ISSN 2336-3193
DOI: 10.2478/cszma-2020-0002

Acta Mus. Siles. Sci. Natur., 69: 21-28, 2020
Published: online 31 March 2020, print March 2020

Species distribution model for *Buxbaumia viridis,* identifying new areas of presumed distribution in the Czech Republic.

Lukáš Číhal, Lucie Fialová & Vítězslav Plášek

Species distribution model for *Buxbaumia viridis,* identifying new areas of presumed distribution in the Czech Republic. – Acta Mus. Siles. Sci. Natur. 69: 21-28, 2020.

Abstract: There is a large number of organisms in nature that are hard to find in the field, and one of these is also a rare bryophyte species, *Buxbaumia viridis*, which can also only be found in a certain part of the year in the production of sporophytes. Using distribution models, however, it is possible to predict the locations which this species is likely to be present at and to focus the field research directly on such sites. Thanks to the presented model, we have succeeded in identifying areas within the Czech Republic with an increased likelihood of its occurrence, even in some areas where no collection of *B. viridis* has so far been recorded. With the help of the maps created these locations can be identified with the accuracy of approximately 1 km^2, making it easier for terrain research and resulting time and often financial expenses. We have also focused on the study of ecological factors that influence the occurrence of the species, and by testing variables from different categories we have come to some agreement with previously published studies. Whilst we have managed to exclude that air pollution has a major role to its occurrence, which has not yet been the subject of any study.

Key words: *Buxbaumia viridis*, MaxEnt, distribution modelling

Introduction

Buxbaumia viridis (Moug. ex Lam. & DC.) Brid. ex Moug. & Nestl. is an epixylous moss easily distinguishable mainly by conspicuous large green, later pale brown, capsules. They appear mainly in autumn and persist with little changes until the following spring. The occurrence of the species is generally limited to decaying wood, especially in advanced phases of decay (Wiklund 2002, 2003; Plášek 2004).

Nevertheless, the loss of woodland cover over the centuries and, more recently, the intense management of woodland areas has led to a significant loss of habitat for bryophytes. It seems that the crucial factor for the occurrence of the species and its long-term survival are probably suitable humidity, typically near a watercourse, with a sufficient amount of decayed wood.

Although *Buxbaumia viridis* occurs in most of Europe, it is rare in many parts of the continent. The species is considered as vulnerable in Europe (Schumacker & Martiny 1995), and moreover it was included on appendix I of the Bern Convention and is also included in Annex II of the EC Directive 92/43/EEC on conservation of natural habitats and of wild fauna and flora. Outside Europe, it has a scattered distribution in forest habitats across the whole northern hemisphere from south-west Asia across China to North America.

In the Czech Republic there have been recorded more than 300 localities during the last two decades (AOPK ČR 2018). Most of the recently known localities in the Czech Republic are situated in the border mountains, e.g. the Moravskoslezské Beskydy Mts, the Hrubý Jeseník Mts (Plášek 2004; Holá *et al*. 2014), where the species grow mainly in primeval forests. It grows there with respect to very specific habitat requirements - on decaying wood in wet, shaded places.

Although many floristic and ecological data from the Czech Republic were collected and analysed, there are a lot of suitable areas, from which the data about the occurrence of *Buxbaumia viridis* is still missing. That is why the distribution model was prepared. Predictions

sciendo

of the geographical distribution based on available information could be beneficial for future field research (*cf*. Bourg *et al*. 2005; Yu *et al*. 2013), and could accelerate the discovery of next populations of the rare epixylous species.



Fig 1: *Buxbaumia viridis*. Individual stages of sporophyte development in which the species can be reliably idetified in the field. Photos by V. Plášek.

Methods

MaxEnt software version 3.4.1 (https://biodiversityinformatics.amnh.org/open_source/ maxent) (Phillips *et al.* 2017) has been used to create model and to prepare the data and the individual layers, the programme R (R Development Core Team 2018) and QGIS (QGIS Development Team 2018) were used.

As for the analysis, raster layers have been used, which were initially selected in a total of 33 (Tab. 1), namely 19 bioclimatic layers (30 arcsec) - WorldClim version 2 (http://worldclim.org/version2) (Fick *et al*. 2017). In addition, 11 layers taking up the pollutant values from 2007 to 2011 downloaded from the Czech Hydrometeorological Institute website (http://portal.chmi.cz/files/portal/docs/uoco/isko/ozko/ozko_CZ.html, Czech Hydrometeorological Institute), which were in the programme R transferred from vector to raster data using the rasterization from the raster package (Hijmans 2015). Furthermore, it was also a layer of soil cover type, Corine Land Cover (CLC) 2012, version 18.5.1 (https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012?tab=download, European Environment Agency - EEA), whereas the distribution of the land cover into five categories were used for the purposes of the analyses: 1st urbanized areas, 2nd agricultural area, 3rd forests and the semi-natural areas, 4th humid territory and 5th water surface.

Table 1: Complete list of layers

Order number:	Layer:	Explanatory notes:	Units
Bioclimatic variables			
1	BIO1 **	annual mean temperature	°C * 10
2	BIO2	mean diurnal range (mean of monthly (max temp - min temp))	°C * 10
3	BIO3*	isothermality (bio2/bio7) (* 100)	°C * 10
4	BIO4*	temperature seasonality (standard deviation *100)	°C * 10
5	BIO5	max temperature of warmest month	°C * 10
6	BIO6*	min temperature of coldest month	°C * 10
7	BIO7	temperature annual range (bio5-bio6)	°C * 10
8	BIO8 *	mean temperature of wettest quarter	°C * 10
9	BIO9*	mean temperature of driest quarter	°C * 10
10	BIO10	mean temperature of warmest quarter	°C * 10
11	BIO11	mean temperature of coldest quarter	°C * 10
12	BIO12 **	annual precipitation	mm
13	BIO13	precipitation of wettest month	mm
14	BIO14	precipitation of driest month	mm
15	BIO15*	precipitation seasonality (coefficient of variation)	mm
16	BIO16*	precipitation of wettest quarter	mm
17	BIO17	precipitation of driest quarter	mm
18	BIO18	precipitation of warmest quarter	mm
19	BIO19	precipitation of coldest quarter	mm
5-year average concentration of imissions 2007-2011			
20	Arsenic *	annual average concentration of arsenic	[µg.m -3]
21	Benzene *	annual average concentration of benzene	[µg.m -3]
22	Benzo(a)pyrene*	annual average concentration of benzopyrene	[µg.m -3]
23	Cadmium*	annual average concentration of cadmium	[µg.m -3]
24	Lead*	annual average concentration of lead	[µg.m -3]
25	NO2*	annual average concentration of NO2	[µg.m -3]
26	Nickel*	annual average concentration of nickel	[µg.m -3]
27	PM10	annual average concentration of airborne dust up to 10 µm	[µg.m -3]
28	PM10 _24	annual average concentration of airborne dust maximum values up to 10 µm in 24 hours per calendar year	[µg.m -3]
29	PM25	annual average concentration of airborne dust up to 25 µm	[µg.m -3]
30	SO2*	annual average concentration of SO2 maximum values in 24 hours per calendar year	[µg.m -3]
Sheet strata			
31	Corine land cover*	soil cap map	
32	Treecover*	the ratio of the area covered by the branches and the leaves of the trees projected on the soil cap	%
Distance layers			
33	Waterdist*	distance from the nearest watercourse	m

* Layers used for final analysis.
** Layers removed before the hidden correlation check.

Next the Percent Tree Cover (https://github.com/globalmaps/gm_ve_v2, Vegetation (Proceeding Tree Cover) - Version 2 © Chief University of Chemical Technology, Chiba University and Collaborating Organizations), which has been resampled with the bilinear interpolation method using the resample function from the raster package (Hijmans 2015). Finally, the layer of the distance from a watercourse was created and a fine water section layer (A03) from the digital database of water management data (DIBAVOD 2007) has been used for its calculation (available at http://www.dibavod.cz/index.php? id = 27) using the distance function from the raster package (Hijmans 2015). However, the two layers (BIO1, BIO12), which were considered inappropriate for analysis, have been removed from these variables at the very beginning, because it is preferable to select limit variables for analyses as they have more biological significance for species distribution than the factors such as annual average rainfall, temperatures, etc. (Pradhan 2016).

Before using layers in own models, we first removed the correlation between layers with a hidden correlation check, which calculates the variability inflation factors of all predictors in regression models, using the vifcor function of the usdm package (Naimi *et al*. 2014) with setting up the correlation factor to r= 0.8 recommended by Menard (2002). A total of 13 variables from 31 input variables had a problem with collinearity and thus were removed from the following analyses: BIO2, BIO5, BIO7, BIO10, BIO11, BIO13, BIO14, BIO17, BIO18, BIO19, PM10, PM10_M36 and PM25. The remaining 18 layers were found to be suitable for further analysis: Arsenic, Benzo (a) Pyrene, Benzene, BIO3, BIO4, BIO6, BIO8, BIO9, BIO15, BIO16, Cadmium, Corine land cover, Lead, Nickel, NO2, SO2, Treecover, Waterdist (Table 1). The set of these layers was then used for the analysis in the programme Maxent and included in the final model all the layers that contributed to the model with a value greater than 5%.

To reduce the bias, filtering through a 0.1 x 0.1-degree grid (WGS84 format), was used when only one sample was left per pixel of the raster using the gridSample feature of the dismo package (Hijmans *et al*. 2017). The AOPK ČR PLA Administration (AOPK ČR 2018) was used as the source of these data, which provided us with a total of 313 findings. From the database, however, only the findings that had a GPS position and were not older than 1999 (inclusive) have been used for the research purposes. In total 76 collections were used after the bias reduction and filtering the collections.

In order to monitor the data structure, we used the dudi.pca function of the ade4 package (Dray *et al*. 2007) that allowed us to perform a PCA analysis for the entire study area and thus we were able to check the data due to the occurrence of potential outliners (points located alone or far from others) that could negatively affect further analysis. After the inspection it became clear that there were no "outliners" in the data, so there was no need to edit the data again.

For the specification of background area (the area where the variables are studied) we have used the whole Czech Republic area, which seems to be suitable for its size primarily because the background points (the point specifying the variables in the area studied) are not selected from either too large or too small areas. Czech Republic is also well-researched area with good accessibility of the individual layers used for modelling. 10.000 background points were used for the analysis.

AUC evaluation score based on a 10-fold cross-validation was used to validate the models. For each of the AUC scores the values are: <0.6 invalid, 0.6-0.7 weak, 0.7-0.8 fair, 0.8-0.9 good, and 0.9-1.0 excellent (Swets 1988). The selection of the features was performed automatically and for regularization multipliers were selected β value of 1.

Results

Based on the receiver operating characteristic (ROC) curve averaged over the 10-fold cross-validation runs, the average test AUC for the replicate runs was 0.914. This means that the model falls into an excellent category (Swets 1988).

As factors with the highest relative value of the environmental variables in Maxent model were evaluated tree cover (44.6%), corine land cover (22.4%), Bio16 (11.7%), BIO9 (11.1%) and waterdistance (10.3%). According to the curves, which show how the predicted probability of presence changes according to environmental variable, the dependence of the probability of occurrence on the resulting factors was as follows: With the increasing value of the tree cover variable the probability of the occurrence increased, the highest probability of occurrence in the corine land cover was in category 3 (forests and the semi-natural areas), the higher rainfall in the most humid quarter (BIO 16) also increased the probability, but the highest values of this variable were slightly decreased, while higher values for the meanest temperature of the driest quarter (BIO9) reduced the probability and with increasing the distance from the watercourses the probability first decreased sharply and then increased slightly.

The border zones of the mountain ranges from the Jeseníky and Beskydy Mountains to the Rychlebské Mountains, the Kralický Sněžník Mountains, the Orlické Mountains, the Krkonoše Mountains, the Jizerské Mountains, the Krušné Mountains and the Šumava Mountains to the Novohradské Mountains proved to be the regions that are most suitable for the occurrence of the species studied. In addition to the above-mentioned mountains, it was also the area of Nízký Jeseník, military area Libavá, Hostýnsko Vsetínská hornatina, Drahanská vrchovina, Chřiby, Žďárské vrchy and Javořická vrchovina. The most interesting areas of these are the locations where the species studied has not been collected yet, namely the Krušné Mountains, the Jizerské Mountains, the Novohradské Mountains, Libavá, Chřiby and Žďárské

vrchy. Using the predictive model, we have thus been able to identify the areas that should be targeted when searching for the species studied as they show a strong resemblance to previously known locations.

As the part of the research, we also provide geographic map of the model identifying regions that have similar environmental conditions to those in which the species are currently populated (Fig. 1).

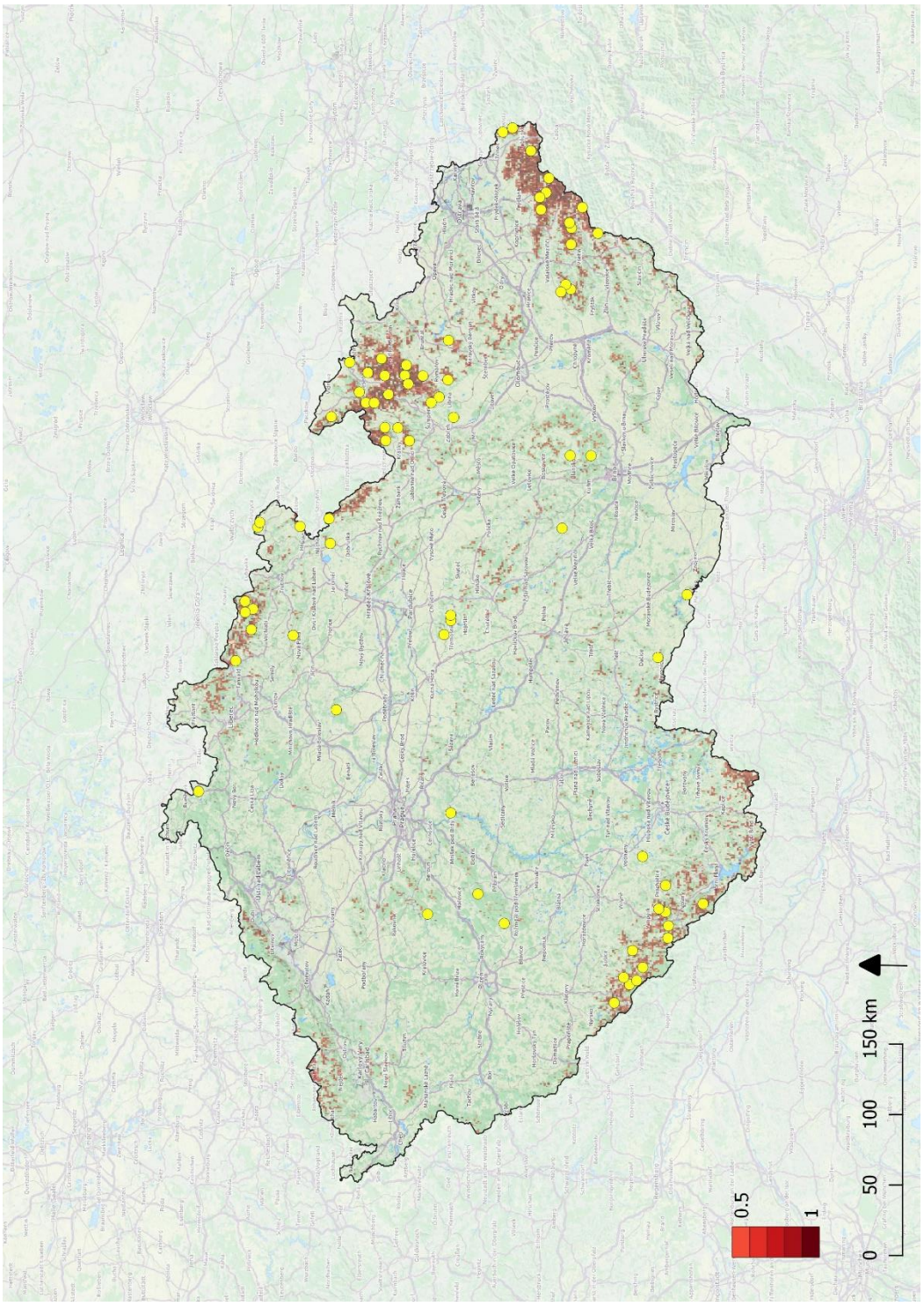


Fig 1: Map identifying regions that have similar environmental conditions to currently found populations of *Buxbaumia viridis* visualized on a scale from 0.5 to 1 (higher than random) with displayed samples (unbiased collection points) on OpenStreetMap background.

Discussion

Thanks to the fact that the species *Buxbaumia viridis* is classified as vulnerable in Europe (Schumacker & Martiny 1995) as well as in Annex I of the Bern Convention and it is also included in Annex II of the EC Directive 92/43 / EEC, it is the species that special attention is paid to. On the other hand, an equally important fact, which greatly impairs obtaining the information about the distribution of this species is that it is easily recorded only upon the production of sporophytes. That is why the collection of all biological and population-ecological data directly in the field is difficult. Therefore, most of the data collecting had to be based solely on observation of the capsules and *Buxbaumia viridis* can be a much more common species than we can admit, but it can produce sporophytes only very rarely or very irregularly.

Three distribution models have already been developed to help address this issue. The first was a distribution model focused on the eastern central Sweden (Wiklund 2002), the second model was focused on the territory of the Czech Republic (Holá *et al*. 2014) and the last on the Alpine region of South Tyrol, located in north-eastern Italy (Spitale & Mair 2015).

The basic difference between the newly created model and the older ones is the fact that all the earlier models were modelled by logistic regression (GLM). However, these models require two types of data (present and absent) and as Spitale & Mair (2015) admits, the models do not take into account the imperfect detection of the species and consider all the absences as true, but this may not be the case because it may happen that the species is only overlooked in the field or it is not - as a sterile population without capsules - found and such data become false negatives that affect the model in a negative way. In the case of *Buxbaumia viridis*, this error is more likely to happen due to the greater difficulty of its detection. Unlike GLM, we were able to create models based on the data that are only present using maximum entropy modelling (MaxEnt) thus avoiding this.

Contrary to earlier models, other input variables (layers) were tested, with a total of 31 (Tab. 1). The layers were gradually tested for the mutual correlation and then either retained or eliminated according to their contribution to the final model. After gradual removal of all layers that had a total contribution to the final model of less than 5%, we obtained a total of five layers, which largely corresponded to the previously observed results and they refer to a selection of suitable variables that reflect the real demands of the studied species. They were the Tree cover first, then Corine Land cover 2012, Bio 16, Bio9 and lastly water distance. The results thus confirm mainly the importance of Tree cover for the occurrence of *Buxbaumia viridis*. This is also mentioned, for example, in Spitale & Mair (2015), which states that the probability of occurrence with canopy closure increased on a local scale, mainly due to the microclimatic conditions affecting it.

Corine land cover, which has proved to be the second most important factor and refers to the dependency of *Buxbaumia viridis* on certain habitat types, is probably related mainly to need for a sufficient amount of decayed wood to the site (Wiklund 2002). However, it is not recorded in any existing database and therefore, the variable is not applicable to direct modelling (Holá *et al*. 2014). Other two environmental layers that have proved to be important for predicting *Buxbaumia viridis* were BIO 9 (Mean Temperature of Driest Quarter) and BIO16 (Precipitation of Wettest Quarter). Regarding the fact that most precipitation in the Czech Republic falls in June or July, and the least amount of rainfall in January or February (Tolasz *et al*. 2007), we can assume that the dependence of *Buxbaumia viridis* on higher precipitation during the summer months (BIO16) plays its part because they keep good humidity conditions at the site during elevated summer temperatures. The dependency of the species on lower average temperatures in the winter months (BIO9) is not obvious, and it may initiate a discussion. However, both variables are likely to depend on a colder environment. In our case, the distance from the watercourse was the last variable that influenced the resulting model,

whereas Spitale & Mair (2015) claims this variable may reflect the relationship with the geomorphology of the studied area, especially if the coarse sections of the flows are included in the model, as this is our case. However, the distance from the river influences the local climate and is an important environmental variable that affects species distribution (Vončina 2011; Holá *et al*. 2014).

The evaluation of the variables influencing the occurrence of the studied species in the Czech Republic in our model corresponds to a certain extent to the results of previous analyses, even though the layers that had not been previously included in the models were used. On the contrary, it was not confirmed that air pollution, which proved to be marginal, affected the occurrence. Compared with previous models, we were not influenced by false negatives, which could result from the difficulty of detecting the studied species directly in the field, due to the use of the MaxEnt method. We have also achieved much greater details of the resultant map due to its resolution of 30 arcsec suitable for field use. For comparison, a previously published map for the Czech Republic (Holá *et al*. 2013) had a resolution of 5 × 3 arcmin. Thanks to this map we are not only able to predict areas with increased probability of occurrence, but we are also able to select specific locations that can be explored using the predictive map directly in the field (cf. Číhal & Kaláb 2017). Exploring recently identified locations with a higher probability of occurrence and with no collections yet would be the most appropriate.

Acknowledgements: The research was financially supported by the Ministry of Culture of the Czech Republic by institutional financing of long term conceptual development of the research institution (the Silesian Museum, MK000100595), internal grant of the Silesian Museum No. IGS201801/2018.

References

AOPK ČR (2018): Nálezová databáze AOPK ČR. URL: http://portal.nature.cz/ (Accessed on 2018-10-24).

Bourg N.A., McShea W.J. & Gill D.E. (2005): Putting a CART before the search: successful habitat prediction for a rare forest herb. – Ecology 86(10): 2793-2804.

Číhal L. & Kaláb O. (2017): Species distribution models for critically endangered liverworts (Bryophyta) from the Czech Republic: a guide to future survey expeditions. – Acta Musei Silesiae, Scientiae Naturales 66(2): 101-110.

DIBAVOD (2007): Výzkumný ústav vodohospodářský T. G. Masaryka, v.v.i., odbor ochrany vod a infiormatiky, odd. GIS. Available from url: http://www.dibavod.cz.

Hijmans R.J., van Etten J., Cheng J., Mattiuzzi M., Sumner M. *et al*. (2015): Package 'raster'. R package.

Hijmans R.J., Phillips S., Leathwick J., Elith J. & Hijmans M.R.J. (2017): Package 'dismo'. Circles 9.

Holá E., Vrba J., Linhartová R., Novozámská E., Zmrhalová M., Plášek V. & Kučera J. (2014): Thirteen years on the hunt for *Buxbaumia viridis* in the Czech Republic: still on the tip of the iceberg? – Acta Societatis Botanicorum Poloniae 83(2): 137-145.

Menard S. (2002): Applied logistic regression analysis. Vol. 106. Sage.

Naimi B., Hamm N.A., Groen T.A., Skidmore A.K. & Toxopeus A.G. (2014): Where is positional uncertainty a problem for species distribution modelling? – Ecography 37(2): 191-203.

Phillips S.J., Dudík M. & Schapire R.E. (2017): Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed 2018-10-24.

Plášek V. (2004): The moss *Buxbaumia viridis* (Bryopsida, Buxbaumiaceae) in the Czech part of the Western Carpathians – distribution and ecology. Bryological studies in the Western Carpathians. Poznań: Sorus, pp. 37-44.

Pradhan P. (2016): Strengthening MaxEnt modelling through screening of redundant explanatory bioclimatic variables with variance inflation factor analysis. Researcher 8(5): 29-34.

QGIS Development Team (2018): QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available from http://qgis.osgeo.org

R Development Core Team (2018): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/.S

Schumacker R. & Martiny P.H. (1995): Threatened bryophytes in Europe including Macaronesia. Red data book of European bryophytes. Part 2., pp. 29-193.

Spitale D. & Mair P. (2017): Predicting the distribution of a rare species of moss: The case of *Buxbaumia viridis* (Bryopsida, Buxbaumiaceae). Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology 151(1): 9-19.

Swets J.A. (1988): Measuring the accuracy of diagnostic systems. – Science 240(4857): 1285-1293.

Tolasz R., Míková T., Valeriánová A. & Voženílek V. (2007): Atlas podnebí Česka. ČHMÚ, UPOL.

Vončina G., Cykowska B., & Chachuła P. (2011): Rediscovery of *Buxbaumia viridis* (Bryophyta, Buxbaumiaceae) in the Tatra and Gorce in the Polish Western Carpathians. Chorological studies on Polish Carpathian bryophytes. Poznań: Sorus, pp. 171-176.

Wiklund K. (2002): Substratum preference, spore output and temporal variation in sporophyte production of the epixylic moss *Buxbaumia viridis*. Journal of Bryology 24(3): 187-195.

– (2003): Phosphorus concentration and pH in decaying wood affect establishment of the red-listed moss *Buxbaumia viridis*. Canadian Journal of Botany 81(6): 541-549.

Yu J., Ma Y.H. & Guo S.L. (2013): Modeling the geographic distribution of the epiphytic moss *Macromitrium japonicum* in China. – Annales Botanici Fennici 50(1-2): 35-42.

Authors' addresses: Lukáš Číhal, Silesian Museum, Nádražní okruh 31, CZ-746 01 Opava, Czech Republic. E-mail: cihal@szm.cz

Lucie Fialová & Vítězslav Plášek, University of Ostrava, Dept. of Biology & Ecology, Chittussiho 10, Ostrava, Czech Republic.
E-mail: fialova.lucie92@gmail.com; vitezslav.plasek@osu.cz