mackay, 1980; Fuller et al., 1983; Wallace, 1975; Wallace et al., 1977) have a special importance in shaping the longitudinal zonation of Central-European Hydropsyche, while some indicators of ecological quality, such as the annual mean of dissolved oxygen (Becker, 1987; Philipson, 1954; Philipson & Moorhouse, 1974), the inorganic forms of phosphorous and nitrogen (Buczyńska, 2013; Camargo et al., 2005; Pirvu et al., 2015; Vuori, 1995), salinity (Piscart et al., 2005; Sala et al., 2016; Zinchenko & Golovatyuk, 2013), heavy metal content (Bonada et al., 2005; Tszydel et al., 2016; van der Geest et al., 1999) as well as organic pollution in general (Engels et al., 1996; Higler & Tolkamp, 1983; Stuijfzand et al., 1999; Vuori, 1995) could still be defined determining. Biotic factors, like species-specific traits, habitat and resource partitioning, staggering life cycles or competition influence primarily the coexistence of species in the same stretch, some of them, such as net-building characteristics, or differences in food paticle size preferences, however, could have and indirect effect on longitudinal distribution patterns (Alstad, 1982; Malas & Wallace, 1977; Wallace, 1975; Wallace et al., 1977).

Species distribution modelling (SDM) has a central role in ecological and conservation studies to explain species occurrence based on environmental variables (Franklin, 2010). SDMs could reveal and evaluate the relationship between a certain environmental predictor and the presence (or absence) of species, while also being suitable for the estimation of effects that changes of these predictors cause. Models based on carefully selected, adequately complex set of environmental variables can predict the occurrence of species in formerly undiscovered or uninhabited locaions.

After focusing mainly on linear, descriptive statistics and ordination techniques as key tools of SDM, attention has recently turned towards the application of machine learning (ML-) algorhytms due to their ability to handle missing or imbalanced data (a frequent feature of ecological studies), flexibility of fitting functions, automatic variable importance estimation and powerful predictive performance (Elith et al., 2006; Guisan et al., 2002, Knudby et al., 2010; Olden et al., 2008, Šmilauer & Lepš, 2014, Valavi et al., 2021). Furthermore, single ML methods can be combined to work as an ensemble in order to improve performance and precision by learning from the predictions of base learners (Rokach, 2010, Wolpert, 1992).

2. AIMS

Due to the worldwide similarity of the sequential zonation of species and also to the wide range of sensitivity to organic pollution and other adverse anthropogenic factors, larvae of the genus Hydropsyche are excellent bio-indicators, and ideal model organisms for studying species distribution in relation of environmental factors. Although members of the family Hydropsychidae are often used in applied ecological research, as well as in ecological quality assessment methods based on aquatic invertebrates, there is still more to reveal about the types and extent of impacts certain environmental factors have in shaping their sequential distribution.

The aims of our research were:

(1) to identify the main physical, chemical, habitat related and landuse parameters that influence the distribution of 10 Hydropsyche species occurring in the area investigated,

(2) cluster species into groups based on similar ecological preferences,

(3) to build an ensemble of machine learning (ML-) models based on the selected parameters that could explain and predict the distribution patterns of these groups.

By the evaluation of the results obtained from the above mentioned steps, we could get an insight into the direct and indirect effects of the most important environmental factors, that shape the distribution of species. Using these factors as variables in an accurately built distribution model, occurrence of species or species groups can be predicted for habitats with unknown assemblages. Expected changes of distribution patterns can also be predicted in case of different environmental scenarios.

3. MATERIALS AND METHODS

Field samplings were carried out at 94 different sampling sites, all located on the right hand catchment of river Tisza in North-Hungary and the Northern Great Plain between 2009 and 2018. Net-spinning caddisfly larvae were collected as part of the benthic invertebrate fauna investigations determined by the EU's Water Framework Directive (WFD) according to the pro-rata multi-habitat AQEM protocol (AQEM Consortium, 2002).

Physical and chemical parameters of water were determined by the Department of Public Health, Laboratory for Environmental Protection, Government Office of Borsod-Abaúj-Zemplén County according to certified standards, while habitat-related parameters were recorded at the time of sampling of the biota. Some hydrological parameters were provided by the North-Hungarian Water Conservancy Directorate (ÉMVIZIG).

Data were analized using three main types of statistical methods: dimesionality reduction techniques, decision tree based machine leaning (ML-) algorithms and regression analysis.

Reduction of data dimensionality – like in the case of e.g. redundancy analysis (RDA), or hierarchical clustering – means the transformation of a high dimensional data set – containing, in our case, a high number of environmental predictors – into a low dimensional space, using the linear combination of the original features. It also includes the identification of features, according to which certain elements of the data set can be separated from each other and/or grouped into clusters.

Decision tree based machine learning techniques include predictive algorithms, that first create base-learner decision trees with weak predictive power, then combine, improve or summarize their predictions in order to achieve higher accuracy, specificity, sensitivity, etc. The algorithms used in our study – *Random Forest* (RF – Breiman, 2001), *Bagged Adaptive Boosting* (AdaBag – Freund & Schapire, 1997), *Gradient Boosting* (GBM – Friedman, 2001) and *Extreme Gradient Boosting* (XGB – Chen & Guestrin, 2016) – apply two basic types of ensemble: bootstrap aggregating („bagging"), and boosting.

As one of the final steps of analysis we applied *Penalized Multinomial Logistic Regression* (PMLR – Anderson & Blair, 1982), a type of regression analysis, in which the probabilty of a possible outcome is predicted as a linear combination of one or more independent variables (predictors).

After data preprocessing the top most important predictors per species equal to 1/10 the count of cases indicating presence were determined according to the results of a 5-fold cross-validated Random Forest model (Liaw & Wiener, 2002).

For the clustering of species to ecologically relevant groups the dissimilarity matrix of species scores on the first four axes of a redundancy analysis (RDA – Oksanen et al., 2020) was subjected to hierarchical clustering (Kassambara & Mundt, 2020). Species belonging to the same cluster were then handled together as groups in subsequent analyses.

„Stacking ensemble" technique (Wolpert, 1992) was used to model the relationship between species(-groups) and the retained environmental variables, where the 5-fold cross-validated predictions of four different base learner (Level 1) models – RF, GBM, AdaBag and XGB – trained on the stratified 3/4 split of the modelling dataset were “stacked” as predictors in a final „meta”-model.

Level 2 models, with hyperparameters, determined during Level 1 were used to give predictions on the hold-out 1/4 split of the dataset. This resulted in a test set of data in which the dependent variable is the dominant species(-group), while the independent variables are predictions of Level 2 models.

As a final step, predictions of a PMLR-model – trained on the stacked predictions of base leraners, thus „combining their predictive power” – were tested on the predictions of Level 2 ones. Performance of each model were assessed calculating accuracy, Cohen’s kappa (κ) and the area under the Receiver Operating Characteristic (ROC-) curves.

4. RESULTS AND DISCUSSION

Faunistical results

A total 9376 larval specimens of ten different *Hydropsyche* species – *H. angustipennis, H. bulbifera, H. bulgaromanorum, H. contubernalis, H. fulvipes, H. incognita, H. instabilis, H. modesta, H. pellucidula* and *H. saxonica* – were recorded from 278 samples during the ten-year sampling period. Less sensitive species with higher tolerance of adverse anthropogenic effects were found most frequently and had the highest relative abundancies as well. Aggregation of environmental parameters and species’ abundance data yielded a final raw dataset of 215 cases for further analyses.

The ten species of the net-spinning caddisfly genus *Hydropsyche* found in the investigated area represent a wide range of environmental preferences. H. *fulvipes, H. instabilis* and *H. saxonica* inhabit small and medium sized streams of the crenal and upper rhithral zones, *H. incognita* and *H. pellucidula* are characteristic all along the rhithral zones, while *H. contubernalis* and *H. bulgaromanorum* are typical large-river species, living in the potamal zone. *H. angustipennis*, *H. bulbifera* and *H. modesta* are known to be among the most tolerant hydropsychids in Central Europe, usually occurring in the middle sections of the stream–river continuum (Waringer and Graf, 2011). Faunistical results reflecting the sequential zonation patterns that are characteristic for the genus.

The influence of environmental factors on the distribution of net-spinning caddisfly larvae

13 environmental variables were identified to be relevant to the presence of one or more *Hydropsyche* species. Members of the group *fulv_inst_saxo* (*H. fulvipes, H. instabilis* and *H. saxonica*) prefer sites at higher altitudes, that have larger particle sized substrates under waters with lower average temperature, lower salinity and lower levels of nitrate, while characteristic species of the potamal region (H. modesta, H. contubernalis and H. bulgaromanorum) unsurprisingly occur in larger rivers with higher discharge, farther from sources. Species inhabiting small and medium sized streams of rhithral regions have wider range of preferences, the higher tolerace of the least sensitive ones, however – especially that of *H. angustipennis* in case of chemical oxygen demand (COD) and orthophosphate-phosphorous content (PO4P) – is markedly obvious.

It is also remarkable that the 13 most important variables are measured in various spatial scales. They include spot-like measurements (such as conductivity, chemical oxygen demand, orthophosphate-phosphorous content, etc.), parameters, recorded for a 50-100 m reach of watercourse (e.g. substrate particle size composition) as well as variables defined at the range of geographical scales (e.g. above mean sea level, land use variables).

Clustering of species to ecologically relevant groups

First four axes of the redundancy analysis (RDA) explained 38.23%, 19.44%, 18.94% and 11.91 % of variance in the species abundance data constrained by the selected environmental variables.

The clustering of species resulted in the formation of 6 different “groups”, three of which contained only one species (H. bulgaromanorum, H. contubernalis and H. modesta), while others had two (H. incognita + H. pellucidula and H. angustipennis + H. bulbifera) or three (H. fulvipes + H. instabilis + H. saxonica) members. The groups created by clustering distinctly reflect the sequential zonation, ecological preferences and also the distinction in the tolerance of species according to Dohet (2002), Lechthaler & Stockinger (2005), Pitsch (1993) and Waringer & Graf (2011). The only member of the *bulg* group, H. bulgaromanorum is a characteristic species of the potamal zones of large lowland rivers. Species of the closely related clusters *cont* and *mode* (containing H. contubernalis and H. modesta repectively) inhabit similar habitats, typically at higher altitudes in the investigated area. Members of the angu_bulb group, H. angustipennis and H. bulbifera rarely occur in the potamal zones, but mostly prefer small and medium sized rivers in the rhithral zone (Lechthaler & Stockinger, 2005; Waringer & Graf, 2011). Similar zonal preferences characterize the group inco-pell, the sensitivity of its members, H. incognita and H. pellucidula, however, are much higher than that of the former group. H. fulvipes, H. instabilis and H. saxonica form a well-defined, distinct group – these species inhabit headwaters and small streams at the highest altitudes while being the less tolerant of environmental degradation, water pollution and adverse anthropogenic effects (Edington & Hildrew, 1995; Graf et al., 2008, Higler & Tolkamp, 1983; Statzner & Dolédec, 2011).

Results of modelling the relationship between environmental variables and the dominant occurance of species(groups)

Level 1 (base learner) models showed mediocore or poor performance in predicting the dominant occurrence of the clusters above. The list of the most important environmental variables in these models, however, were highly similar, and their multi-scalar nature remained the same, despite the decrease in the detail of the dependent variable, i.e. the dominant taxon, as a result of hierarchical clustering. Bagged Adaptive Boosting performed the worst (AUC = 0.715, Accuracy = 0.475, Kappa = 0.269) followed by Extreme Gradient Boosting (AUC = 0.760, Accuracy = 0.468, Kappa = 0.296) and Generalized Boosting (AUC = 0.777, Accuracy = 0.506, Kappa = 0.332) while the best performance was measured for the Random Forest base learner (AUC = 0.802, Accuracy = 0.526, Kappa = 0.347). The weak performance of base learners – at least compared to their predictive power reported by e.g. Valavi et al., 2022 and Zhang et al., 2019 – is most probably the result of classs imbalance (Benkendorf et al., 2013), the relatively small number of cases (van Proosdij et. al., 2016) and the strict validation criteria applied.

All three evaluation metrics showed a significant increase in the case of the final (meta-) model (AUC = 0.847, Accuracy = 0.769, Kappa = 0.686), which appeared to have maximum sensitivity (TPR – true positive rate) for groups bulg (TPR = 1.000) and cont (TPR = 1.000), while being still impressively high for the group angu_bulb (TPR = 0.908). Balanced accuracy, used for an „overall", group-wise evaluation of the model performance and expressed as the mean of sensitivity and specificity, was the highest in case of groups bulg, cont and fulv_inst_saxo (1.000, 0.978 and 0.888 respectively), while being the lowest in case of groups angu_bulb, inco_pell and mode (0.856, 0.689 and 0.621 respectively).

5. SUMMARY

According to our results, gradient-like environmental parameters are the most important predictors of the longitudinal zonation of Hydropsyche species, while some indicators of ecological quality – whose values are typically related to certain acute or diffuse anthropogenic impacts – can also have significant effect.

Investigating the complex effects of the most important environmental variables on the distribution of Hydropsyche species resulted in clusters of single species or species-groups, reflecting the similar ecological demands of their members.

According to the performance of our final, machine learning (ML-) ensemble model that can accurately predict the presence or absence of spcies and/or species-groups based on selected environmental variables we can conclude, that

– the distribution patterns of Hydropsyche larvae in the river network of a (sub)basin can be adequately modelled by machine learning algorithms based on carefully selected environmental parameters;

– due to the heterogeneity of microhabitats and the peculiar longitudinal connectivity of a dendritic stream-river network, it is particularly important to include variables measured on various spatial scales in case of modelling riverine ecosystems;

– the mediocre or poor performance of ML-algorithms with strictly supervised 5-fold cross-validation – applied to compensate the natural bias in species occurrence data – can be highly improved by stacking ensemble to give accurate predictions;

– a model, constructed this way, could accurately predict the dominant occurrence of Hydropsyche species/groups at sites with previously unknown fauna;

– it could also be used to examine or predict the changes of distribution patterns or areas of species under different environmental or climate change scenarios;

– the modelling approach we applied here can be suitable to study the distribution patterns of further species or taxa in other aquatic invertebrate groups constrained by their environmental preferences.

6. REFERENCES

ALLAN, J. D. (1995): Stream ecology. – Chapman & Hall, London, 388 pp.

ALSTAD, D. N. (1980): Comparative biology of the common Utah Hydropsychidae (Trichoptera). – American Midland Naturalist 103: 167-174.

ALSTAD, D. N. (1982): Current speed and filtration rate link caddisfly phylogeny and distributional patterns on a stream gradient. – Science 216: 553-554.

ANDERSON, J. A. – BLAIR, V. (1982): Penalized maximum likelihood estimation in logistic regression and discrimination. – Biometrika 69(1): 123-136.

AQEM CONSORTIUM (2002): Manual for the application of the AQEM method. A comprehensive method to assess European streams using benthic macroinvertebrates, developed for the purpose of the Water Framework Directive. Version 1.0, February 2002.

BECKER, G. (1987): Net-building behaviour, tolerance and development of two caddisfly species from the river Rhine (Hydropsyche contubernalis and H. pellucidula) in relation to the oxygen content. – Oecologica (Berlin) 73: 242-250.

BENKENDORF, D. J. – SCHWARTZ, S. D. – CUTLER, D. R. – HAWKINS, C. P. (2023): Correcting for the effects of class imbalance improves the performance of machine-learning based species distribution models. – Ecological Modelling 483: 110414.

BONADA, N. – ZAMORA-MUÑOZ, C. – RIERADEVALL, M. – PRAT, N. (2004): Ecological profiles of caddisfly larvae in Mediterranean streams: implications for bioassessment methods. – Environmental pollution 132: 509-521.

BONADA, N. – VIVES, S. – RIERADEVALL, M. – PRAT, N. (2005): Relationship between pollution and fluctuating asymmetry in the pollution-tolerant caddisfly Hydropsyche exocellata (Trichoptera, Insecta). – Archiv für Hydrobiologie 162: 167–185.

BREIMAN, L. (2001): Random Forests. – Machine Learning 45: 5-32.

BUCZYŃSKA, E. (2013): The influence of a dam reservoir on caddisflies (Trichoptera) of an upland river on the example of the River Bystrzyca (south-eastern Poland). – Ochrona Środoviska I Zasobów Naturalnych 24: 17-22.

CAMARGO, J. A. – ALONSO, A. – SALAMANCA, A. (2005): Nitrate toxicity to aquatic animals: a review with new data for freshwater invertebrates. – Chemosphere 58: 1255-1267.

CHASE, J. M. – LEIBOLD, M. A. (2003): Ecological niches: Linking classical and contemporary approaches. – University of Chicago Press, Chicago, 221 pp.

CHEN, T. – GUESTRIN, C. (2016): XGBoost: A Scalable Tree Boosting System. – KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2016: 785–794.

DOHET, A. (2002) - Are caddisflies an ideal group for the biological assessment of water quality in streams? – In: W. MEY (ed.): Proceedings of the 10th International Symposium on Trichoptera, Potsdam, Germany – Nova Supplementa Entomologica, Keltern 15: 507-520.

EDINGTON, J. M. – HILDREW, A. G. (1995): A revised key to the caseless caddis larvae of the British Isles, with notes on their ecology. – Freshwater Biological Association Scientific Publications 53: 1-119.

ELITH, J. – GRAHAM, C. H. – ANDERSON, R. P. – DUDÍK, M. – FERRIER, S. – GUISAN, A. – HIJMANS R. J. – HUETTMAN, F. – LEATHWICK, J. R. – LEHMANN, A. – LI, J. – LOHMANN, L. G. – LOISELLA, B. A. – MANION, G. – MORITZ, C. – NAKAMURA, M. – NAKAZAWA, Y. – OVERTON, J. MCC. M. – TOWNSEND PETERSON, A. – PHILLIPS, S. J. – RICHARDSON, K. – SCHACHETTI-PEREIRA, R. – SCHAPIRE, R. E. – SOBERÓN, J. – WILLIAMS, S. – WISZ, M. S. – ZIMMERMANN, N. E. (2006): Novel methods improve prediction of species' distributions from occurrence data. – Ecography 29(2): 129-151.

ENGELS, S. – NEUMANN, D. – LÖBBEL, H. – BRÜHNE, M. (1996): Waiting for Hydropsyche – why has only one of at least four local Hydropsyche species returned into the Lower Rhine? – Large Rivers 10: 313–314.

FRANKLIN, J. (2010): Mapping species distributions: spatial inference and prediction. – Cambridge University Press, 340 pp.

FREUND, Y. – SCHAPIRE, R. E. (1997): A decision-theoretic generalization of on-line learning and an application to boosting. – Journal of Computer and System Sciences 55(1): 119-139.

FRIEDMAN, J. H. (2001): Greedy Function Approximation: A Gradient Boosting Machine. – The Annals of Statistics 29(5): 1189-1232.

FULLER, R. L. – MACKAY, R. J. (1980): Feeding ecology of three species of Hydropsyche (Trichoptera: Hydropsychidae) in southern Ontario. – Canadian Journal of Zoology 58: 2239-2251.

FULLER, R. L. – MACKAY, R. J. – HYNES, H. B. N. (1983): Seston capture by Hydropsyche betteni nets (Trichoptera, Hydropsychidae). – Archiv für Hydrobiologie 97: 251-261.

GORDON, N. D. – MCMAHON, T. A. – FINLAYSON, B. L. (1992): Stream hydrology. – Wiley, Chichester, 526 pp.

GRAF, W. – MURPHY, J. – DAHL, J. – ZAMORA-MUÑOZ, C. – LÓPEZ-RODRÍGUEZ, M. J. (2008): Trichoptera. – In: SCHMIDT-KLOIBER, A. – HERING, D. (eds.): Distribution and Ecological Preferences of European Freshwater Organisms vol. 1., Pensoft Publishers, Sofia-Moscow, 388 pp.

GUISAN, A. – EDWARDS, T. C. – HASTIE, T. (2002): Generalized linear and generalized additive models in studies of species distributions: setting the scene. – Ecological Modelling 157: 89-100.

HIGLER, L. W. G. – TOLKAMP, H. H. (1983): Hydropsychidae as bio-indicators. – Environmental Monitoring and Assessment 3: 331-341.

HILDREW, A. G. – EDINGTON, J. M. (1979): Factors facilitating the coexistence of Hydropsychid caddis larvae (Trichoptera) in the same river system. – Journal of Animal Ecology 48: 557-576.

HUGHES, J. M. – SCHMIDT, D. J. – FINN, D. S. (2009): Genes in streams: using DNA to understand the movement of freshwater fauna and their riverine habitat. – BioScience 59: 573-583.

HUTCHINSON, G. E. (1978): An introduction to population ecology. – Yale University Press, New Haven and London, 271 pp.

HYNES, H. B. N. (1970): The ecology of running waters. – University of Toronto Press, Toronto, Ontario, Canada, 555 pp.

ILLIES, J. (1961): Versuch einer allgemeinen biozönotischen Gliederung der Fließgewässer. – Internationale Revue der gesmten Hydrobiologie und Hydrographie 46: 205-213.

KASSAMBARA, A. – MUNDT, F. (2020): factoextra: Extract and Visualize the Results of Multivariate Data Analyses. – R package, version 1.0.7.

KNUDBY, A. – BRENNING, A. – LEDREW, E. (2010): New approaches to modelling fish-habitat relationships. – Ecological Modelling 221(3): 503-511.

LECHTHALER, W. – STOCKINGER, W. (2005): Trichoptera – Key to Larvae from Central Europe (DVD)

LIAW, A. – WIENER, M. (2002): Classification and Regression by randomForest. – R News, 2(3), 18-22.

MACARTHUR, R. – LEVINS, R. (1967): The limiting similarity, convergence, and divergence of coexisting species. – The American Naturalist 101(921): 377-385.

MALAS, D. – WALLACE, J. B. (1977): Strategies for coexistence in three species of net-spinning caddisflies (Trichoptera) in second-order southern Appalachian streams. – Canadian Journal of Zoology 55: 1829-1840.

MAYR, E. (1947): Ecological factors in speciation. – Evolution 1(4): 263-288.

OKSANEN, J. – BLANCHET, F. G. – FRIENDLY, M. – KINDT, R. – LEGENDRE, P. – MCGLINN, D. – MINCHIN, P. R. – O'HARA, R. B. – SIMPSON, G. L. – SOLYMOS, P. – STEVENS, M. H. H. – SZOECS, E. – WAGNER, H. (2020): vegan: Community Ecology Package. – R package, version 2.5-7.

OLDEN, J. D. – LAWLER, J. J. – POFF, N. L. (2008): Machine learning methods without tears: A primer for ecologists. – The Quarterly Review of Biology 83(2): 171-193.

PHILIPSON, G. N. (1954): The effect of water flow and oxygen concentration on six species of caddis fly (Trichoptera) larvae. – Proceedings of the Zoological Society of London 124: 547-564.

PHILIPSON, G. N. – MOORHOUSE, B. H. S. (1974): Observations on ventilatory and net-spinning acitvities of larvae of the genus Hydropsyche (Trichoptera, Hydropsychidae) under experimental conditions. – Freshwater Biology 4: 525-533.

PÎRVU, M. – ZAHARIA, C. – SATMARI, A. – PÂRVULESCU, L. (2015): Spatial ecology of Hydropsyche incognita (Trichoptera: Hydropsychidae) in the Carpathians. – European Journal Of Entomology 112: 106-113.

PISCART, C. – LECERF, A. – USSEGLIO-POLATERA, P. – MORETEAU, J.-C. – BEISEL, J.-N. (2005): Biodiversity patterns along a salinity gradient: the case of net-spinning caddisflies. – Biodiversity and Conservation 14: 2335–2349.

PITSCH, T. (1993): Zur Larvaltaxonomie, Faunistik und Ökologie mitteleuropäischer Fließwasser-Köcherfliegen (Insecta: Trichoptera). – Landschaftsentwicklung und Umweltforschung – Schriftenreihe des Fachbereichs Landschaftsentwicklung – Sonderheft S 8: 316 pp. – Technische Universität Berlin.

ROKACH, L. (2010): Ensemble-based classifiers. – Artificial Intelligence Review 33: 1-39.

ROUX, C. – TACHET, H. – BOURNAUD, M. – CELLOT, B. (1992): Stream continuum and metabolic rate in the larve of five species of Hydropsche (Trichoptera). – Ecography 15: 70-76.

SALA, M. – FARIA, M. – SARASÚA, I. – BARATA, C. – BONADA, N. – BRUCET, S. – LLENAS, L. – PONSÁ, S. – PRAT, N. – SOARES, A. M. V. M. – CAÑEDO-ARGUELLES, M. (2016): Chloride and sulphate toxicity to Hydropsyche exocellata (Trichoptera, Hydropsychidae): Exploring intraspecific variation and sub-lethal endpoints. – Science of the Total Environment 566–567: 1032–1041.

ŠMILAUER, P. – LEPŠ, J. (2014): Multivariate Analysis of Ecological Data using CANOCO 5. (2nd edition) – Cambridge University Press, 376 pp.

STATZNER, B. – DOLÉDEC, S. (2011): Phylogenetic, spatial and species-trait patterns across environmental gradients: the case of Hydropsyche (Trichoptera) along the Loire River. – International Review of Hydrobiology 96: 121-140.

STUIJFZAND, S. C. – ENGELS, S. – VAN AMMELROOY, E. – JONKER, M. (1999): Caddisflies (Trichoptera: Hydropsychidae) used for evaluating water quality of large European rivers. – Archives of Environmental Contamination and Toxicology 36: 186-192.

TACHET, H. – PIERROT, J. P. – ROUX, C. – BOURNAUD, M. (1992): Netbuilding behaviour of six Hydropsyche species (Trichoptera) in relation to current velocity and dsitribution along the Rhône River. – Journal of the North American Benthological Society 11: 350-365.

TSZYDEL, M. – MARKOWSKI, M. – MAJECKI, J. (2016): Larvae of Hydropsyche angustipennis (Trichoptera, Hydropsychidae) as indicators of stream contamination by heavy metals in Łódź agglomeration. – Zootaxa 4138: 127-138.

VALAVI, R. – ELITH, J. – LAHOZ-MONFORT, J. J. – GUILLERA-ARROITA, G. (2021): Modelling species presence-only data with random forests. – Ecography 44(12): 1731-1742.

VALAVI, R. – GUILLERA-ARROITA, G. – LAHOZ-MONFORT, J. J. – ELITH, J. (2022): Predictive performance of presence-only species distributionmodels: a benchmark study with reproducible code. – Ecological Monographs 92(1): e01486.

VAN DER GEEST, H. G. – GREVE, G. D. – DE HAAS, E. M. – SCHEPER, B. B. – KRAAK, M. H. S. – STUIJFZAND, S. C. – AUGUSTIJN, K. H. – ADMIRAAL, W. (1999): Survival and behavioural responses of larvae of the caddisfly Hydropsyche angustipennis to copper and diazinon. – Environmental Toxicology and Chemistry 18: 1965-1971.

VAN PROOSDIJ, A. S. J. – SOSEF, M. S. M. – WIERINGA , J. J. – RAES, N. (2016): Minimum required number of specimen records to develop accurate species distribution models. – Ecography 39: 542–552.

VANNOTE, R. L. – MINSHALL, G. W. – CUMMINS, K. W. – SEDELL, J. R. – CUSHING, C. E. (1980): The River Continuum Concept. – Canadian Journal of Fisheries and Aquatic Sciences 27: 130-137.

VUORI, K-M. (1995): Species- and population-specific responses of translocated hydropsychid larvae (Trichoptera, Hydropsychidae) to runoff from acid sulphate soils in the River Kyrönjoki, western Finland. – Freshwater Biology 33: 305-318.

WALLACE, J. B. (1975): Food partitioning in net-spinning Trichoptera larvae: Hydropsyche venularis, Cheumatopsyche etrona and Macronema zebratum. – Annals of the Entomological Society of America 68: 463-472.

WALLACE, J. B. – WEBSTER, J. R. – WOODALL, W. R. (1977): The role of filter-feeders in flowing waters. – Archiv für Hydropbiologie 79: 506-532.

WARINGER, J. – GRAF, W. (2011): Atlas der mitteleuropäischer Köcherfliegenlarven – Atlas of Central European Trichoptera Larvae. – Erik Mauch Verlag, Dinkelscherben, 468 pp.

WOLPERT, D. H. (1992): Stacked generalization. – Neural Networks 5: 241-259.

ZHANG, L. – HUETTMANN, F. – ZHANG, X. – LIU, S. – SUN, P. – YU, Z. – MI, C. (2019): The use of classification and regression algorithms using the random forests method with presence-only data to model species' distribution. – MethodsX 6: 2281-2292.

ZINCHENKO, T. D. – GOLOVATYUK, L. V. (2013): Salinity tolerance of macroinvertebrates in stream waters (Review). – Arid Ecosystems 3: 113-121.

7. PUBLICATIONS

Publications related to the thesis

FICSÓR, M. – CSABAI, Z. (2021): Longitudinal zonation of larval Hydropsyche (Trichoptera: Hydropsychidae): abiotic environmental factors and biotic interactions behind the downstream sequence of Central European species. – Hydrobiologia 848: 3371–3388. **(IF= 2.822, SJR= Q1)**, DOI: 10.1007/s10750-021-04602-0

FICSÓR M. – CSABAI, Z. (2023): Machine learning model ensemble based on multi-scale predictors confirms ecological segregation and accurately predicts the occurrence of net-spinning caddisfly larvae species groups (Trichoptera: Hydropsychidae) at catchment-scale. – Ecological Indicators 146: 109769. **(IF= 6.9 (2022), SJR= D1)**, DOI: 10.1016/j.ecolind.2022.109769

Cumulative impact factor: 9,722

Independent citations: 10

Oral presentations related to the thesis

FICSÓR M. (2015): Adatok Észak-Magyarország és az Észak-Alföld szövőtegzes-faunájához (Trichoptera: Hydropsychidae) a fajok elterjedési mintázatának vizsgálatával lárvaadatok alapján. – XII. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Csapod, 2015.

FICSÓR, M. (2016): The influence of abiotic factors on the regional distribution of larval net-spinning caddisfly species (Trichoptera: Hydropsychidae) in Northern Hungary and in the Northern Great Plain. – 2nd Central European Symposium for Aquatic Macroinvertebrate Research, Pécs, 2016.

FICSÓR M. – VÁRBÍRÓ G. (2017): Abiotikus környezeti tényezők hatásának vizsgálata hegy- és dombvidéki előfordulású szövőtegzes-lárvák elterjedési viszonyaira észak-magyarországi vízfolyásokban. – XIII. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Pécs, 2017.

FICSÓR M. (2017): A *Hydropsyche modesta* Navás, 1925 (Trichoptera: Hydropsychidae) szövőtegzes-faj ökológiai igényeinek vizsgálata észak-magyarországi vízfolyásokban lárvaadatok alapján. – LIX. Hidrobiológus Napok, Tihany, 2017.

Ficsór M. (2019): Koegzisztencia vagy kompetíció? Hegy- és dombvidéki szövőtegzes-lárvák együttélési viszonyainak vizsgálata észak-magyarországi előfordulási adatok alapján. – XIV. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Debrecen, 2019.

Ficsór M. (2021): Mit kérdezhetünk, és milyen válaszokat várhatunk gépi tanulási (machine learning – ML) modellünktől fajok és fajcsoportok elterjedési mintázatának vizsgálata során? – Szövőtegzes-lárvák (Trichoptera: Hydropsychidae) példája észak-magyarországi vízfolyásokban. – XV. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Agárd, 2021.

Ficsór, M. – Csabai, Z. (2021): Seeing the forest for the trees – experiences of the application of non-linear machine learning algorithms in modelling the distribution of net-spinning caddisfly larvae (Trichoptera: Hydropsychidae) in North-Hungarian streams and rivers. – LXII. Hidrobiológus Napok, Tihany, 2021.

Ficsór, M. – Csabai, Z. (2022): Hydropsyche in the multiverse of rivers – a multi-scale, multi-model ensemble approach to explain the longitudinal distribution of larval net-spinning caddis. – LXIII. Hidrobiológus Napok, Tihany, 2022.

Ficsór M. – Csabai Z. (2023): Koegzisztencia vagy kompetíció II. - Elkülöníthetők-e az azonos ökológiai preferenciákkal jellemezhető Hydropsyche-fajcsoportok tagjai a környezeti változók hatásaira adott válaszaik alapján? – XVI. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Tihany, 2023.

Other publications

Ficsór M. – Nagy K. (2009): Referencia- és „kvázi"-referenciahelyek makroszkopikus vízi gerinctelen közösségeinek vizsgálata az Észak-magyarországi Környezetvédelmi, Természetvédelmi és Vízügyi Felügyelőség illetékességi területén. – Acta Biologica Debrecina Supplementum Oecologica Hungarica 20: 87-98.

Ficsór M. (2011): Adatok a folyami szitakötők (Odonata: Gomphidae) Északkelet-magyarországi előfordulásához lárvavizsgálatok alapján. – Acta Biologica Debrecina Supplementum Oecologica Hungarica 26: 67-74.

Ficsór, M. - Szabó, A. (2011): Contribution to the aquatic macroinvertebrate fauna of Szinva and its tributaries, NE Hungary. – Acta Biologica Debrecina Supplementum Oecologica Hungarica 26: 75-88.

BORZA, P. – CZIROK, A. – DEÁK, CS. – **FICSÓR, M.** – HORVAI, V. – HORVÁTH, ZS. – JUHÁSZ, P. – KOVÁCS, K. – SZABÓ, T.– VAD, CS. F. (2011): Invasive mysids (Crustacea: Malacostraca: Mysida) in Hungary: distributions and dispersal mechanisms. North-Western Journal of Zoology 7(2): 222-228.

VÁRBÍRÓ, G. – FEKETE, O. – ORTMANN-AJKAI, A. – **FICSÓR, M.** – CSER, B. – KOVÁCS, K. – KISS, G. – CZIROK, A. – HORVAI, V. – DEÁK, CS. (2011): Developing a multimetric macroinvertebrate index on montainous, small and medium sized water bodies. – Acta Biologica Debrecina Supplementum Oecologica Hungarica 26: 211-220.

FICSÓR, M. (2013): Contribution to the aquatic mollusc fauna of Northern Hungary and the Northern Great Plain. Part I: Gastropoda. – Acta Biologica Debrecina Supplementum Oecologica Hungarica 31: 41-62.

FICSÓR, M. (2014): Contribution to the freshwater Mollusc fauna of Northern Hungary and the Northern Great Plain. Part II: Bivalvia. – Acta Biologica Debrecina Supplementum Oecologica Hungarica 32: 51-65.

CSERCSA, A. – BOZÓKI, T. – KRASZNAI, E. Á. – **FICSÓR, M.** – VÁRBÍRÓ, G. (2015): Contribution to the aquatic macroinvertebrate fauna of the Eger-apatak (eger stream) in Northern Hungary. – Folia Historico-Naturalia Musei Matraensis 39: 5-16.

KRASZNAI, E. Á. – BODA, P. – CSERCSA, A. – **FICSÓR, M.** – VÁRBÍRÓ, G. (2016): Use of self-organizing maps in modelling the distribution patterns of gammarids (Crustacea: Amphipoda). – Ecological Informatics 31: 39-48.

MÁLNÁS, K. – KOVÁCS, K. – **FICSÓR, M.** – JUHÁSZ, P. – MÜLLER, Z. – OLAJOS, P. – KISS, B. (2016): Appearances of the non-indigenous Helobdella europaea Kutschera, 1987 (Hirudinea, Glossiphoniidae) in Hungarian watercourses. – Folia Historico-Naturalia Musei Matraensis 40: 17-20.

Other oral and poster presentations

FICSÓR M. – NAGY K. – IMRE A. (2006): Az Észak-magyarországi Környezetvédelmi, Természetvédelmi és Vízügyi Felügyelőség illetékességi területén 2005-ben a Víz Keretirányelv keretén belül végrehajtott makrozoobenton-vizsgálatok. – III. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Göd, 2006.

FICSÓR M. – NAGY K. (2007): Makroszkopikus vízi gerinctelen közösségek vizsgálata különböző biológiai vízminősítő módszerek (ASTERICS, karakterfaj-elemzés, BMWP) tükrében az ÉMI-KTVF illetékességi területén 2006-ban. – IV. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Tihany, 2007.

FICSÓR M. – NAGY K. (2009): Referencia- és „kvázi”-referenciahelyek makroszkopikus vízi gerinctelen közösségeinek vizsgálata az Észak-magyarországi Környezetvédelmi, Természetvédelmi és Vízügyi Felügyelőség illetékességi területén. – VI. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Villány, 2009.

FICSÓR M. – SZABÓ A. (2011): Hosszú távú és szezonális vizsgálatok a Szinva-patak és befolyóinak makroszkopikus vízi gerinctelen faunáján. – VIII. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Jósvafő, 2011.

FICSÓR M. (2011): Adatok a folyami szitakötők (Odonata: Gomphidae) északkeletmagyarországi előfordulásához lárvavizsgálatok alapján. – VIII. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Jósvafő, 2011.

VÁRBÍRÓ G. – FEKETE O. – ORTMANN-NÉ AJKAI A. – **FICSÓR M.** – CSER B. – KOVÁCS K. – KISS G. – CZIROK A. – HORVAI V. – DEÁK CS. (2011): Makrogerinctelen multimetrikus index kidolgozása hegyvidéki vízfolyásokra. – VIII. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Jósvafő, 2011.

FICSÓR M. – SZABÓ A. (2012): A Szinva patak vízminőségének vizsgálata a makrogerinctelen fauna alapján – A Multimetrikus Makrozoobenton Index (HMMI) a gyakorlatban. – IX. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Gyula, 2012.

BORZA P. – CSER B. – CZIROK A. – DEÁK CS. – **FICSÓR M.** – HORVAI V. – HORVÁTH ZS. – KOVÁCS K. – PETRI A. – VAD CS. F. (2013): Adatok a síkvidéki felszíni Niphargus-fajok (Crustacea, Amphipoda, Niphargidae) magyarországi elterjedéséhez. – X. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Szalafő, 2013.

FICSÓR M. (2013): Adatok Észak-Magyarország és az Észak-Alföld vízi puhatestű faunájának ismeretéhez: I rész. Csigák. – X. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Szalafő, 2013.

KOVÁCS K. – **FICSÓR M.** (2013): Új Helobdella faj Magyarországon? – X. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Szalafő, 2013.

FICSÓR, M. (2014): Contribution to the aquatic Mollusc fauna of Northern Hungary and the Northern Great Plain. Part II: Bivalvia. – 1st Central European Symposium for Aquatic Macroinvertebrate Research, Szarvas, 2014.

CSERCSA A. – BOZÓKI T. – KRASZNAI E. – VÁRBÍRÓ G. – **FICSÓR M.** – BODA P. (2015): Az Eger-patak hidrozoológiai vizsgálata. – XII. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Csapod, 2015.

BOZÓKI T. – CSERCSA A. – **FICSÓR M.** – KRASZNAI E. – VÁRBÍRÓ G. – BODA P. (2017): Az urbanizáció hatása a vízi makrogerinctelen közösségekre az Eger-patakon. – LIX. Hidrobiológus Napok, Tihany, 2017.

SZELES J. – BOZÓKI, T. – **FICSÓR M.** – B-BÉRES V. – DRENOVÁCZ M. – BÉKÉSI CS. – NAGY S. A. – VÁRBÍRÓ G. (2021): Síkvidéki kisvízterek válasza az eltérő vízszintekre. – XV. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Agárd, 2021.

VÁRBÍRÓ G. – **FICSÓR M.** – ZAGYVA A. T. – BODA P. (2021): A harmadik vízgyűjtő-gazdálkodási terv makroszkopikus vízi gerinctelenekre vonatkozó eredményei. – XV. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Agárd, 2021.

SZELES, J. – BOZÓKI T. – **FICSÓR M.** – B-BÉRES V. – NAGY S. A. – DRENOVÁCZ M. – BÉKÉSI CS. – VÁRBÍRÓ G. (2022): Eltérő minőségű síkvidéki vízterek makroszkopikus gerinctelen élőlényközösségének trait alapú elemzése. – LXIII. Hidrobiológus Napok, Tihany, 2022.

BOÓZ B. – **FICSÓR M.** – PERNECKER B. – MÓRA A. – CSABAI Z. (2023): Elhanyagolt kétszárnyúak: Megjósolhatjuk-e a figyelmen kívül hagyott családok előfordulási jellemzőit a gyakran vizsgált és ismertebb családok alapján? – XVI. Makroszkopikus Vízi Gerinctelenek Kutatási Konferencia, Tihany, 2023.