

METODIKA PREDIKČNÍHO MODELOVÁNÍ VÝSKYTU VZÁCNÝCH A OHROŽENÝCH DRUHŮ MECHOROSTŮ A LIŠEJNÍKŮ NA ÚZEMÍ ČR



Botanický ústav
AV ČR, v. v. i.
Zámek 1
252 43 Průhonice
Česká republika

Oddělení GIS a DPZ
<http://labgis.ibot.cas.cz>

Autoři:

Mgr. Matěj Man, matej.man@ibot.cas.cz
doc. Ing. Jan Wild, Ph.D.

T A
Č R

Tato metodika (*Metodika predikčního modelování výskytu vzácných a ohrožených druhů mechorostů a lišejníků na území ČR*) byla vytvořena s finanční podporou TA ČR v rámci projektu TH02030644, Systém predikčního modelování výskytu ohrožených druhů mechorostů a lišejníků pro aplikované využití v ochraně přírody.

Foto na titulní straně: *Buxbaumia viridis*. Evropsky významný druh mechorostu, jeden z ohrožených druhů, pro který může tato metodika pomoci odhalit nové lokality výskytu v ČR. Autor fotografie: Štěpán Koval, 2014, <https://www.bryo.cz>, upravil Matěj Man.

Obsah

1	Abstrakt	4
2	Cíle metodiky	6
3	Teoretický úvod	7
3.1	Znalost výskytu mechorostů a lišejníků	7
3.2	Ochrana mechorostů a lišejníků	9
3.3	Vzácné, nebo jen málo probádané?	10
3.4	Vyhledávání vzácných druhů v ČR	11
3.5	Zahraniční zkušenosti se zacíleným průzkumem	12
3.6	Modelování potenciální vhodnosti habitatu	13
3.6.1	Koncept, předpoklady, potenciální limity	14
3.6.2	Environmentální data	17
3.6.3	Trénovací výskytová data	20
3.6.4	Pseudo-absence	22
3.6.5	Modely a algoritmy	24
3.6.6	Performance a validace modelů	28
3.6.7	Limity a rizika SDM	30
3.6.8	Další užitečné zdroje a literatura	30
4	Praktické vyhledávání vzácných druhů	31
4.1	Vstupní výskytová data	33
4.2	Environmentální data	38
4.3	Oblast predikce	40
4.4	Interakce uživatele s predikční aplikací	42
4.5	Model a jeho výstup	44
4.6	Mapa potenciální vhodnosti habitatu	48
4.7	GIS a terénní práce	49
4.7.1	Vizualizace	50
4.7.2	Plánování terénního průzkumu	53
4.7.3	Metoda biodiversity hotspots	58
4.8	Případová studie NP České Švýcarsko	59
5	Novost metodiky	60
6	Popis uplatnění metodiky	60
7	Ekonomické aspekty	61
8	Citovaná literatura	62
9	Použité zkratky	66

1 Abstrakt

Metodika nabízí certifikovaný postup, jak využít modelování potenciální vhodnosti habitatu pro praktické vyhledávání nových lokalit vzácných druhů mechorostů a lišejníků na území ČR. Metodika je postavena na online predikčním software, který byl za tímto účelem vyvinut autory metodiky. Online software je zdarma dostupný široké veřejnosti a poskytuje výpočetní algoritmy, výpočetní výkon i environmentální data. Uživatel metodiky nahraje do software trénovací výskytová data cílového druhu, nebo si je vybírá z Národní Výskytové Databáze Mechorostů a Lišejníků – Database of Lichens and Bryophytes (DaLiBor)¹, do které má software přístup. Výstupem je georeferencovaná mapa, kterou je možné dále vyhodnotit a vizualizovat v geografickém informačním software. Metodika zavádí postup, jak využít tyto mapy k plánování terénního průzkumu za účelem nalezení zájmových druhů mechorostů a lišejníků. Cílem této metodiky je poskytnout jednotný nástroj pro orgány ochrany přírody i odbornou veřejnost, který by měl podpořit bryologický a lichenologický průzkum na území ČR. Metodika pomůže tento průzkum zacílit a tím i zefektivnit co do lidských i finančních zdrojů. Metodika se ve své první části věnuje metodám a postupům z teoretického hlediska, staví na výsledcích současného výzkumu, na které odkazuje. Připojené jsou také příklady úspěšných aplikací ze zahraničí. Na tomto teoretickém základu pak staví praktická část metodiky, která je koncipována jako návod či popis postupu, jak by měl uživatel krok za krokem získat co nejkvalitnější mapu potenciální vhodnosti habitatu a jak ji efektivně využít při terénním průzkumu. Metodika se vyhýbá využití komerčního software, aplikací a datových formátů. Proto je postavena na využití open source software, zdarma dostupných aplikací a svobodných licencích.

Klíčová slova: mechorosty, lišejníky, modelování distribuce druhů, vzácné druhy, terénní průzkum

¹ <https://dalibor.ibot.cas.cz>

Abstract:

The methodology introduced certified procedure of using species distribution modelling to support discovery of new localities with presence of rare species of bryophytes and lichens in Czech Republic. The procedure is based on the prediction software developed by the authors. The software is online, free available and open source. It is available for general public, it provides algorithms, computation power and environmental data. User provides training data – coordinates of species presence. From the software interface it is also possible to request the coordinates directly from Database of Lichens and Bryophytes (DaLiBor)². Georeferenced map is the output form the software. This map is ready to be visualised and analysed in geographic information software (GIS). This methodology introduced the procedure how to use these maps for planning the field survey to discover new localities with presence of rare species of bryophytes and lichens. The aim of the methodology is to provide integrated tool for conservation planning and research to promote the bryological and lichenological field survey in Czech Republic. The methodology will help to focus the survey which could make the survey more efficient and cheaper according to money but also human resources. First half of the methodology brings the theoretical background based on results of recent research. Examples of successful abroad case studies are included. Practical part of the methodology builds upon that theoretical background. Practical part is a tutorial how to get the best map of potential habitat suitability and how to effectively use the map for the field survey. The methodology avoids using a commercial software, proprietary data formats and applications which could be limiting for some users. Therefore, it is based exclusively on free software, mostly open source and free licence.

Key words: bryophytes, lichens, species distribution modelling, rare species, field survey

² <https://dalibor.ibot.cas.cz>

2 Cíle metodiky

Dobrá znalost rozšíření vzácných či ohrožených druhů je zásadním předpokladem jejich efektivní ochrany. Ať už je to ochrana legislativního či technického rázu. U některých vzácných či ohrožených druhů mechorostů a lišejníků je stále velký potenciál k objevování nových lokalit v porovnání například s cévnatými rostlinami, kde je prozkoumanost našeho území v tomto ohledu výrazně větší. To může být dáno menší komunitou odborníků (bryologů a lichenologů) či složitější determinací mechorostů a lišejníků v terénu. Na našem území navíc tradičně probíhá spíše inventarizační průzkum orientovaný na cenná území než na konkrétní druhy. To je většinou z ochránářského hlediska v pořádku a odpovídá to poptávce po informacích sloužících orgánů ochrany přírody. Inventarizační či obecně floristické průzkumy mají navíc velkou metodickou i softwarovou podporu. Nové lokality vzácných či ohrožených druhů mechorostů a lišejníků jsou tak nalézány spíše náhodně při floristických průzkumech, i když iniciativy zaměřené na cílené vyhledávání konkrétních vzácných druhů na našem území také existují, zejména u druhů evropsky významných (*Orthotrichum rogeri*, *Hamatocaulis vernicosus*, *Nototrypa orbicularis*, *Dicranum viride*, *Mannia triandra* a *Buxbaumia viridis*). Cílené vyhledávání vzácných či ohrožených druhů je však v ČR v pozadí a má slabou metodickou i informatickou podporu.

Cílem této metodiky je poskytnout postup, který metodicky i výpočetně podpoří bryologický a lichenologický průzkum na území ČR. Tato metodika pomůže průzkum zacílit a tím i zefektivnit, přináší reprodukovatelný a dobře dokumentovaný postup, jak s pomocí nástrojů modelování druhové distribuce naplánovat návštěvu lokalit s vysokou pravděpodobností výskytu vhodného habitatu zájmového druhu. Představujeme zde uživatelsky nenáročný rozhraní pro výpočet a validaci predikčních modelů a tvorbu mapy pravděpodobnosti výskytu. Navrhujeme také metodický postup, jak s výslednou predikční mapou pracovat, jak ji kriticky zhodnotit a na co si při jejím používání dát pozor. Metodické kroky jsou zpracované jako detailní návod popisující jednotlivé kroky včetně konkrétních příkladů využití volně dostupného software a aplikací pro chytrá zařízení. To vše s cílem poskytnout co nejúplnější podporu přípravě podkladů pro terénní průzkum, od prvotní koncepční úvahy nad tím, pro které druhy je tato metodika použitelná, až po příklady, jak zpracovat a nahrát výsledná data do terénního zařízení.

Cílem této metodiky není popřít či zpochybnit expertní přístup k bryologickému a lichenologickému průzkumu. Bryolog, znalec daného území bude pravděpodobně schopen velice efektivně a přesně připravit podklady pro terénní průzkum zacílený na konkrétní vzácné druhy pouze na základě své zkušenosti. To je pro praktickou ochranu přírody i pro vědu velice cenné. Předkládaná metodika je doplňkem tohoto expertního přístupu a pomocníkem v případech, kdy daný odborník není k dispozici, ať už z důvodů personálních, kapacitních či finančních. Tato metodika cílí zejména na pracovníky orgánů ochrany přírody, ať už na státní či regionální úrovni, kteří plánují či realizují floristické (bryologické a lichenologické) průzkumy.

3 Teoretický úvod

3.1 Znalost výskytu mechorostů a lišejníků

Mechorosty a lišejníky jsou většinou v závěsu za cévnatými rostlinami, co se týká jejich výzkumu a ochrany. To se tradičně přisuzuje menšímu počtu specialistů a obtížné determinaci často vyžadující mikroskop a preparace morfologických determinačních struktur, u lišejníků dokonce s využitím metod chemotaxonomie, například v podobě tenkovrstvé či plynové chromatografie. Relativně menší znalost rozšíření vzácných mechorostů a lišejníků může přitom působit problémy v praktické ochraně přírody a způsobovat nečekané kolize zájmů ochrany přírody se zájmy využití území člověkem. Dobrá znalost rozšíření vzácných druhů hraje důležitou roli při nastavování právního rámce jejich ochrany a ochrany biotopů, kde se vyskytují, a to jak na Evropské, tak národní úrovni.

V této metodice je nezbytné se zabývat také formou a strukturou dat, které reprezentují znalost rozšíření druhů. V obecné rovině je významná expertní znalost reprezentovaná bryologicko-lichenologickou komunitou následovaná literárními a herbářovými doklady. Tyto tradiční zdroje jsou přes svou relativně vysokou spolehlivost jen velmi omezeně využitelné pro strojové zpracování, dnes již samozřejmé takřka ve všech biologických oborech. Pro strojové zpracování jsou zřizovány databáze výskytů různých skupin organismů a to na úrovni světové, např. GBIF (<https://www.gbif.org>), ale i lokální, např. Nálezová databáze ochrany přírody NDOP (<https://portal.nature.cz/nd>), Pladias (<https://pladias.cz>), nebo ČNFD (<https://botzool.cz/vegsci/phytosociologicalDb/>). Na jedné straně to s dostupností dat o výskytu mechorostů a lišejníků v ČR ve formě počítačových databází není vůbec špatné. V roce 2020 bylo v NDOP k nalezení asi 350 000 záznamů mechorostů a 80 000 lišejníků, v ČNFD pak asi 150 000 záznamů mechorostů a lišejníků. Desetitisíce záznamů obsahují databáze herbářů, např. CBFS nebo BRNM. Na straně druhé, v porovnání s 13 miliony záznamů pro cévnaté rostliny v databázi Pladias stále ještě mechorosty a lišejníky zaostávají.

Kromě faktu, že jsou data o výskytu elektronicky uložena ve strojově čitelné formě, je také důležitá jejich standardizace. Jak bylo zmíněno výše, jen v ČR existuje několik nezávislých institucionálních databází shromažďující digitalizovaná data o výskytu mechorostů a lišejníků. Každá má jiný taxonomický standard, rozsah, formát a systém ukládání dat, a ve většině případů nejsou uzpůsobené pro přímé poskytování dat pro strojové zpracování v reálném čase. Proto vznikla v roce 2018 databáze DaLiBor (Database of Lichens and Bryophytes) <https://dalibor.ibot.cas.cz>, která poskytuje sjednocující standard pro správu výskytových dat mechorostů a lišejníků na národní úrovni. DaLiBor je přístupná pro registrované uživatele (žádost o registraci se odesílá e-mailem správci databáze). Databáze poskytuje jednotné prostředí pro správu a ukládání dat, jejich validaci a další strojové zpracování. DaLiBor je vybaven efektivními taxonomickými převodníky vystavenými experty Janem Kučerou a Jiřím Malíčkem. To umožňuje automatizovaný převod taxonů z ostatních databází. DaLiBor poskytuje také systém kontrol, jako například lokalizace záznamu v ČR či validity nomenklatury a standardizaci formátu vstupních dat, jako například formát zeměpisných souřadnic, formát data nálezu, standardizované kategorie substrátů v místech výskytu.

DaLiBor nejen sdružil existující databáze, ale také obsahuje vlastní excerpty literatury či některé osobní nálezové databáze bryologů a lichenologů. V roce 2020 poskytuje DaLiBor asi půl milionu záznamů výskytů mechorostů a lišejníků ve standardizované podobě, spravovaných pod licencí Creative Commons. Tabulka 1.

Tabulka 1 Počet záznamů mechorostů a lišejníků v databázi DaLiBor v roce 2020 podle původu, odkud byla data do databáze importována, a zkratky používané pro identifikaci datového zdroje.

Počet záznamů	Zdroj dat	Zkratka
387817	Nálezová databáze ochrany přírody	NDOP
149855	Česká národní fytoocenologická databáze	ČNFD
16939	Excerpce literatury – lišejníky	Excerpce (L)
10950	Herbář Jihočeské university	CBFS
6902	Excerpce Bryonora – lišejníky	Bryonora (L)
4052	Veřejné herbáře – lišejníky	Herbáře (L)
3890	Herbář & databáze J. Malíčka	Malíček
2649	Inventarizační průzkumy – lišejníky	Inventarizáky (L)
881	Excerpce Bryonora – mechorosty	Bryonora (M)
359	Osobní nálezové database	Osobní nálezy

Jádro databáze DaLiBor tvoří standardní relační databáze a základní jednotkou je záznam jednoho výskytu v čase a prostoru. Minimální záznam obsahuje pouze vědecké jméno druhu, lokalizaci, datum nálezu, autora a zdroj dat. Lokalizace je ukládána bodově jako zeměpisná délka a šířka v souřadnicovém systému WGS84. Na základní jednotku záznamu se váže bohatá struktura dalších nepovinných polí, která umožní záznam obohatit o údaje k lokalitě (název, přesnost lokalizace, nadmořská výška) nebo sběru (herbářový doklad, determinace), ale i pro bezcévné rostliny specifických údajů, jako jsou například typy substrátů či chemická data. Kde je to možné, jsou pole databáze řešena editovatelnými číselníky, které omezují duplicity a udržují jednotnost údajů.

Data vstupují do databáze buď expertním způsobem administrátora databáze přes standardní rutiny SQL, nebo, což je častější a preferovaný způsob, přes importní soubor formátu MS Excel. Podmínkou je samozřejmě jednotný formát vstupního souboru. V průběhu importu dochází ke kontrole polí proti číselníkům, zejména vědeckého jména druhu, autora a případně zdroje. V případě přítomnosti dalších indikátorů lokality (obce, okresy, síťové mapování) je možné provést křížovou kontrolu oproti souřadnicím.

K databázi je možné přistupovat přes rozsáhlé grafické rozhraní (GUI) pro editaci i validaci záznamů. Databáze vyžaduje přihlášení konkrétního uživatele s jeho přihlašovacím jménem a heslem. Proto jsou veškeré editační kroky ukládány v historii, a je tak možné dohledat změnu a jejího autora, stejně jako filtrovat záznamy podle historie editace nebo editora. Důležitým prvkem je rozhraní pro validaci záznamu, které umožňuje na souběžném zobrazení detailu záznamu a mapy označit záznam jako správný, chybný nebo nejistý. Součástí validačního procesu jsou i komentáře, které lze vkládat k jednotlivým záznamům, a na které jsou upozorněni správci databáze, případně správci jednotlivých druhů. Tento validační nástroj přímo navazuje na možnost tvorby mapy výskytu druhu. Tu lze jednak vizualizovat přímo v grafickém rozhraní databáze, ale také provést exporty podkladů pro úpravu map v externím software (např. QGIS nebo ArcGIS) nebo přímo prohlížet v geografickém informačním software jako virtuální mapové vrstvy.

DALIBOR
Mapy » AdminMapy » Ke stažení

Přejít na mapu taxonu...

Přejít na záznam ID...

matej.man@ibot.cz

+

Pro změnu souřadnic klikněte do mapy

(zpět na výpis mapovacího pole)

Testujeme, co nahrajete tak časem smázneme...)

+

Floristický záznam

Taxon:

Acarospora nitrophila

Substrát:

siliceous rock

Chemická data:

nevyplněno - doubleclick pro editaci

Hlavní kategorie substrát:

savikolní

Uprušující kategorie substrát:

silikát

Originální jméno:

nevyplněno - doubleclick pro editaci

Ze souřadnic odvozená obec:

Lokalita:

Mohelno, Mohelenská hadcová step National Nature Reserve

Poznámka:

nevyplněno - doubleclick pro editaci

Nadm. výška (min):

275

Nadm. výška (max):

375

Datum:

2011-03-31

uvádějte ve formátu (YYYY, YYYY-MM, YYYY-MM-DD)

Nálezce(i):

Halda, Josef

-- přidat dalšího --

+

Doplňkové informace

Příznaky pro mapu

Záznam do mapy:

NE

Záznam herb

NE

Kvalita:

Původní výskyt:

neurčeno

Revize záznamu:

Nerevidovaný záznam

Zdroje

Pramen:

nevyplněno - doubleclick pro editaci

Herbář(e):

-- přidat další --

Mapové čtverce:

Technické info

ID:

dávka: 162

záznam: 500000104

nahrál: Malíček Jiří

2018-11-30 09:53:09.448

Fytochor. doporučen:

false

Nadm. výška aprox.:

false

Kvadrant doporučen:

true

Původ:

implicitní projekt | Botanický ústav AV ČR, v. v. i.

Poznámky z cizích zdrojů

Environment:

Poznámka k excerpci:

Poznámka - pochyby:

Poznámka - ostatní:

Determinoval:

+

Lokalizace nálezů

	původní lokace	nová lokace
GPS	16.183333333333334, 49.10805555555556	
DMS	49°06'29"N, 16°11'00"E	
Zdroj souřadnic:	odečteno z mapy	
Polohová přesnost[m]:	500	
Ze souřadnic odvozená obec:		
Okres:	Třebíč	
Mapové čtverce:		
Kvadranty:	6863c	

3.2 Ochrana mechorostů a lišejníků

Přesto, že můžeme od roku 2010 na národní úrovni pozorovat konzistentní snahu o zařazení mechorostů a lišejníků do seznamu chráněných druhů ve vyhlášce 395/1992 Sb. (Horydská et al. 2011) ani po deseti letech se to stále nepodařilo. Jediná legislativní ochrana se vztahuje na vybraných několik druhů mechorostů a vyplývá z Evropské směrnice o stanovištích (92/43/EHS), respektive z její přílohy II, která stanovuje druhy vyžadující zvláštní územní ochranu, která je realizovaná v podobě tzv. evropsky významných lokalit (EVL). Z území České republiky je známo celkem osm druhů mechorostů uvedených v příloze II. Recentní výskyt je potvrzen pouze u šesti z nich (*Buxbaumia viridis*, *Dicranum viride*, *Hamatocaulis vernicosus*, *Mannia triandra*, *Notothylas orbicularis* *Orthotrichum rogeri*). Zbýlé dva druhy – játrovka *Scapania carinthiaca* var. *massalongia* a mech *Meesia longiseta* – jsou v současné době řazeny mezi regionálně vyhynulé (Holá & Turoňová 2014). Lišejníky nepožívají legislativní ochrany ani na Evropské ani na národní úrovni. Možná je změna národní legislativy ohledně lišejníků a mechorostů tak zdlouhavá právě proto, že není dostatečně známé jejich rozšíření na našem území.

lišejníků, a to i mimo tradiční regiony jejich výskytu. To může na straně biologické souviset s populační dynamikou druhů, a na straně metodické s obtížnou terénní determinací nebo se zacílením terénních průzkumů na lokality, u nichž se na základě expertního odhadu předpokládá výskyt „zajímavých druhů“, nemluvě o skvělé prozkoumanosti chráněných území v ČR. Problematickou situací může být nečekaný nález vzácných druhů mechorostů nebo lišejníků například při biologických průzkumech v rámci hodnocení vlivu staveb na životní prostředí. To může přirozeně nastat pro jakékoliv ohrožené organismy, nicméně u těch s menší znalostí rozšíření je potenciální kolize zájmů investora a ochrany přírody pravděpodobnější.

Díky dobré znalosti rozšíření vzácných a chráněných druhů mechorostů a lišejníků můžeme na jedné straně minimalizovat kolize s investičními zájmy, ale především na straně druhé můžeme efektivně chránit a spravovat lokality, na kterých se druhy vyskytují. Management lokalit s výskytem vzácných druhů a sledování jejich populací je v ČR dobře metodicky podchycen a na rozdíl od vyhledávání nových lokalit se mu věnují alespoň nějaké lidské zdroje a finance, a to zejména prostřednictvím Agentury ochrany přírody a krajiny ČR (AOPK). Za všechny můžeme jmenovat například kosení či zásahy do půdního krytu na rašeliništích v případě druhu *Hamatocaulis vernicosus* (Štechová & Kučera 2007) či ponechávání tlejícího dřeva v rámci péče o lokality s *Buxbaumia viridis* nebo například disturbance lokalit na Božím daru ve prospěch populace lišejníku *Baeomyces placophyllus*. Nejen v případě těchto, ale i mnoha dalších ohrožených druhů mechorostů a lišejníků nepochybně existují v ČR lokality výskytu, o kterých nevíme a kterým by aktivní péče mohla prospět, nebo je dokonce zachránit před zánikem. Metodika, kterou právě čtete, by měla pomoci zefektivnit vyhledávání lokalit s výskytem vzácných druhů, kde by mohlo případně dojít k ochrannému zásahu.

3.3 Vzácné, nebo jen málo probádané?

Formalizovaný přístup k vyhledávání nových lokalit vzácných druhů, který tato metodika přináší, může pomoci nejen k případnému zahájení managementu na dosud neznámých lokalitách, může ale také pomoci k poznání reálné distribuce daného druhu. Vzácné druhy daného území jsou přirozeně ty, které jsou jen málo rozšířené, typicky jsou vázané na málo rozšířený substrát nebo ojedinělou kombinaci environmentálních podmínek či specifický historický vývoj lokality. Druh může být také považován za vzácný proto, že není jeho rozšíření dostatečně známé, ať už kvůli nedostatečné bryologické a lichenologické prozkoumanosti území (příkladem mohou být opět druhy *Hamatocaulis vernicosus*, *Buxbaumia viridis* či *Baeomyces placophyllus*), nebo kvůli pomíjivému biotopu, na kterém se vyskytuje.

V první dekádě 21. století se soustředila značná pozornost na průzkum lokalit zacílený na výše zmíněné druhy evropsky významných lokalit. V několika případech, například *Buxbaumia viridis*, *Hamatocaulis vernicosus* a *Dicranum viride* (Zmrhalová & Holá 2010; Jandová & Holá 2012; Holá & Turoňová 2014) se ukázalo, že jsou druhy rozšířenější, než se původně předpokládalo, protože byly nalezeny na mnoha nových lokalitách. Z toho je cítit jistý potenciál pro cílené vyhledávání nových lokalit výskytu nejen těchto, ale i dalších druhů. V tom může být k užitku tato metodika.

3.4 Vyhledávání vzácných druhů v ČR

Tradiční metodou vyhledávání a mapování vzácných mechorostů a lišejníků je terénní průzkum. V oblasti mapování a sledování vývoje stávajících populací jsou k dispozici zavedené metodiky spravované zejména zadavateli průzkumů jako je AOPK či správy chráněných území.

Obecně pro mechorosty například:

- *Obecná metodika monitoringu a mapování mechorostů*, Lucie Zemanová, verze 8.2018, AOPK ČR
- *Metodika monitoringu a mapování ohrožených druhů lišejníků*, Lucie Zemanová, verze 8.2018, AOPK ČR
- *Příprava a zavedení sledování stavu předmětů ochrany evropsky významných lokalit z roku 2016*, která obsahuje část věnovanou pěti druhům mechorostů a stojí za ní autorky Eva Holá a Táňa Štechová.
- *Monitoring ohrožených rašeliništních mechorostů a péče o jejich lokality*, Štechová T., Holá E., Ekrťová E., Manukjanová A. & Kučera J. z roku 2014, AOPK ČR
- *Metodika inventarizačního průzkumu: Mechorosty*, Eva Holá, Táňa Štechová, Eliška Vicharová, Ondřej Popelka, verze 2019, AOPK ČR
- *Metodika druhové ochrany bezcévných rostlin*, Soubor metodik ochrany a monitoringu pro různé habitaty, Táňa Štechová, Zdeněk Soldán, Eliška Vicharová, Julie Jandová, David Svoboda, výsledek projektu TAČR Beta: TB050MZP005, 2016

Obecně pro lišejníky například:

- *Metodika inventarizačních průzkumů – lišejníky*, Ondřej Peksa, Lucie Zemanová, David Svoboda verze 2019, AOPK ČR
- *Metodika monitoringu a mapování ohrožených druhů lišejníků*, Lucie Zemanová, verze 8.2018, AOPK ČR
- *Metodika monitoringu pro terikolní makrolišejníky*, AOPK ČR

Pro vybrané zájmové druhy pak existují samostatné metodiky s podrobně rozpracovaným postupem monitoringu v terénu, metodou sledování populační dynamiky a popisem úkonů a parametrů, které je na lokalitách třeba vykonat či zaznamenat. Zde se jedná zejména, avšak nejen, o druhy evropsky významných lokalit, například

- *Srpnatka fermežová (*Hamatocaulis vernicosus*)*, metodika monitoringu, Renata Linhartová, Táňa Štechová, editor: Eva Holá, verze 2011
- *Dvouhrotec zelený (*Dicranum viride*)*, metodika monitoringu, Eva Holá, Julie Jandová, Renata Linhartová, editor: Eva Holá, verze 2011
- *Návrh metodiky monitoringu pro mech *Neckera pennata* Hedw. [Bryophyta, Hypnales]*, Zbyněk Hradílek, verze 1.2006, AOPK ČR
- *Podrobná metodika monitoringu druhu *Cladonia stellaris* (sect. *Cladina*)*, Jana Kocourková, 2006, AOPK ČR
- *Metodika monitoringu lišejníku *Fulgensia fulgens**, Jana Kocourková, 2006 AOPK ČR
- *Návrh metodiky monitoringu pro lišejník *Alectoria sarmentosa* (Ach.) Ach. [Lecanorales, Alectoriaceae]*, Jaroslav Šoun, Jan Vondrák, 2006, AOPK ČR

V oblasti vyhledávání nových lokalit cílových druhů neexistuje na našem území takřka žádná metodická podpora na rozdíl od výše popsané bohaté metodické podpory pro monitoring. Možná i

proto, že je vyhledávání nových lokalit vzácných mechorostů a lišejníků na první pohled těžko metodicky popsatelné, klade se v této oblasti tradičně důraz na expertní znalost. Což je zpravidla efektivní, může to však narazit na personální limity. Například neobsazenost pozic či neochota najímat externí bryology a lichenology v institucích, které zaštiťují monitoring a správu lokalit vzácných druhů, jako je AOPK či správy chráněných oblastí. Často tak průzkum probíhá v okolí známých lokalit nebo ověřováním lokalit se známým historickým výskytem. To může vést k efektivnímu nalézání nových lokalit, avšak pouze na lokální úrovni. K objevení nových lokalit vzácných druhů dochází tak často náhodně, při exkurzích bryologické a lichenologické obce, či jako vedlejší produkt ekologických či populačních výzkumů mechorostů a lišejníků.

3.5 Zahraniční zkušenosti se zacíleným průzkumem

Východisko, jak formalizovat přístup k vyhledávání nových lokalit vzácných druhů mechorostů a lišejníků, mohou představovat Geoinformační systémy (GIS), metody dálkového průzkumu Země (DPZ) či metody predikce výskytu vhodných habitatů (SDM).

Geoinformační systémy jsou v ochraně přírody využívány hojně, a to pravděpodobně i k vytipování potenciálních lokalit výskytu vzácných druhů, protože orgány státní správy jako je AOPK či správy chráněných oblastí mají přístup k environmentálním datům pro potřeby geoinformačních analýz. Tyto instituce zpravidla mají také GIS odborníky, kteří mohou mít přístup například k digitálnímu modelu terénu a reliéfu poskytovanému Českým úřadem zeměměřickým a katastrálním či k meteorologickým datům poskytovaným Českým hydrometeorologickým institutem a mohou je tak efektivně využívat jako podklad pro plánování terénních průzkumů. Mnohé geoinformační vrstvy pak AOPK i správy chráněných oblastí sami generují, rozvíjejí a spravují. Například vrstvy z mapování biotopů či konsolidovaná vrstva ekosystémů, obě informačně velice hodnotné pod správou AOPK. Dále například lokální produkty bezpilotního průzkumu, často pořizované správami národních parků.

Zahraniční zkušenosti nám ukazují, že mohou být k vyhledávání cílových druhů použitelná i data z bezpilotních systémů osazených multi- či hyper-spektrálními senzory. To jsou však metody aplikovatelné pouze v bezlesí, ideálně i na plochách s řídkou vegetací. V našich podmínkách například pro vrcholové partie horských oblastí či post průmyslové oblasti v raných fázích sukcese jako haldy či opuštěné lomy. Na případové studii z roku 2016 na písčných dunách ostrova Sylt v severním Německu můžeme vidět úspěšnou aplikaci takového přístupu při vyhledávání a mapování lokalit s výskytem invazivního mechu *Campylopus introflexu* (Skowronek et al. 2017).

S rozvojem výpočetní kapacity a přibývajícím množstvím dostupných environmentálních dat se rozvinul celý obor zabývající se modelováním výskytu organismů. S jistým zpožděním oproti cévnatým rostlinám se začaly tyto metody aplikovat po roce 2000 i u mechorostů a lišejníků. Nejprve dosti naivně s využitím poměrně hrubých podkladových dat na škále kilometrů čtverečních např. (Sérgio et al. 2007), v současné době již s využitím sofistikovaných predikčních a validačních nástrojů a environmentálních dat s vysokým rozlišením v řádu desítek metrů.

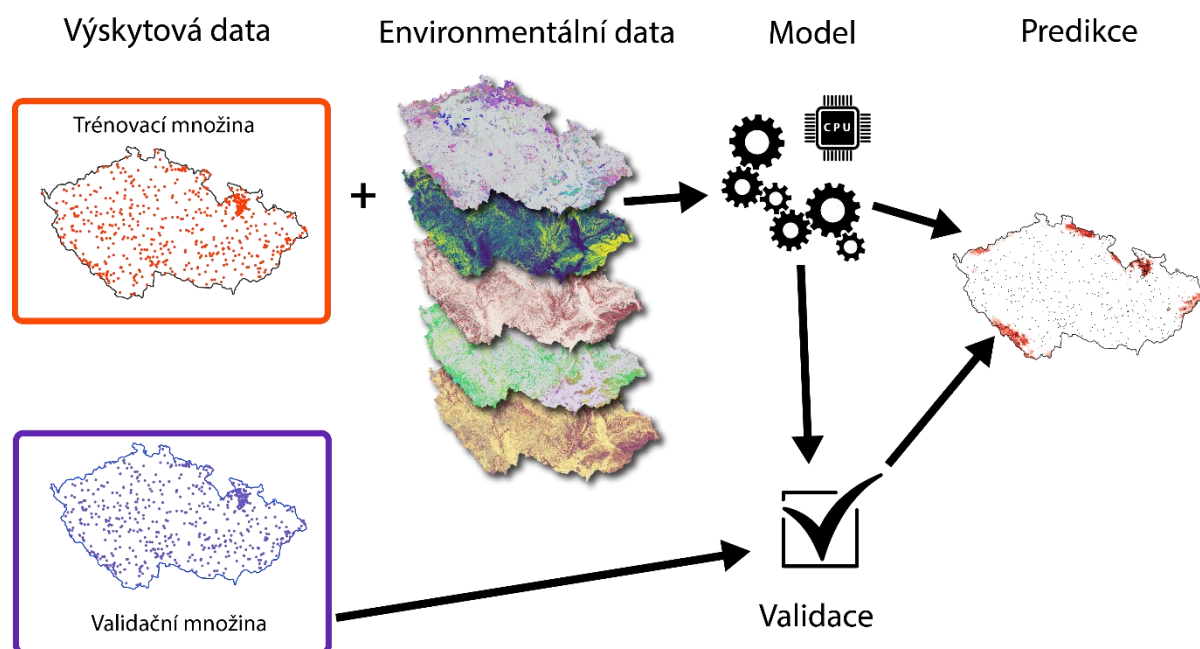
Zde předkládaná metodika navazuje na úspěchy využití SDM při vyhledávání nových lokalit a zacílení terénních průzkumů, například (Callaghan & Ashton 2008). V této studii na regionální úrovni v měřítku 2x2 km predikovali diverzitu mechorostů a porovnali se známou diverzitou. Z toho odvodili čtverce, kde je vysoce pravděpodobně podceněn bryofloristický průzkum, a na které by bylo dobré se zaměřit při dalších terénních průzkumech. Využití SDM v případové studii z Virginie z roku 2005 pro zacílení terénního průzkumu při vyhledávání nových lokalit vzácného druhu *Xerophyllum asphodeloides* přineslo také své ovoce v podobě nalezení dosud neznámých lokalit, odkud byl druh následně potvrzen (Bourg et al. 2005). Podobně úspěšní byli také kolegové ze Švýcarska, kteří objevili dosud neznámé lokality vzácného druhu *Eryngium alpinum* díky zacílení terénního výzkumu pomocí

SDM (Guisan et al. 2006). Nejblíže je naše metodika svým přístupem a zejména zacílením na mechorosty nejspíš případové studii z Tyrolských Alp, kde v roce 2015 autoři využili predikčních modelů a geoinformačních analýz k nalezení nových lokalit vzácného druhu mechu *Buxbaumia viridis* (Spitale & Mair 2015).

3.6 Modelování potenciální vhodnosti habitatu

Species distribution modelling (SDM) v širším slova smyslu můžeme česky opsat jako modelování výskytu druhů, predikční modelování výskytu druhů, modelování ekologické niky druhů či modelování potenciální vhodnosti habitatu pro výskyt daného druhu. V rámci této metodiky je nejvhodnější použití názvu „modelování potenciální vhodnosti habitatu pro výskyt daného druhu“ (Guisan & Zimmermann 2000) což asi nejlépe vystihuje používaný přístup. Na druhou stranu je to spojení dosti krkolomné a dlouhé, proto se bude často záměnou za tento přesný název objevovat dále převážně zkratka SDM.

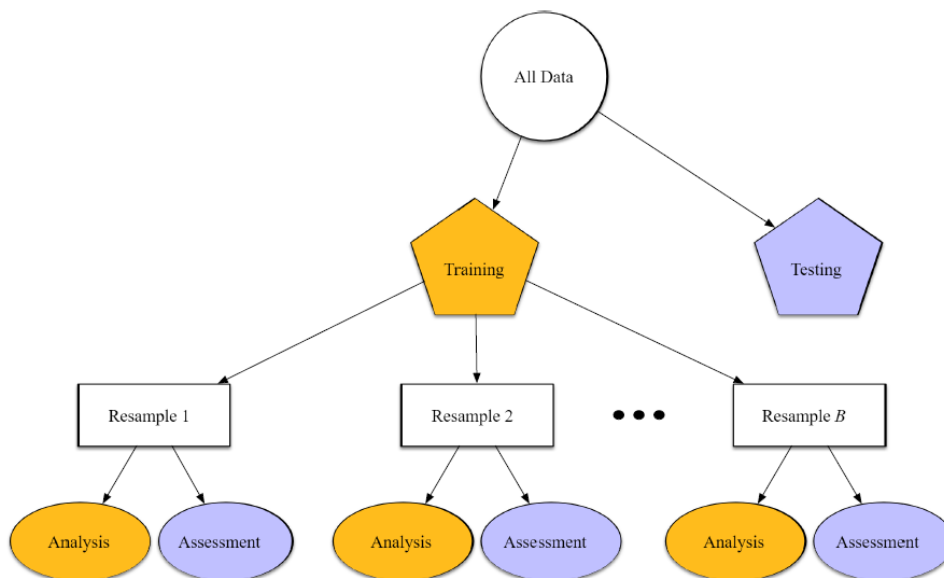
Základním předpokladem pro modelování druhové distribuce je dostupnost dat o výskytu organismu



Obrázek 2 Schéma postupu při modelování potenciální vhodnosti habitatu pro výskyt cílového druhu. Trénovací a validační množina jsou odlišné, nezávislé datasety.

(prezence a absence či pseudo-absence), environmentální data a robustní predikční algoritmus běžící na výpočetním stroji (server, PC). Data jsou rozdělena na trénovací a validační množinu. Trénovací množina výskytových dat je využita dále k parametrizaci modelu na základě environmentálních podmínek a jejich interakcí v místech známého výskytu daných druhů. Přesnost predikce výsledného modelu je pak hodnocena na validační množině výskytů, kde se porovnává pravděpodobnost výskytu předpovězená modelem se skutečností (obrázek 3). Většina algoritmů pro SDM, zejména pak metody strojového učení provádí další dělení trénovacích dat, při kterém vyčleňuje testovací podsoubor, použitý pro validaci konkrétního běhu modelu. Toto dělení trénovacího souboru a fit predikčního modelu se zpravidla provádí několikrát tzv. násobná křížová validace pro stabilizaci výsledku (obrázek 4). Teprve poté se provádí testování výsledného modelu na validační množině dat, která nevstupovala do žádné fáze trénování modelu.

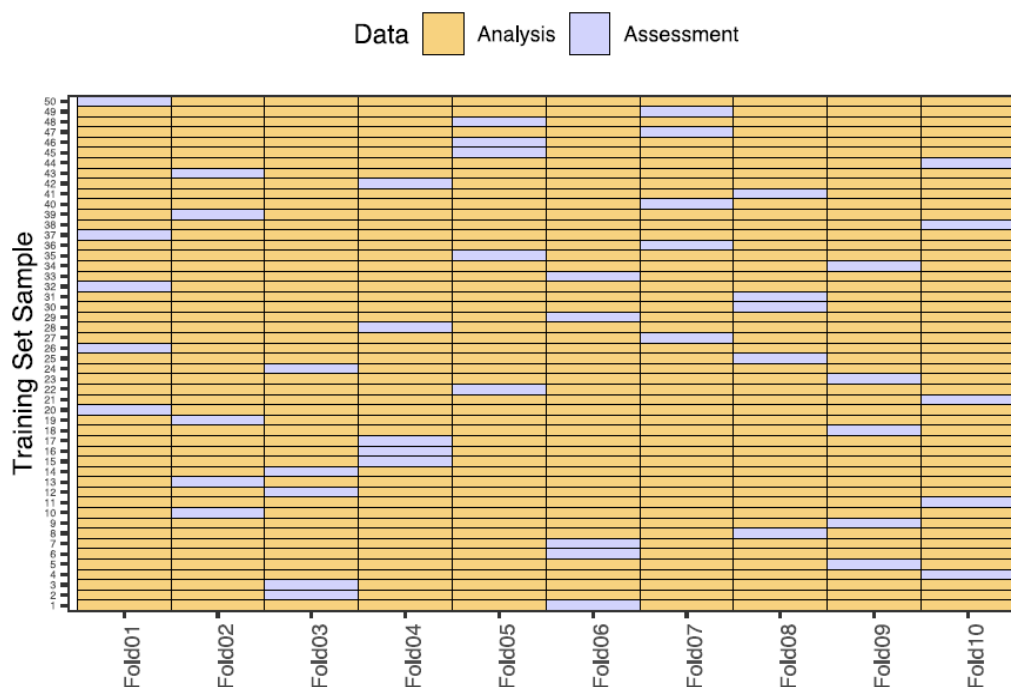
Pro popis přesnosti výsledných modelů se používají různé metriky, podle toho, jaký typ chyby nás zajímá. Na základě validace je proveden výběr nejúspěšnějšího modelu a ten je použit k výsledné tvorbě predikční mapy. Celý tento stručně popsany proces je graficky znázorněn na Obrázku 3 a je dále rozveden podrobněji v následujících kapitolách.



Obrázek 3 Schéma rozdělení dat na trénovací a testovací dataset (pětiúhelníky) a další dělení trénovacího datasetu na trénovací a validační podsoubory pro násobnou křížovou validaci. Zdroj: https://github.com/rstudio-conf-2020/applied-ml/blob/master/Part_4.pdf

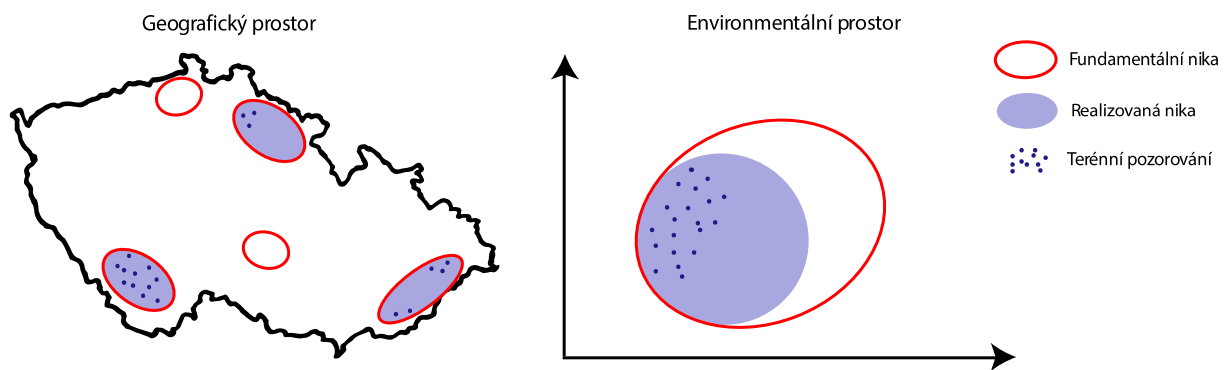
3.6.1 Koncept, předpoklady, potenciální limity

Tradičně je výskyt druhů popisován ve dvojrozměrném, geografickém prostoru reprezentovaném pomocí map či souřadnic. Pro modelování habitatu potenciálně vhodného pro výskyt druhů musíme však kromě geografického prostoru vzít v úvahu také prostor environmentální, reprezentovaný proměnnými, na které organismy přímo či nepřímo reagují. Myšlenka environmentálního prostoru, ve kterém se druhy vyskytují a na který reagují, má oporu v teorii ekologické niky (Obrázek 5). Koncept ekologické niky prošel jistým vývojem a je různými autory modifikován či rozšiřován (Polechová & Storch 2018). SDM vychází z konceptu ekologické niky, který popsal v 60 letech dvacátého století britský ekolog G. E. Hutchinson (Hutchinson 1957). Ten definuje *fundamentální ekologickou niku* druhu jako soubor environmentálních podmínek, ve kterých může druh dlouhodobě přežívat. Fundamentální niku si můžeme představit jako n-rozměrný prostor, ve kterém je každý bod reprezentován souborem environmentálních podmínek, které umožňují přežívání daného druhu. Kromě fundamentální niky operují ekologové často s označením realizovaná environmentální nika, což je environmentální prostor, ve kterém se v dané chvíli daný druh skutečně vyskytuje. Realizovaná nika je podmnožinou niky fundamentální.



Obrázek 4 Schéma desetinásobné křížové validace při tvorbě predikčních modelů. Výskytová data jsou desetkrát (fold) rozdělena na trénovací podsoubor a soubor pro křížovou validaci. Model se počítá a validuje desetkrát. Zde máme 50 pozorování, kdy se vždy data náhodně rozdělí. Zdroj: https://github.com/rstudio-conf-2020/applied-ml/blob/master/Part_4.pdf

Zjednodušeně řečeno, modely druhové distribuce předpokládají, že je geografické rozšíření daného organismu mimo jiné determinováno environmentálními podmínkami a že se organismus může teoreticky vyskytovat všude tam, kde jsou pro jeho výskyt vhodné environmentální podmínky. Ačkoliv koncept ekologické niky operuje s biotickými interakcemi jako je kompetice či parazitismus, které mohou výrazně ovlivňovat rozšíření organismů, přímo do modelů druhové distribuce jsou biologické interakce zahrnuty jen vzácně z důvodu nedostatku či nedostupnosti podkladových dat. Nepřímo se biotické interakce, případně limity šíření druhu projevují v pozorovaném výskytu. Ten ovšem kromě všech těchto faktorů zahrnuje i míru prozkoumanosti území a tyto faktory od sebe nelze jednoduše oddělit. Distribuce druhů ze skupiny mechorostů a obzvláště pak lišejníků může být do značné míry určována právě kompeticí, a to zejména s cévnatými rostlinami, ale i kompeticí v rámci skupiny. Na druhou stranu u většiny druhů mechorostů a lišejníků se předpokládají velké areály rozšíření. Ty jsou formovány velkou schopností propagulí mechorostů a lišejníků šířit se na velké vzdálenosti (Barbé et al. 2016; Patiño & Vanderpoorten 2018), a je tedy možné předpokládat, že mají šanci obsadit jakýkoliv geografický prostor, a proto by se bez biologických interakcí jejich realizovaná nika mohla blížit té fundamentální. V případě snadného šíření na velké vzdálenosti jsou druhy spíše limitované environmentálními podmínkami, což je situace obzvláště výhodná pro využití SDM.



Obrázek 5 Ilustrace výskytu organismu v geografickém a environmentálním prostoru.

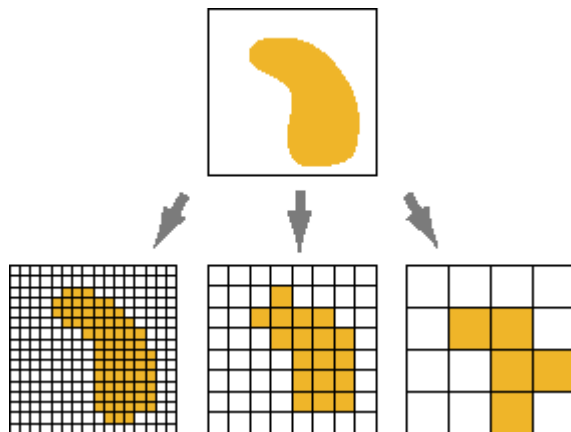
Při SDM je třeba mít na paměti několik věcí, vyvozených z konceptu ekologické niky, které nám pomohou uchopit a interpretovat výsledné predikce (mapy).

- 1.) Mezi skutečnou distribucí daného druhu a jejich známými nálezy z terénu je zpravidla velký rozdíl. Dostupná pozorování (v Obrázku 5 body) velmi pravděpodobně nepokrývají celou realizovanou niku (V Obrázku 5 plná plocha). To může mít významný vliv na kvalitu predikce, zejména pokud je pozorovaných výskytů velmi málo, nebo pokud pochází většina pozorování z malého území oproti geografickému pokrytí výsledných predikcí (map). Podrobněji v kapitole 3.6.3.
- 2.) Biotické interakce, historie a limity šíření některých druhů mohou mít vliv na jejich realizovanou niku, která nemusí odpovídat jejich nice fundamentální. Pokud pro charakteristiku niky daného druhu používáme SDM, jedná se vždy o predikci na základě neúplné reprezentace kombinace environmentálních faktorů, které představují minimální kombinaci podmínek pro přežití daného druhu na daném místě. Výsledné mapy distribuce habitatů potenciálně vhodných pro výskyt daného druhu se tak mohou lišit od skutečné distribuce, neboť se zakládají na modelu, který se liší od realizované niky (viz bod 1.) i od fundamentální niky.
- 3.) Environmentální data používaná při SDM jen stěží vystihují všechny environmentální podmínky, které skutečně determinují fundamentální niku daného druhu. V případě, kdy nepostihují dostupná environmentální data hlavní faktory determinující rozšíření druhů, pak mohou být výsledné predikce (mapy) zavádějící. Podrobněji v kapitole 3.6.2.

3.6.2 Environmentální data

Pro praktické modelování výskytu vhodného habitatu je důležité popsat mnohorozměrný environmentální prostor obývaný daným druhem formálně tak, aby byl tento popis strojově dobře čitelný pro daný software či algoritmus. Zpravidla ve formě digitální mapy, kde je každá souřadnice či každý čtverec reprezentovaný konkrétní hodnotou environmentálního parametru. Například průměrnou roční teplotou, sklonem svahu, typem lesa či krajinného pokryvu. Z technického hlediska to bývají většinou rastry, tedy geografické informační vrstvy reprezentované čtvercovou sítí, která má definovaný souřadnicový systém, počátek (zpravidla souřadnice levého horního rohu), počet sloupců, počet řádku, počet čtverců a velikost čtverce. Každý čtverec (= pixel, buňka) nese hodnotu daného environmentálního parametru jako například nadmořská výška, průměrný úhrn srážek, průměrná roční teplota, typ krajiny.

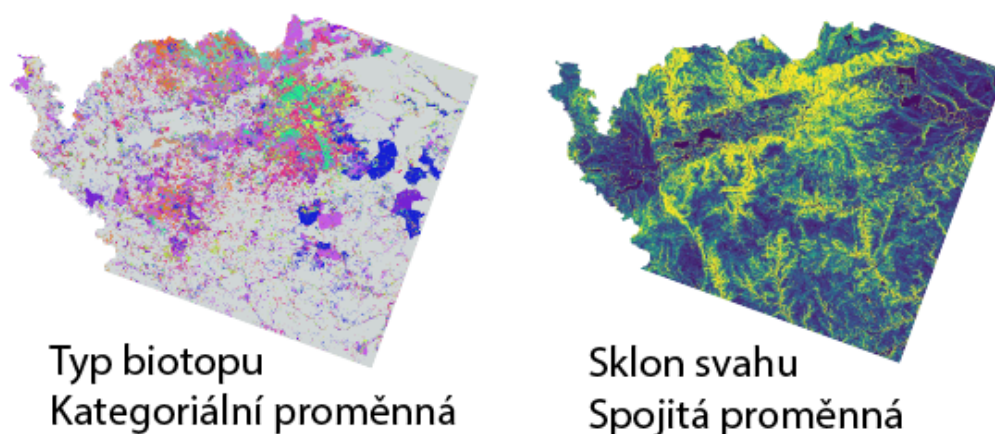
Velikosti buňky rastru environmentálních proměnných hraje velice důležitou roli při modelování potenciálního výskytu vhodného habitatu. Spolu s přesností lokalizace výskytových dat totiž určuje prostorovou škálu celého modelovacího procesu. Nelze jednoznačně říct, jaká škála je pro SDM optimální. Vždy to souvisí s účelem modelování a dostupností dat. V začátcích SDM se velikost buněk podkladových rastrů pohybovala v řádku kilometrů i při modelování na regionální úrovni, protože podrobnější data buď nebyla dostupná, nebo by jejich použití bylo výpočetně velice náročné. Díky satelitnímu snímkování jsou dnes dostupná některá environmentální data s rozlišením desítek metrů pro celý svět, dokonce jsou dostupná i některá data založená na interpolaci terénních pozorování pro celý svět v rozlišení stovek metrů, například půdní data (Hengl et al. 2017) či klimatická data (Karger et al. 2017; Fick & Hijmans 2017). Na regionální úrovni se dnes často používají rastry s velikostí buněk řádově několik metrů i méně, což je užitečné například v případě, kdy organismy obývají malé území, determinované lokálními mikroklimatickými podmínkami.



Obrázek 6 Rastrová reprezentace objektu v různém rozlišení (s různou velikostí buňky). Je zřejmé, že s větší velikostí buňky dochází ke generalizaci daného jevu, ale zároveň to vede k menší výpočetní náročnosti predikcí.

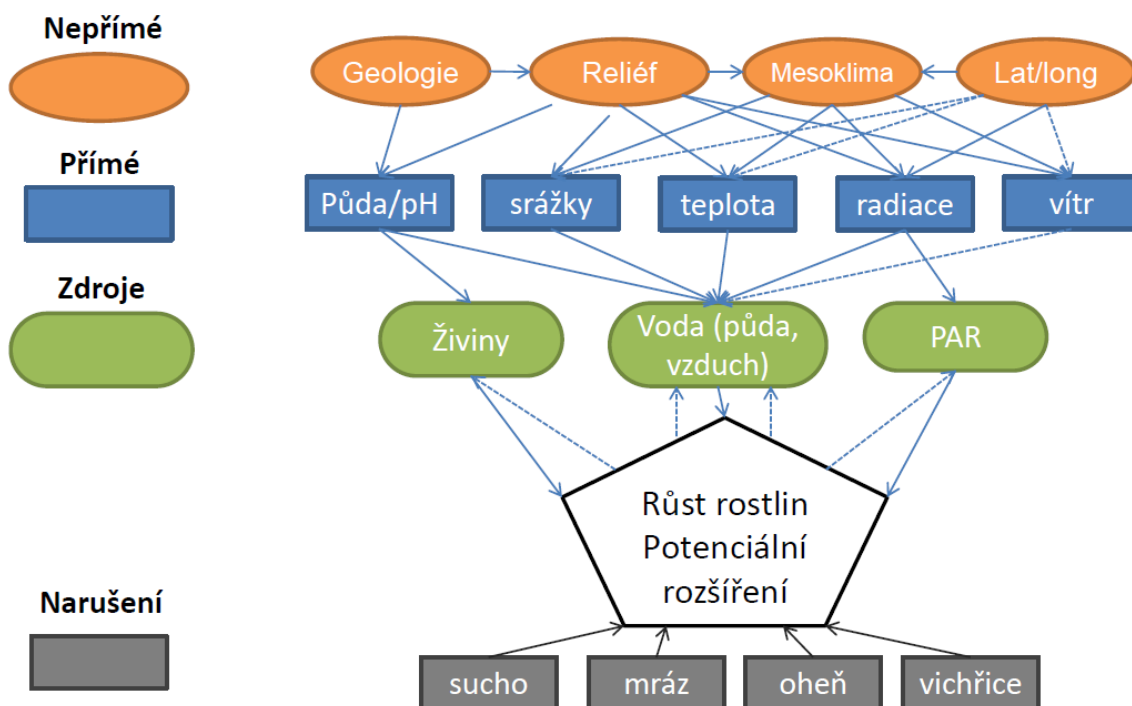
Zdroj: <https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/what-is-raster-data.htm>

Environmentální proměnné můžeme kromě velikosti buňky dále hodnotit a roztřídit podle různých kritérií, která nám pomohou interpretovat výsledné predikce. Například zda se jedná o proměnné kategoriální či spojité, abiotické či biotické, přímé či nepřímé. Také některé predikční algoritmy jsou citlivé na použití kategoriálních environmentálních vrstev, nebo nedokáží korektně pracovat v případě, kdy použijeme dohromady kategoriální a spojité vrstvy. Predikčním algoritmům je pak dále věnovaná samostatná kapitola. Pro kategoriální proměnné nemají také smysl některé operace, které můžeme provádět pro spojité proměnné – typicky matematické operace jako průměr, suma a další. To je třeba mít na paměti při operacích s kategoriálními rastry při jejich agregaci a interpolaci (např. při transformaci rastru vždy použít pro přiřazení hodnot metodu nejbližšího souseda – nearest neighbour). Kategoriální environmentální vrstvy nabývají diskrétních hodnot, které nemají numerický význam. Můžeme si to představit jako číselný převodník, kdy definujeme např., že: 1 = bezlesí, 2 = lužní les, 3 = pastvina... (Obrázek 7). Zde může docházet při modelování k chybám, ať už zaviněných nesprávnou interpretací proměnné algoritmem, nebo lidskou nepozorností, kdy není taková kategoriální proměnná řádně kódovaná jako tzv. faktor.



Obrázek 7 Příklad kategoriální a spojité proměnné. Kategoriální proměnná nabývá diskrétních hodnot, které kódují typ krajiny (např. 1 = bezlesí, 2 = lužní les, 3 = pastvina ...).

Kromě zmíněných technických aspektů spojených s environmentálními proměnnými je dobré mít také na paměti aspekty biologické či ekologické. A to zejména, zda se jedná o proměnnou, která popisuje přímý fyziologický efekt na zájmový organismus. V případě rostlin jsou takové přímé proměnné například teplota, množství živin, půdní a vzdušná vlhkost, příjem fotosynteticky aktivního záření. Ty mohou být reprezentované rastry s měřením mikroklimatu. (Austin & Smith 1989; Guisan & Zimmermann 2000; Mod et al. 2016). Nepřímou proměnnou je například elevace, která bývá při SDM často jednou z nejdůležitějších proměnných. Než aby rostliny byly ovlivněny počtem metrů nad mořem, jedná se spíše o reprezentaci gradientu teplotního, srážkového či radiačního, které elevace reprezentuje, a které přímo ovlivňují fyziologickou odpověď jedinců. Podrobněji schéma na obrázku 8. Podle toho, zda používáme proměnné přímé, nepřímé nebo oboje, je třeba pak interpretovat výsledné predikce potenciálního výskytu habitatu a zejména důležitost jednotlivých proměnných, které se na predikcích podílely.



Obrázek 8 Konceptuální model vztahu mezi zdroji, přímými a nepřímými environmentálními proměnnými a jejich vlivu na růst, vývoj a potenciální distribuci cévnatých rostli. Upraveno podle Guisan A. et Zimmermann N.E. (2000): *Ecological Modelling*, Obr. 3

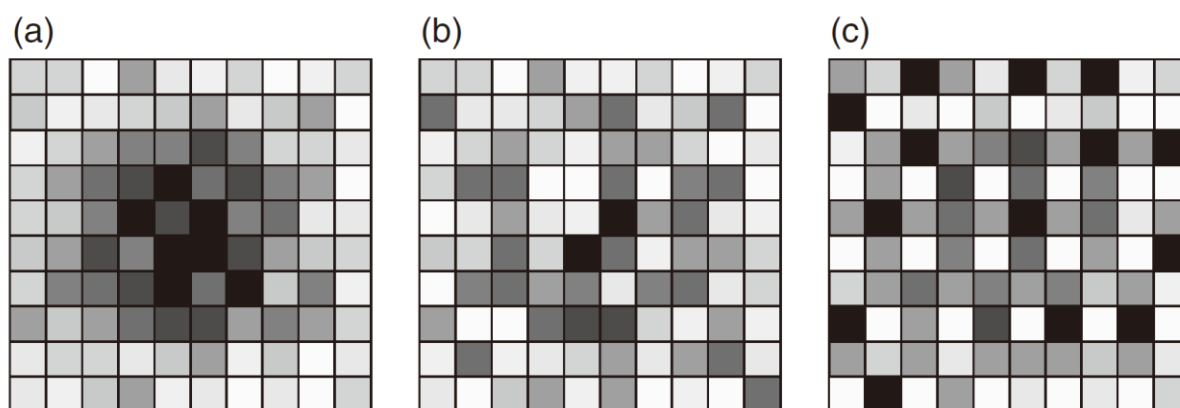
3.6.3 Trénovací výskytová data

Data o výskytu druhů jsou základním kamenem při modelování distribuce potenciálních habitatů (SDM). Ve většině případů jsou to bodová pozorování reprezentovaná geografickými souřadnicemi (bodem v mapě), o kterém máme informaci, zda se tam v daném čase cílový druh vyskytuje, či nevyskytuje. Pomineme-li technické aspekty správy a ukládání výskytových dat, kterému se věnuje samostatná kapitola této metodiky, pak je třeba se zabývat aspekty, jako je: i.) počet pozorování, ii.) přesnost a datum lokalizace, iii.) dostupnost absencí z terénního průzkumu, iv.) prostorová autokorelace výskytových dat a v.) pokrytí environmentálního gradientu v geografickém prostoru habitatových predikcí.

- i.) Počet pozorování, který použijeme pro SDM, může mít vliv na kvalitu výsledných predikcí. Některé studie ukázaly, že počet dat o výskytu druhů použitých pro SDM pozitivně koreluje s predikční schopností modelů (Hirzel & Guisan 2002; Reese et al. 2005), jiné studie tvrdí, že pro úspěšné modelování výskytu druhů je dostatečný počet pozorování v řádech desítek (Wiszniewski et al. 2008; Kremen et al. 2008). Některé studie říkají, že stačí 50 pozorování (Coudun & Gégout 2007). Odjinud se zase dozvídáme, že pokud nemáme k dispozici stovky pozorování, je na modelu stále co zlepšovat (Liu et al. 2019; Gábor et al. 2020). Z toho je zřejmé, že neexistuje univerzální rada, kolik pozorování je potřeba pro robustní odvození predikční mapy. Počet potřebných pozorování pro SDM závisí na mnoha faktorech, jako je jejich rozložení v prostoru (pokud je dobře pokrytý environmentální gradient, stačí méně pozorování než velký počet výskytů v jednom prostorovém klastru), volba predikčního algoritmu (některé algoritmy vyžadují větší množství pozorování například kvůli internímu rozdělování dat), velikost modelovaného území či proporce prezencí ku všem pozorováním (prevalence). I prevalence sama o sobě může mít vliv na SDM, zejména na výpočet ukazatelů kvality modelu. Někteří autoři tvrdí, že vyvážený poměr prezencí a absencí v datasetu, tedy prevalence kolem 0.5, je optimální (Reese et al. 2005), jiné studie říkají, že by prevalence měla odpovídat odhadované prevalenci v přírodě, tedy pokud je druh vzácný a nachází se na několika málo lokalitách, prevalence v datech pro SDM by pak měla být nižší, naopak pro běžné druhy vyšší (Meynard & Kaplan 2012). Specificky při SDM pro vzácné druhy je třeba se otázkou počtu pozorování zabývat, protože vstupních dat nebývá mnoho. Úskalím predikčních modelů je fakt, že jsou schopné vyprodukovat mapky potenciální vhodnosti habitatu pro daný druh i pro velká území na základě velice nízkých počtů pozorování – v řádu jednotek výskytů, a je pak na osobě, která výsledky takových modelů interpretuje, aby takové modely buď rovnou vyloučila, nebo interpretovala maximálně kriticky a s maximální opatrností. Modely na základě malého počtu pozorování mohou být užitečné jako vodítko pro zacílení terénního výzkumu, nebudou ale pravděpodobně použitelné pro zjišťování důležitosti jednotlivých environmentálních faktorů, které distribuci druhů determinují (Smith & Santos 2020).
- ii.) Při práci s výskyty druhů v SDM, reprezentovanými dvojicí geografických souřadnic, bychom měli mít na paměti, že data mají vždy nějakou prostorovou chybu. První úvaha by měla vést k cílovému organismu. Jak je organismus velký, je mobilní či sesilný, je ovlivňován bezprostředním či spíše širším okolím? Měli bychom se také ptát, zda je nutné používat pro daný organismus pouze recentní výskytová data, nebo je možné se zabývat také historickými záznamy, má smysl předpokládat, že se druh stále vyskytuje, nebo může vyskytovat na historických lokalitách? Dále můžeme odhlédnout od studovaného organismu a uvažovat o technických aspektech. Prostorová chyba vzniká při zaznamenání či odečtení souřadnic a jejich případné transformaci. Pokud používáme existující elektronické databáze, často bývá informace

o prostorové chybě součástí metadat, pokud používáme například údaje zaměřené turistickou GPS, informace o prostorové chybě bývá také součástí metadat, navíc je však třeba brát v potaz prostorovou chybu GPS modulu, která nemusí být započtena v prostorové chybě, kterou ukazuje GPS přístroj. Garmin³ například udává prostorovou chybu 15 m. S ohledem na prostorovou chybu je třeba si tato data představit spíše jako kružnice s poloměrem prostorové chyby, ve kterých se cílový jev nachází. Pokud je prostorová chyba příliš velká, environmentální podmínky v daném bodě mohou být s ohledem na členitost terénu dramaticky odlišné, což může vnášet značnou chybu do výsledných predikcí. Je také potřeba přihlídnout k rozlišení rastrů, které reprezentují environmentální podmínky v SDM. Pokud hodláme vytvářet mapy s velikostí buňky 1 x 1 km, nemusíme se příliš trápit s prostorovou chybou kolem 500 m. Pokud ale má být výsledná mapa podrobnější, například s velikostí buňky 10 x 10 m, pak už mohou být ve členitém terénu problematická data o výskytu s chybou kolem 50 m. Podobně jako u počtu pozorování nedá se ani u prostorové chyby dat jednoznačně říci, jaká prostorová chyba je v SDM ještě únosná a co už je příliš. Vždy záleží na účelu a interpretaci modelů.

- iii.) Při SDM má člověk často v ruce data, která sestávají z ověřených výskytů daných organismů – presence, ale neobsahují ověřenou terénní informaci o tom, že se druh na daném místě nevyskytuje – absence. To není pro modelování ideální situace, protože modely vystavěné na prezenčně-absenčních datech jsou patrně nejpřesnější (Elith et al. 2006), nicméně existují prozkoumané postupy, jak si poradit v situaci, kdy nejsou absenční data k dispozici. Nejčastěji za pomoci vygenerování tzv. pseudo-absencí. Jelikož se jedná o komplexní problém, který má vliv na SDM a výslednou interpretaci predikcí, je pseudo-absencím věnovaná samostatná podkapitola této metodiky.
- iv.) Prostorová autokorelace pojednává o podobnosti prvků v geografickém prostoru. Obrázek 9. Jak říká Toblerovo první pravidlo geografie: „všechno souvisí se vším, ale bližší věci jsou si podobnější než ty vzdálenější.“ Prostorová autokorelace je často zodpovědná za jistou část variability ve výskytových datech, která pak může mylně připadnout na vrub faktoru, který je pro výskyt daného druhu biologicky naprosto nevýznamný, ale je prostorově auto korelovaný (Fourcade et al. 2018).



Obrázek 9 Prostorová distribuce reprezentovaná v pravidelné čtvercové síti. a) pozitivní autokorelace, b) náhodná distribuce, c) negativní prostorová autokorelace. Převzato z (Zhu & Liu 2018)

Prostorová distribuce pozorování může mít například charakter izolovaných skupin (klastrů), s vyšší koncentrací výskytových dat na ploše. To může na jedné straně odrážet reálné rozšíření druhu, jenž je vázán na limitující faktor prostředí s klastrovou distribucí (výskyt jarních mrazů,

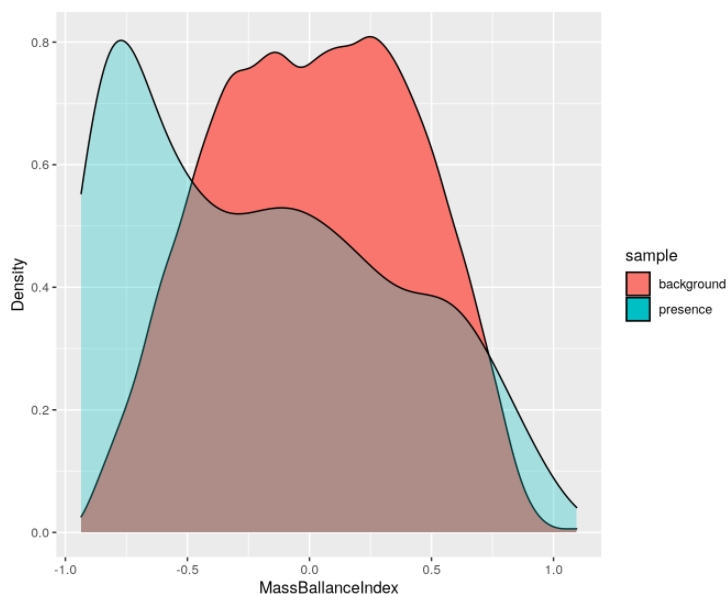
³ <https://support.garmin.com/en-US/?faq=aZc8RezeAb9LjCDpJpITY7>

nízké letní teploty, výskyt specifické horniny), například horské druhy, vápnomilné druhy, či druhy tolerující hadce. Na straně druhé může být klastrování výskytových dat způsobeno celkově malou prozkoumaností území na výskyt cílového druhu a tendence k vyhledávání lokalit v okolí známého výskytu. Proto je vždy potřeba taková data vizualizovat v mapě a zamyslet se nad jejich prostorovou distribucí a stejně tak je třeba se podívat na výsledné predikce potenciálního výskytu vhodných habitatů s promítnutím pozorovaných výskytů a zhodnotit, zda nevznáší prostorová distribuce dat do výsledku chyby, například v podobě promítnutí designu sběru dat.

- v.) U dat o výskytu druhů použitých pro SDM je také důležité přemýšlet o tom, jaký gradient v environmentálním prostoru reprezentují a zda se pokrytý gradient liší či odpovídá gradientu v prostoru plánované predikce. Od toho se pak často odvíjí kvalita výsledných map. Dobrým nástrojem, který nám může poskytnout představu o pokrytí gradientu, je vynesení frekvence hodnot daného gradientu pro výskytová data a pro náhodně vygenerované body pokrývající geografický prostor zamýšlené predikce (Obrázek 10). V případě, že vynášíme pouze prezence proti hodnotám v náhodně distribuovaných bodech v území plánované predikce, můžeme mírný posun gradientů vůči sobě vnímat jako biologicky relevantní signál svědčící o preferencích cílového druhu. V případě výrazného vzájemného posunu gradientů je třeba věnovat zvýšenou pozornost definici území, pro které chceme odvodit mapové výstupy s údaji o potenciálním výskytu vhodných habitatů. Může to být signál, že je cílové území příliš velké a výskytová data pokrývají jen úzkou část environmentálního gradientu v geografickém prostoru. V případě, že vynášíme denzitu hodnot environmentálního parametru pro místa pozorování prezencí i absencí cílového druhu, pak se dá očekávat větší překryv histogramů než pro čistě prezenční data.

3.6.4 Pseudo-absence

Jak již bylo zmíněno výše, při SDM se často potýkáme se situací, kdy nejsou k dispozici ověřené



Obrázek 10 Histogram denzity environmentálního faktoru Mass Balance Index (viz. kapitola 4.2) na místech ověřeného výskytu modelovaného druhu (modře) a náhodně rozložených bodů v oblasti predikce (červeně). Mass Balance Index ukazuje trend u cílového druhu osidlovat místa s nižší hodnotou, než je hodnota požadová tohoto faktoru.

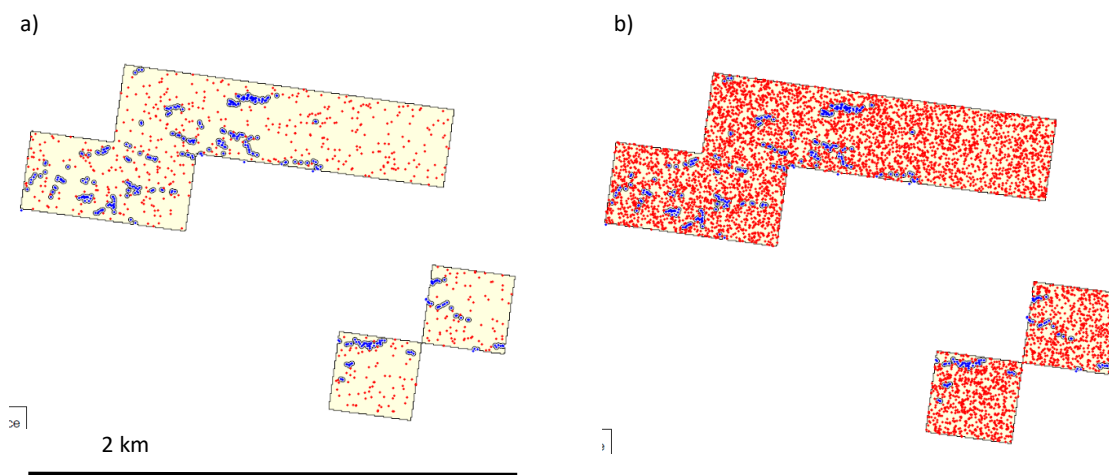
terénní absence zájmového druhu. Většinou nám ověřené absence chybí v případě, kdy používáme data, která nepocházejí z terénních průzkumů s výzkumným designem pro SDM nebo vůbec jakýmkoliv designem. Základní otázkou floristického výzkumu většinou je, kde se cílový druh vyskytuje, než to, kde chybí.

Existují i metody SDM založené pouze na prezenčních datech, ale ty jsou používány okrajově a ani tato metodika jich nevyužívá. Jedná se zejména o metody popisující a srovnávající environmentální prostor (environmental envelope models), ve kterém se druhy vyskytují, a hledající analogy mimo území známé prezence, z nejznámějších například Genetic algorithms for rule production (GARP) (Stockwell 1999) nebo Ecological niche factor analysis (ENFA) (Hirzel & Guisan 2002). Na tomto místě považujeme za vhodné krátce zmínit také algoritmus maximální entropie implementovaný v populárním SDM software MaxEnt (Phillips et al. 2006), který bývá často označován jako metoda pracující pouze na základě prezencí druhů. Pravdou však je, že MaxEnt sice od uživatele žádá vstupní absenční data nevyžaduje, nicméně v průběhu modelování jsou generována požadovaná pozorování, podobně jako pseudo-absence u jiných metod.

Většina metod SDM zpravidla vyžaduje data, která obsahují nějakou protiváhu ověřených prezencí, a pokud to nemohou být ověřené absence, pak to musí být pseudo-absence či přesněji řečeno požadovaná pozorování. To je i případ algoritmů využitých v této metodice. Jelikož se obecně jedná o umělé generování absencí, je třeba mu věnovat patřičnou pozornost, protože zvolený přístup může mít vliv na výsledky SDM. Podobně jako mohou být algoritmy citlivé na celkový počet dat o výskytu (prezencí, absencí), která jsou použita k trénování, mohou být algoritmy citlivé i na poměr prezencí ku pseudo-absencím, a mohou reagovat jinak než na poměr reálných prezencí a absencí. Optimální počet pseudo-absencí může navíc být specifický pro konkrétní algoritmy. Pro generalizované či aditivní modely se jako optimální jeví poměrně vysoký počet pseudo-absencí v řádech tisíců, až 10 000 bez ohledu na počet reálných prezencí (Barbet-Massin et al. 2012) viz Obrázek 11. Pro metody strojového učení se dle nejaktuálnějších poznatků jeví jako optimum vygenerovat nízký několikanásobek pseudo-absencí ku prezencím (Liu et al. 2019), což se obrací k prevalenci menší než 0,5 diskutované výše. To může být pro vzácné druhy situace skutečně blízká realitě.

Kromě samotného počtu pseudo-absencí je dobré se věnovat jejich prostorové distribuci. Nejčastějším způsobem, jak pseudo-absence vygenerovat, je náhodné rozložení tak, že každá lokalita má potenciálně stejnou šanci se stát pseudo-absencí bez ohledu na to, jaké tam panují environmentální podmínky, nebo zda se tam druh vyskytuje, či nikoliv. Při náhodném výběru pseudo-absencí se může stát, že i lokality s ověřenou prezencí padnou do náhodného výběru pseudo-absencí. Kromě základního přístupu se zcela náhodným generováním můžeme přistoupit k informovanějším technikám, jako je náhodný stratifikovaný výběr, kde jsou definované buď geografické oblasti (např. čtvercová síť) nebo oblasti na základě podobnosti environmentálních podmínek (např. výstup z klastrové analýzy environmentálních vrstev) a do těchto kategorií se pak náhodně generuje buď stejný počet, nebo proporční počet pseudo-absencí k velikosti prostorové třídy. Kromě náhodného či náhodného stratifikovaného výběru můžeme aplikovat také výběr řízený, například na základě znalosti environmentálních podmínek výskytu druhů můžeme hledat podmínky co možná nejvíce odlišné a tam umístit pseudo-absence, nebo je možné pseudo-absence umístit na základě expertní znalosti území do míst, kde předpokládáme, že se druh nemůže vyskytovat.

Ačkoliv by se na první pohled mohlo zdát, že řízený výběr pseudo-absencí dává smysl a měl by přispět k lepším výsledkům SDM, opak může být pravdou. Řízeným výběrem pseudo-absencí vnášíme do modelů subjektivní předpoklady, které mohou mít vliv zejména na ukazatele kvality modelů. Odhady kvality modelů pak mohou být falešně příliš optimistické a modely se mohou jevit kvalitnější a predikce přesnější, než ve skutečnosti jsou. Proto se spíše doporučují metody náhodného či stratifikovaného náhodného výběru pseudo-absencí (Barbet-Massin et al. 2012).



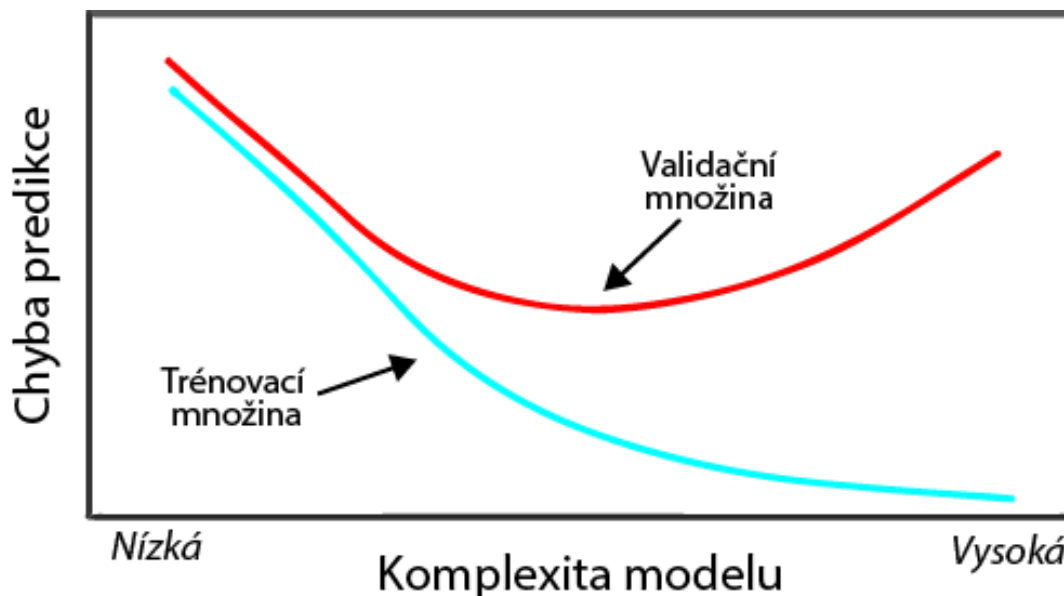
Obrázek 11 Příklad několika čtvercových výřezů mapy výskytových dat Rojovníku bahenního. Ověřené prezence (modře), pseudo-absence (červeně). a) 1 000 náhodně vygenerovaných pseudo-absencí, b) 10 000 náhodně vygenerovaných pseudo-absencí.

3.6.5 Modely a algoritmy

K modelování výskytu druhů či habitatů vhodných pro jejich výskyt můžeme použít dva přístupy. Přístup mechanistického nebo korelativního (deterministického) modelování. Mechanistické modely se zaměřují především na fyziologické limity daného organismu. Musíme znát, jaké podmínky toleruje, jaké hodnoty a kombinace environmentálních tlaků jsou pro daný organismus optimální a jaké jsou letální. Pak sestavujeme matici pravidel, či matematickou formuli, kterou aplikujeme na environmentální podmínky, například (Chaine & Beaubien 2001). Většinou se však při SDM využívají spíše korelativní metody, kdy se hledají analogy environmentálních podmínek, ve kterých se daný organismus vyskytuje na místech, kde o jeho výskytu nemáme informaci bez ohledu na fyziologii daného druhu. Korelativní metody jsou využívány také v této metodice.

Základním úkolem algoritmů v SDM je identifikovat potenciálně komplexní, nelineární vztahy v mnohorozměrném prostoru environmentálních proměnných a výskytu druhů či jejich habitatů. V současné době je k dispozici celá džungle algoritmů, které můžeme pro modelování potenciálního habitatu druhů využít, navíc je možné je kombinovat a vytvářet predikce na základě výsledků více modelů (ensemble, stack). Situaci ilustruje populární knihovna pro SDM „caret“ (Kuhn 2008), která jako jedna z mnoha rozšiřuje statistický software „R“ (R Core Team 2016) a nabízí v současné době 238 algoritmů⁴ pro SDM. Mnohé z nich jsou implementací téhož základního algoritmu s různými možnostmi pro nastavení a optimalizaci, nicméně stále je tu několik stovek možností....

⁴ <http://topepo.github.io/caret/available-models.html>



Obrázek 12 Chyba predikce v závislosti na komplexitě modelu. Červená linie reprezentuje skutečnou chybu vypočtenou na základě validačních dat, která nikdy nevstoupila do trénování modelu. Ta ukazuje, že příliš komplexní modely se dokonale přizpůsobí trénovacím datům. Upraveno podle (Hastie et al. 2009).

S rozvojem výpočetních kapacit se dostaly ke slovu i velice náročné metody zahrnující mnoho iterací (opakování) při tvorbě a optimalizaci modelu za použití velice komplexních algoritmů, které se dokáží takřka dokonale adaptovat jakýmkoliv vstupním datům a odhalit i tu nejhlubší souvislost. To s sebou však nese riziko přetrénování modelu, který je dokonale uzpůsobený trénovacímu datasetu, ale při validacích selhává, protože trénovací data zpravidla nepostihují celou ekologickou niku daného druhu. To ilustruje obrázek 12, kde je vynesena chyba predikce v závislosti na komplexitě modelu. Chyba predikce se zdánlivě limitně blíží nule pro velice komplexní modely, pokud spoléháme na ukazatele predikční chyby odvozené pouze z trénovacího datového souboru (modrá linie), ale skutečná predikční chyba odvozená z validačního datového souboru, který nebyl nikdy použit k trénování modelu, se zvyšuje právě kvůli přetrénování modelu. To může nastat zejména u modelů s využitím metod strojového učení, které jsou schopné se přizpůsobit trénovacím datům velice flexibilně.

Dále zde uvádíme alespoň základní rozdělení dostupných algoritmů pro SDM se zaměřením na některé nejčastěji používané a podrobným popisem těch, které využívá předkládaná metodika. Korelativní techniky s prezenčními a absenčními (pseudo-absenčními) daty modelování potenciálního výskytu druhů či jejich habitatů můžeme hrubě rozdělit na dvě skupiny, avšak hranice mezi nimi není ostrá:

1. Regresní metody
2. Metody strojového učení

Z regresních metod se při SDM nejčastěji setkáváme s: generalized linear models (GLM), generalized additive models (GAM), multivariate adaptive regression splines (MARS).

Regresní metody vycházejí v základu z principů lineární regrese, která predikuje závislou proměnnou jako funkci jedné či několika vysvětlujících proměnných a jejich interakce, pokud je v modelu explicitně definovaná. Lineární regrese předpokládá lineární vztah mezi závislou a vysvětlující proměnnou, tedy že konstantní změna ve vysvětlující proměnné vyvolává konstantní reakci závislé proměnné. Předpokladem pro lineární regresi je normální rozložení závislé proměnné. Tento předpoklad zpravidla není v biologii obecně ani v SDM naplněn, a proto se využívá zobecněných lineárních modelů (GLM) nebo zobecněných aditivních modelů (GAM).

V této metodice je jako zástupce regresních metod využit generalizovaný lineární model (GLM). Ten je aplikovatelný i na jiné rozložení závislé proměnné než pouze normální. Na rozdíl od lineární regrese, která je konstruována metodou nejmenších čtverců, generalizované modely operují s metodou maximální věrohodnosti (maximum-likelihood estimation), který maximalizuje pravděpodobnost, že pozorované hodnoty pocházejí z předpokládaného rozdělení pravděpodobnosti. Modely jsou tedy flexibilnější, vyžadují ale správnou specifikaci rozdělení, ze kterého pochází závislá proměnná, což je ale u SDM zpravidla rozdělení binomické (0,1). Díky link funkci, která spojuje závislou a vysvětlující proměnnou, jsou tyto modely schopné efektivně popsat i nelineární vztahy mezi nimi. V Případě SDM, kdy je závislá proměnná binomická (prezence, absence, 0,1), je link funkcí logit, což je přirozený logaritmus pravděpodobnosti daného jevu (0,1). Zde hovoříme o logistické regresi, která je pouze speciálním případem zobecněného lineárního modelu (GLM). Zobecněný lineární model je etalonem pro SDM. Je to nejdéle používaná metoda v SDM, ověřená mnoha studiemi s dobrým matematickým i statistickým zázemím.

Mezi výhody generalizovaných lineárních modelů patří zejména:

- Vrací stabilní predikce
- Jsou schopné výrazné generalizace modelovaných jevů
- Nejsou zpravidla výpočetně příliš náročné
- Výsledky jsou dobře interpretovatelné
 - vztahy mezi závislou a vysvětlujícími proměnnými jsou matematicky poměrně snadno popsitelné
 - Příspěvek jednotlivých proměnných k celkové predikci (důležitost proměnných) je dobře definovatelný

Jejich nevýhodami pak jsou:

- Neschopnost popsat příliš komplexní vztahy
- Jsou velice citlivé na kolinearitu vysvětlujících proměnných
- Nejsou schopné odhalit vzájemné interakce vysvětlujících proměnných, pokud nejsou interakce v modelu explicitně definované předem
- Předpokládají specifické rozložení dat
- Práce s kategoriálními vysvětlujícími proměnnými vyžaduje zvýšenou pozornost a speciální přístup, který není vhodný pro proměnné s mnoha kategoriemi, zejména pakliže se počet kategorií blíží počtu pozorování

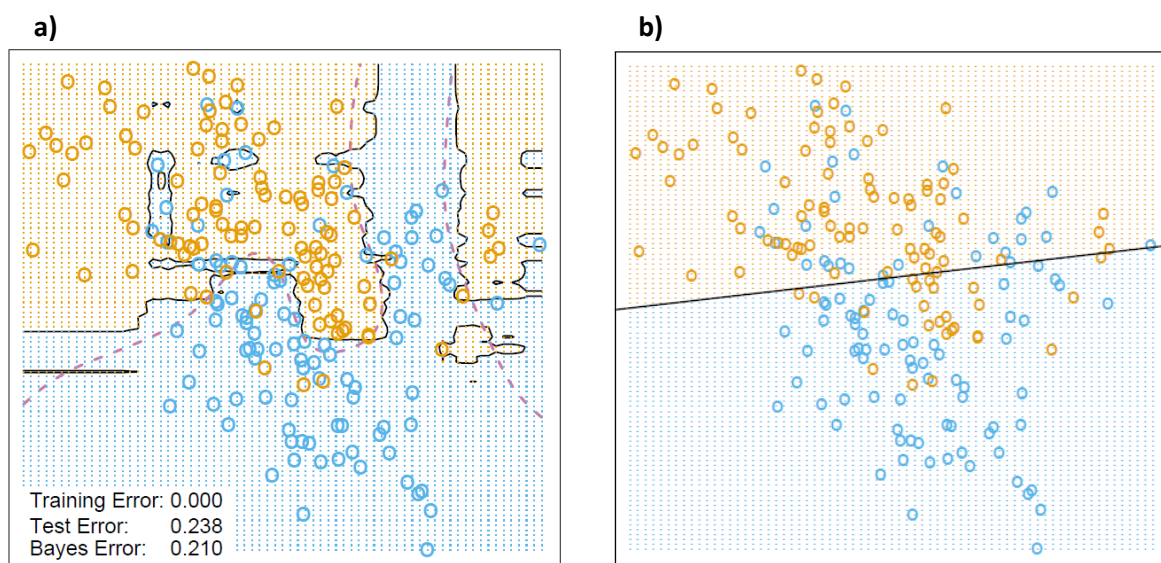
Z metod strojového učení se při SDM nejčastěji setkáváme s: Artificial Neural Networks (ANN), Random Forests (RF), Boosted Regression Trees (BRT), Support Vector Machines (SVM), MaxEnt.

Metody strojového učení (Machine learning) jsou v současné době v SDM velice populární a hojně využívané, ostatně jako v mnoha jiných odvětvích lidské činnosti. Strojovým učením zde rozumíme schopnost počítačového algoritmu / systému se automaticky měnit tak, aby zefektivnil zachycení daného jevu. Algoritmy strojového učení produkují matematické modely či soubory pravidel na základě vstupních dat – učí se z nich. Metody strojového učení jsou schopné predikovat či rozhodovat o cílovém jevu, aniž by pro rozhodování o něm byly předem naprogramované. V případě predikčního modelování se zpravidla jedná o učení řízené / učení učitelem (supervised learning), kde člověk jasně specifikuje zájmová pozorování v podobě trénovacích dat. Algoritmus se v případě SDM snaží vysvětlit pravděpodobnost prezence či absence druhu, respektive pravděpodobnosti vhodnosti habitatu pro daný druh. Většina algoritmů strojového učení umožňuje další ladění / nastavení

„tuning“ parametrů chování, které mají vliv na výsledné predikce. Například metoda křížové validace, počet iterací či mezních hodnot koeficientů jako je např. u random forest počet stromů, velikost koncového nodu atp. Ladění metod strojového učení může být prováděno buď ručně, na základě rozhodnutí modeláře, ale může být také prováděno automaticky či poloautomaticky, kdy algoritmy iterativně „hledají“ optimální nastavení na základě stanoveného kritéria, např. přesnosti predikce (Kappa, ROC, RMSE).

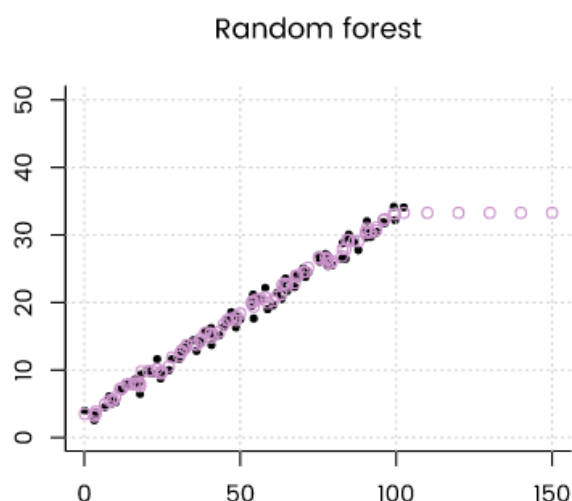
Z algoritmů strojového učení není snadné vybrat jeden, který by byl pro SDM optimální. Různé algoritmy mohou vracet různě přesné predikce v závislosti na jejich dalším ladění, ale také v závislosti na geografických podmínkách či specifických vlastnostech modelovaného organismu. Kvůli častokrát velice odlišným přístupům, které jednotlivé metody strojového učení využívají, není často ani možné jednoduše porovnat kvalitu jejich výstupů. Pro tuto metodiku jsme se rozhodli využít algoritmus random forest jako jeden z nejpopulárnějších a nejčastěji využívaný v SDM v praktických aplikacích posledních let (Liu et al. 2013; Mi et al. 2017; Pecchi et al. 2019; Cerrejón et al. 2020).

Random forests (RF) je algoritmus, který opakovaně vytváří klasifikační či regresní stromy na vybraném podsouboru trénovacích dat. Podsoubor trénovacích dat, který se pro každé opakování mění (Obrázek 4). Pro každý nod, ve kterém se rozhodovací strom větví, je použit náhodně vybraný podsoubor prediktorů. Jednotlivé stromy jsou pak agregovány do výsledného průměrného stromu. Během agregace jednotlivých stromů probíhá výpočet důležitosti jednotlivých vysvětlujících proměnných (Hollings et al. 2017). Jednotlivé klasifikační a regresní stromy jsou schopné se „natrénovat“ velice přesně na vstupní data (obrázek 12 a obrázek 13).



Obrázek 13 Příklad klasifikace binární proměnné (modré = 0, oranžové =1). Černá linie reprezentuje hranici, podle které se model rozhoduje, do které třídy data patří. Random Forest (a) lineární model (b). Převzato z (Hastie et al. 2009).

Proto se provádí výpočet několika stovek stromů na podsouborech dat navíc v každém dělení stromu s podsouborem vysvětlujících proměnných. To zvyšuje možnosti zobecnění (predikce, interpolace) pomocí výsledků Random Forest. Problematická je situace s extrapolací, tedy s predikcí mimo rozsah hodnot pokrytých trénovacími daty (obrázek 14). Problematická je například situace, kdy máme trénovací data z nadmořských výšek mezi 0 a 100 m, ale chceme predikovat i v nadmořských výškách nad 100 m. To je slabá stránka algoritmu Random Forest.



Obrázek 14 Příklad predikce hodnot mimo rozsah s pomocí random forest. Černé body znázorňují trénovací data, kde je vidět jasný lineární trend v prostoru dvou proměnných na ose x a y. Fialové body znázorňují predikci random forest, která je velice přesná v oblasti trénovacích dat, ale pravděpodobně velmi nepřesná mimo rozsah. Převzato z : <http://freerangestats.info/blog/2016/12/10/extrapolation>

3.6.6 Performance a validace modelů

Nedílnou součástí tvorby každého modelu je jeho validace, která umožňuje kvantifikovat, do jaké míry se předpovědi modelu shodují s pozorováními. V optimálním případě je model validován na nezávislém vzorku. Takový vzorek může například pocházet z jiné oblasti, nebo může být nasbírán v jiném období, viz schéma na Obrázku 3. V reálných podmínkách jsou nicméně nezávislá data odpovídající kvality zřídka dostupná. Z tohoto důvodu jsou modely obvykle validovány s využitím křížové validace (krosvalidace), která spočívá v (opakovaném) rozdělení dat reprezentujících výskyt organismů do dvou skupin, z nichž jedna je použita pro natrénování modelu a druhá pro jeho testování viz schéma na Obrázku 4. Obvyklé je například použití dvou třetin dat pro trénování a zbylé třetiny pro testování modelu. Další obvyklý způsob je rozdělení do 5ti nebo 10ti přibližně stejně velkých skupin, z nichž jedna slouží jako testovací a zbylé skupiny jako trénovací. Tento proces se několikrát opakuje, pokaždé s jinou testovací skupinou (tzn. každá skupina je právě jednou použita pro testování). Poslední metodou, která je často používána, je tzv. leave-one-out krosvalidace. Jedná se o n -násobnou křížovou validaci, kdy n je počet všech pozorování. Z dat se vždy vyjme jen jedno pozorování (presence nebo absence druhu) pro testování a zbylá pozorování jsou použita pro trénování modelu. Celý proces se opakuje n -krát. Výhodou leave-one-out krosvalidace je použití téměř všech pozorování pro trénování modelu, zejména v případě, kdy je celkový počet pozorování nízký.

Důležitým předpokladem křížové validace je nezávislost trénovacích a testovacích dat. Pokud není tento předpoklad dodržen, odhad přesnosti modelu bude příliš optimistický a budou upřednostněny

příliš složité modely. Jinými slovy, pokud není splněna podmínka nezávislosti mezi daty použitými pro trénování a testování modelu, modely se jeví spolehlivější, než ve skutečnosti jsou. Což snadno vede k větší důvěře v jejich výstupy, než jakou si ve skutečnosti zaslouží. Pokud jsou modely vytvářeny s cílem předpovídat distribuci v rámci jedné oblasti a za stejných environmentálních podmínek (tj. ve stejném období) a pokud jsou určeny pouze k předpovědi, aniž by byly odvozovány kauzální prediktory, lze záznamy do trénovací a testovací skupiny pro krosvalidaci rozdělit náhodně.

Typickým výstupem modelů je pravděpodobnost výskytu druhu, kterou lze volbou vhodné prahové hodnoty ("*threshold*") transformovat na prezence a absence organismu. S ohledem na výstup z modelů lze do dvou skupin rozdělit i validační metriky: diskriminační metriky a kalibrační metriky. První skupina měří schopnost modelu správně rozlišit prezence a absence organismu a patří k častěji používaným způsobům validace (Lawson et al. 2014; Leroy et al. 2018; Liu et al. 2019). Zatímco druhá skupina odráží schopnost předpovězených pravděpodobností výskytu odrážet pozorovaný podíl míst obsazených studovaným organismem. Níže uvádíme několik běžných metrik používaných pro hodnocení modelů. Nutno však předeslat, že ačkoliv potřeba důkladného vyhodnocení modelu je široce uznávána, existují různé názory na to, jaké vlastnosti modelu jsou důležité a jak je vhodně testovat. Téma validace modelů druhové distribuce je komunitou aktivně diskutováno a je obtížné poskytnout jednoznačná doporučení, snad kromě toho, že modely je vždy dobré validovat s použitím několika metrik spadajících do různých skupin (viz dále).

Existuje několik způsobů dělení validačních metrik. Jedno z takových dělení zohledňuje, zda jsou metriky ovlivňovány prevalencí či nikoliv. Obvykle bývá doporučováno používat metriky, které prevalencí ovlivňovány nejsou ("*prevalence-independent*"). Nicméně relativně nedávno Lawson et al. (2014) ukázal, že takové metriky nemohou být použity při kalibraci modelu (tvorbě modelu). Další dělení zohledňuje výstup z modelů (binární presenci/absenci versus kontinuální pravděpodobnost výskytu). Některé metriky lze aplikovat pouze na binárním výstupu modelů ("*threshold-dependent*"), tedy presence absence, které nicméně mohou vzniknout transformací pravděpodobnosti výskytu volbou vhodné prahové hodnoty ("*threshold*"). Další metriky lze využít přímo pro validaci pravděpodobnosti výskytu ("*threshold-independent*"). Všechny *threshold-dependent* metriky jsou založeny na některých nebo všech prvcích chybové matice ("*error matrix*"). Tedy matici, které porovnává modelované a pozorované prezence a absence organismu.

Lawson et al. (2014) zohlednil dvě výše zmíněné dělení (charakteristiky) a rozlišil čtyři skupiny (tedy zda měří pravděpodobnost výskytu nebo presenci/absenci a zda výstup modelu spadá do kategorie binární nebo zda se jedná o kontinuální pravděpodobnost výskytu). Z nedávno doporučených metrik, které zatím nebyly pro modelování příliš použity, jsou zahrnuty také Jaccardův a Sørensenův index (Leroy et al. 2018).

Binární (*threshold-dependent*) diskriminační metriky:

- Sensitivita, Specificita, TSS (True Skills Statistics), Jaccard's similarity index, Sørensen's similarity index

Binární (*threshold-dependent*) kalibrační metriky:

- Cohen's Kappa

Ačkoliv používání binárních (*threshold-dependent*) metrik je relativně běžné a mnoha studiemi doporučované, najdou se studie, které naopak ukazují, že jejich používání by mělo být omezeno ve prospěch *threshold-independent* metrik (Lawson et al. 2014). Ačkoliv i zde je třeba být k metrikám kritický (Lobo et al. 2008).

Kontinuální (*threshold-independent*) diskriminační metriky:

- AUC (Area Under Curve), Pearson's correlation, Spearman's correlation

Kontinuální (*threshold-independent*) kalibrační metriky:

- RMSE (Root-Mean-Square Error), Deviance, AIC (Akaike's Information Criterion), R-squared

Jak již bylo řečeno výše, žádná metrika neposkytuje univerzální hodnocení kvality modelů. Navíc, uvedené metriky předpokládají dostupnost prezenčně absenčních dat pro validaci modelu a diskriminační metriky také předpokládají, že prevalence vzorku se rovná druhové prevalenci, což nemusí být vždy dodrženo (Leroy et al. 2018). Pokud taková data dostupná nejsou, je třeba důkladně zvážit adekvátnost jejich použití.

3.6.7 Limity a rizika SDM

Jak již bylo na mnoha místech teoretického úvodu řečeno, SDM má svoje limity a k jeho výsledkům je třeba přistupovat kriticky. Vždy se jedná o neúplnou a počítačovou reprezentaci komplexních vztahů v reálném světě. Jak říká britský statistik George Box: *all models are wrong, but some are useful*.

Je také třeba mít na paměti, že výsledné mapy SDM nejsou mapami rozšíření a nezobrazují pravdivý stav o distribuci druhu v terénu, spíše jsou to potenciální vhodnosti daných lokalit s ohledem na známou distribuci druhů.

Výsledky SDM mohou být v některých případech pro daný druh biologicky či ekologicky naprosto nerelevantní, což se často projeví nízkou kvalitou modelu, měřeno například Kappa, TSS, confusion matrix. I ze zcela irelevantního modelu může vzejít líbivá mapka. Proto je třeba věnovat velkou pozornost výstupu modelu před samotnou predikcí a jejím využitím pro terénní průzkum.

3.6.8 Další užitečné zdroje a literatura

- Kniha: Janet Franklin: Mapping Species Distributions, Spatial Inference and Prediction, Cambridge University Press, 2010, ISBN:9780511810602, DOI: doi.org/10.1017/CBO9780511810602
- Kniha: Trevor Hastie, Robert Tibshirani, Jerome Friedman: The Elements of Statistical Learning Data Mining, Inference, and Prediction, Springer, New York, NY, 2009, ISBN: 978-0-387-84858-7, DOI: doi.org/10.1007/978-0-387-84858-7
- Kniha: Pearson, Richard G.: Species' distribution modeling for conservation educators and practitioners, Synthesis. American Museum of Natural History 50 (2007): 54-89.
- Online: Dokumentace R knihovny *caret* využitá v predikčním software této metodiky, <http://topepo.github.io/caret/index.html>
- Online: Návod pro praktické modelování výskytu druhů v R, https://rspatial.org/raster/sdm/1_sdm_introduction.html

4 Praktické vyhledávání vzácných druhů

Tato metodika je postavena na předpokladu, že mapy potenciální vhodnosti habitatu výskytu vzácných druhů mechorostů a lišejníků mohou urychlit a zefektivnit vyhledávání skutečného výskytu těchto druhů. Navrhujeme metodiku několika po sobě jdoucích kroků, které vedou k vytipování lokalit se zvýšenou pravděpodobností nalezení cílových druhů.

>>>>> <https://predikce.ibot.cas.cz> <<<<<<

Pro potřeby této metodiky jsme vytvořili uživatelské rozhraní pro výpočet predikčních modelů výskytu vhodných habitatů cílových druhů. Predikční aplikace je veřejně dostupná na adrese <https://predikce.ibot.cas.cz/>, dostupná pouze v angličtině, a naopak dokáže predikovat pouze výskyty na území ČR. Aplikace očekává vstupní výskytová data cílového druhu, environmentální data jsou součástí aplikace. Aplikace vrací v prvním kroku popisné statistiky a performanci modelů, na základě interakce s uživatelem pak vrací ve druhém kroku georeferencované mapy potenciálního výskytu vhodného habitatu cílového druhu. Z map je možné snadno vyčíst místa s nejvyšší pravděpodobností výskytu habitatu. Ty je možné dále v GIS připravit jako terénní podklady, ať už v podobě tištěné mapy nebo výskytových bodů do GPS či chytrého zařízení.

Vstupní data

csv tabulka

x, y souřadnice
presence, absence (0,1)
v případě čistě prezenčních dat jsou
automaticky generované pseudoabsence.

databáze DaLiBor

výběr druhu z databáze
generují se pseudoabsence

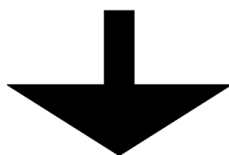
vzniká trénovací a validační dataset
+
environmentální data (topografie, klima, vegetace)

Model

nastavení území pro predikce  definováno daty
celá ČR
+
GLM  algoritmy + trénovací data
Random Forest 
+
ověření modelu na validačním datasetu

GIS a vizualizace

import a vizualizace výsledných map v GIS (QGIS)
generování bodů pro terénní průzkum (QGIS)
import bodů do navigačního zařízení



Průzkum lokalit - terénní šetření

Obrázek 15 Schéma pracovního postupu přípravy podkladů pro vyhledávání vzácných druhů dle této metodiky.

4.1 Vstupní výskytová data

Predikční aplikace může přijímat vstupní data ve dvou formátech.

- Jméno taxonu, který má v databázi DaLiBor alespoň 5 výskytů. (Create new prediction task from DaLiBor records)
- Textová tabulka ve formátu csv, která obsahuje x, y souřadnice a informaci o tom, zda je na dané souřadnice prezence (1) nebo absence (0) daného druhu. (Create new prediction task from CSV table)

V případě predikce z DaLiBor a v případě, kdy uživatel neposkytne žádné absence, aplikace generuje pseudo-absence.

Species distribution online application

This is online software created to build Species Distribution Models (SDM) and compute maps of potential species habitat suitability. The software works in two modes. Firstly you can provide table of coordinates or secondly you can select species from DaLiBor database.

For basic information about species distribution modeling (SDM) in plants conservation we recommend this paper: [Using species distribution models to predict new occurrences for rare plants.](#)

To get some general knowledge about species distribution modeling, we recommend this book: [Mapping Species Distributions.](#)

[Create new prediction task from CSV table](#)

[Create new prediction task from DaLiBor records](#)

Obrázek 16 Grafické rozhraní predikčního software (<https://predikce.ibot.cas.cz>), kde si uživatel jako první krok zvolí, z jakého datového zdroje budou do modelu vstupovat výskytová data.

- Výskytová data z DaLiBor

>>>>> <https://dalibor.ibot.cas.cz/> <<<<<<

Create new prediction task

Taxon from Dalibor	Buxbaumia viridis [425]
Email	
Spatial envelope typ	buffer
Buffer size [m]	3000
Submit prediction task	

Obrázek 17 Rozhraní online predikčního software pro modelování s daty z DaLiBor

Pokud je cílovým druhem mechorost či lišejník, pak je výhodné využít záznamy uložené v databázi DaLiBor, podrobněji popsané v kapitole 3.1. DaLiBor je národní výskytová databáze mechorostů a lišejníků, jejíž zřízení bylo motivováno také touto metodikou. DaLiBor je přístupná pro registrované uživatele (žádost o registraci se odesílá e-mailem správcům databáze). Databáze poskytuje jednotné prostředí pro správu a ukládání dat o výskytu mechorostů a lišejníků na území ČR, jejich validaci a další strojové zpracování. Pro informativní účely, či jako první explorativní analýzu environmentálního

vzorke výskytu vhodného habitatu je možné nekriticky použít záznamy z DaLiBor tak, jak tam jsou. To ale může vést k výrazně méně kvalitnějším predikcím (a to může být falešně příliš optimistický, ale i falešně příliš pesimistický model), protože v DaLiBor jsou skladovaná (zejména v případě mechorostů) i data s velkou nepřesností zaměření a data bez odborné validace.

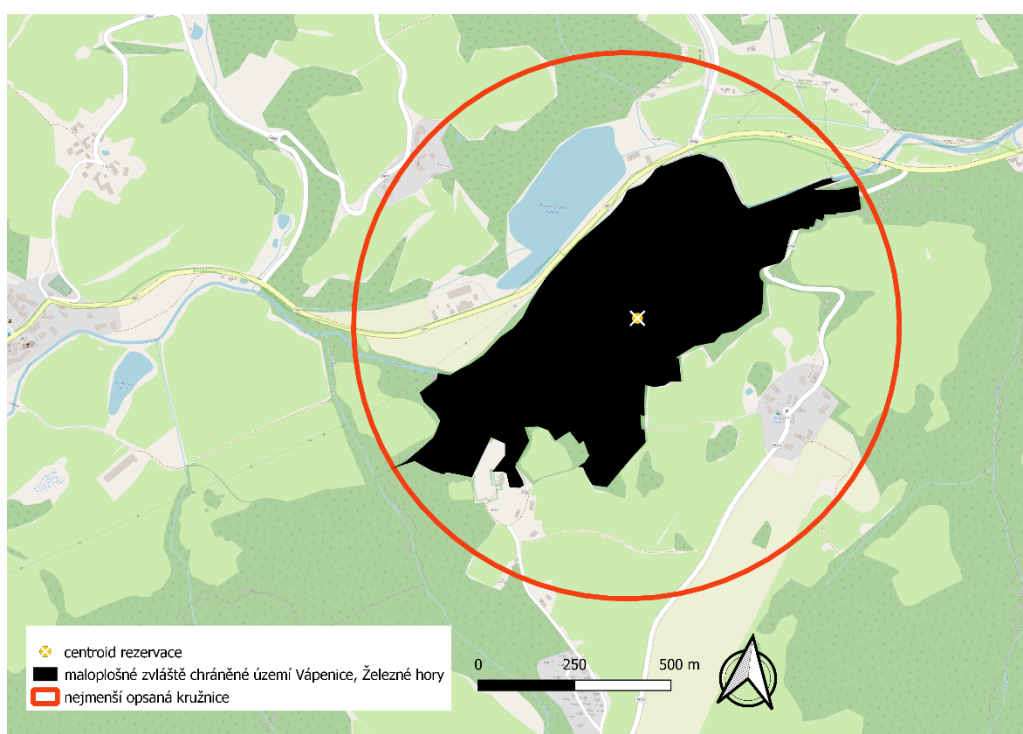
The screenshot shows the DaLiBor search interface. At the top, there is a navigation bar with 'DALIBOR' and links for 'Mapy', 'AdminMapy', and 'Ke stažení'. There are also buttons for 'Přejít na mapu taxonu...' and 'Přejít na záznam ID...'. Below this is the title 'Vyhledávání záznamů nálezových dat'. A search button labeled 'vyhledat' is prominent. To the left of the search filters, there is a list of labels: 'taxon', 'původní jméno', 'obec', 'lokalita', 'lokalita nebo obec', 'nadmořská výška', 'mapovací pole', 'nemá pole ani kvadrant', and 'poznámka'. The search filters themselves are input fields containing: 'ror*pa*/Roripa/roripa*', 'beneš*nad*', 'vápenice', 'vápenice', a range '5542;5432b;5432c;5433*', a checkbox for 'nemá pole ani kvadrant', and '*dřívě*vulgatum*'. To the right of the filters, there is a dropdown menu for 'původ' with options like 'Česká národní fytoecologická da...', 'Excerpce Bryonory – lišejníky', 'Excerpce Bryonory – mechorosty', and 'Excerpce lišejníků – lišejníky'. Below this is a 'validace' field. A red circle highlights the 'Exportovat do excelu' button, with a tooltip that reads: 'Pro tento export platí následující omezení: • maximální počet exportovaných záznamů je omezen na 100000'.

Obrázek 18 V záložce mapy -> vyhledávání záznamů je možné vybrat buď výpis nalezených záznamů do prohlížeče, nebo exportovat do excelu. Pro přípravu výskytových dat pro trénování predikčního modelu je export záznamů a jejich selekce doporučeným postupem, pokud chceme využívat záznamy z DaLiBor.

Pokud v aplikaci zvolíme „Create new prediction task from DaLiBor records“, můžeme si vybrat jeden z nabízených druhů, který je v DaLiBor. Čísla v závorkách ukazují počet záznamů, který v databázi daný druh má, přičemž jsou v nabídce pouze druhy s více než pěti záznamy. To neznamena, že je dobrý nápad snažit se získat predikce pro druh s pěti záznamy pro celé území ČR, viz kapitola 3.6.3. Pro explorativní analýzu můžeme bez dalšího vybrat druh z DaLiBor a nechat si nekriticky spočítat model, který pak prozkoumáme. To ale není vhodný postup pro zacílení terénního výzkumu. Tam je potřeba zohlednit kvalitu záznamů (prostorovou chybu lokalizace či validitu determinace). To můžeme udělat buď přímo v DaLiBor, kdy můžeme po přihlášení záznamy jeden po druhém validovat pomocí nástrojů dostupných v uživatelském rozhraní DaLiBor, a do modelu pak budou vstupovat pouze validní záznamy. Pro účely terénního vyhledávání doporučujeme data cílového druhu po explorativní analýze validovat přímo v DaLiBor, nebo preferenčně exportovat do tabulkového editoru a ručně validovat a editovat lokálně i s ohledem na prostorovou chybu lokalizace.

Je na zvážení uživatele, jaká prostorová chyba lokalizace je pro cílový organismus ještě přijatelná, aby daná lokalizace stále reprezentovala niku daného druhu. Pro některé druhy s úzkou nikou, nebo ve velmi komplexním reliéfu, může být prostorová chyba 30 m již nepřijatelná, pro jiné druhy můžeme modelovat s daty s prostorovou chybou nad 500 m. Jak bylo řečeno v teoretickém úvodu, prostorové chybě souřadnic výskytu daného záznamu je třeba věnovat pozornost, a to mimo jiné kvůli tomu, že DaLiBor přebírá data z jiných zdrojů, kde jsou lokalizace vázané na polygon (např. zvláště chráněné území), ale pro potřeby DaLiBor jsou výskyty převedeny na centroid. Pak je prostorová chyba poloměr nejmenšího opsaného kruhu daného polygonu. Takový centroid může ležet na místě s environmentálními podmínkami dramaticky odlišnými, než jsou ty, kde se druh skutečně vyskytuje, viz obrázek 19 s příkladem MZCHU Vápenice.

Přes různé nástrahy má využití dat z DaLiBor dvě výhody: i) data tam jsou připravená k použití, ii) není třeba řešit formátování dat a zeměpisné projekční systémy, protože to si software načítá přímo z databáze.



Obrázek 19 V případech, kdy do DaLiBor vstupují výskytová data původně vázaná na polygon, převádějí se na centroid. Zde je schéma generování centroidu a kružnice pro výpočet prostorové chyby na příkladu rezervaci Vápenice v Železných horách. Prostorová chyba je pak rovna poloměru nejmenší opsané kružnice. Kvůli tomu pak do predikčního modelu mohou vstupovat environmentální podmínky z centroidu a nikoliv z místa skutečného výskytu daného druhu, kde mohou, ale nemusejí být zcela odlišné podmínky.

b. Výskytová data z CSV tabulky

Pokud není zájmový organismus mechorost nebo lišejník, ale i pokud je a uživatel chce mít kvalitu vstupních dat více pod kontrolou, je možné nahrát do aplikace výskytová data v podobě csv tabulky s přesně definovanou strukturou. Doporučujeme stáhnout si vzorovou tabulku a vzorová výskytová data nahradit vlastními. Tím bude minimalizováno riziko špatného importního formátu. Vzor struktury vstupní tabulky je možné stáhnout zde

>> <https://predikce.ibot.cas.cz/data/sampleData.csv> <<

Uživatel také musí vybrat, v jakém souřadnicovém referenčním systému jsou data importována. Aplikace podporuje import v několika nejběžnějších referenčních systémech používaných v ČR jako je například S-JTSK, ETRS, WGS84, S-42/83. Výchozí je WGS84 (EPSG:4326), nejběžnější pro data z GPS.

Create new prediction task

Email

Upload coordinates table

☒ In the table, true absence coordinates are provided.

EPSG

Spatial envelope typ

Buffer size [m]

Obrázek 20 Rozhraní online predikčního software pro modelování s daty z CSV tabulky.

V případě, kdy je zatržené tlačítko „In the table, true absence coordinates are provided.“, což je výchozí a preferovaný způsob predikce, aplikace používá prezence, absence výhradně ze vstupní tabulky a negeneruje žádné další pseudoabsence, i kdyby byla skutečná absence pouze jeden řádek. Pokud je v některém řádku chybějící hodnota, nebo data v nesprávném formátu, aplikace řádek ignoruje. Pokud jsou data v nesprávném formátu, aplikace vrátí chybu, výpočet predikce neproběhne.

Struktura importní CSV tabulky

- vstupní data musí být v souboru csv se třemi sloupci
- sloupce musí mít hlavičku (názvy) s hodnotami (x;y;pa)
 - x = zeměpisná délka (longitude)
 - y = zeměpisná šířka (latitude)
 - pa = prezence, absence
- oddělovač sloupců je středník (;)
- oddělovač desetín je tečka (.)
- hodnoty ve všech sloupcích musí být čísla
- ve sloupci pa mohou být pouze 0,1 (absence=0, prezence=1)
 - v případě, že má uživatel pouze prezenční data, pak budou ve sloupci pa samé jedničky

Create new prediction task

Email

Upload coordinates table

☐ In the table, true absence coordinates are provided.

Pseudo absence count

EPSG

Spatial envelope typ

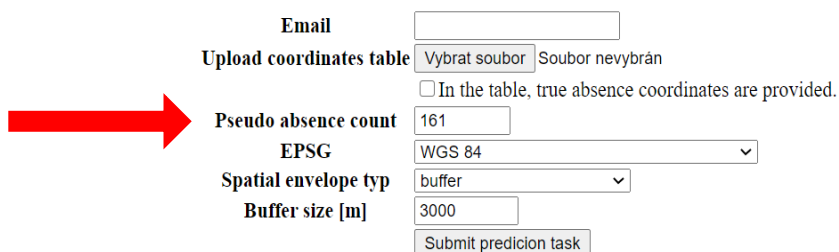
Buffer size [m]

Obrázek 21 Rozhraní online predikčního software pro modelování s daty z CSV tabulky, generování pseudo-absencí.

V případě, že uživatel neposkytne žádné pravé absence a v online rozhraní zruší zatržení „In the table, true absence coordinates are provided.“, objeví se pole „Pseudo absence count“, kde je přednastavená hodnota 161 (ad hoc číslo, pomůcka programátora). To indikuje, že software automaticky vygeneruje pětinašobek pseudoabsencí ku prezencím, stejně jako v případě modelování s DaLiBor. Pokud uživatel nastaví jakékoliv jiné číslo, aplikace vygeneruje právě tolik pseudoabsencí. Tato funkce je implementována s ohledem na možný významný vliv nastaveného počtu vygenerovaných pseudoabsencí na výsledné predikce. Viz kapitolu 3.6.4. této metodiky. Uživatel tak má větší kontrolu a může vygenerovat více modelů s různým počtem pseudoabsencí a sledovat stabilitu výsledků.

V případě predikce z DaLiBor a v případě, kdy uživatel zatrhne, že jeho importní csv neobsahuje žádné pravé absence, aplikace generuje pseudoabsence.

Create new prediction task



The form contains the following fields and controls:

- Email:
- Upload coordinates table:
- ☐ In the table, true absence coordinates are provided.
- Pseudo absence count: (highlighted by a red arrow)
- EPSG:
- Spatial envelope typ:
- Buffer size [m]:
-

Obrázek 22 Defaultně je nastaven počet pseudo absencí na desetinásobek prezencí. Toto defaultní nastavení je indikováno hodnotou 161.

Aplikace vygeneruje defaultně pětinašobek pseudo-absencí ku prezencím a rozmístí je náhodně do oblasti predikce tak, aby ležely alespoň 200 m od prezencí. Defaultní desetinásobek je v aplikaci indikován číslem 161. Uživatel může počet pseudo-absencí ručně definovat v případě modelování z CSV. Jakékoliv jiné číslo než 161. Pseudo-absence jsou vždy rozmístěny náhodně do oblasti predikce. Viz kapitolu 3.6.4.

Jak je zmíněno v teoretické části této metodiky, generování pseudo-absencí má vliv na výsledky SDM a je k dispozici mnoho způsobů a názorů na generování pseudo-absencí a jejich počet. Zdá se, že pro algoritmy s využitím Random Forest (hlavní algoritmus v této metodice) je optimální takový počet pseudo-absencí, jako je nízký několikanásobek prezencí a jejich náhodné rozmístění v prostoru (Liu et al. 2019). Oblast predikce může být definovaná buď vstupními daty, nebo je to celá ČR. Záleží na volbě uživatele. O výběru oblasti predikce se píše podrobněji níže.

4.2 Environmentální data

Environmentální data jsou součástí predikčního software a nejsou uživatelsky přístupná, zejména proto, že jejich použití je vázáno na výzkum Botanického ústavu, kdy je Botanický ústav zpravidla oprávněn data používat, ale není oprávněn je poskytovat třetím osobám. Proto je možné tato data využít uvnitř aplikace. Je to 12 vrstev popisujících topografické, klimatické a krajinné či vegetační charakteristiky. Jsou to rastry (gridy) ve formátu GeoTif s pokrytím ČR sjednocené na stejný geografický referenční systém (CRS), pokrytí a jednotnou velikost buněk.

CRS	+proj=utm +zone=33 +datum=WGS84 +units=m +no_defs EPSG: 32633
Jednotky	metry
Velikost buňky	10 x 10
Prostorový rozsah	284188.6, 790928.6, 5344845.4, 5695595.4
Datový typ	Float32
NoData	-9999
Formát	GTiff

Tabulka 2 Metadata environmentálních vrstev dostupných v predikční aplikaci a výsledných predikčních map.

Těžištěm environmentálních dat jsou parametry odvozené z digitálního modelu terénu. Pro jejich odvození byl využit volně dostupný digitální výškový model ALOSWorld3D 30 m DEM (AW3D30; verze 3.1). Pro potřeby predikčního online software jsme digitální výškový model přepočítali na velikost buňky 10 m metodou B-Spline Interpolation a zároveň jsme vrstvu transformovali z původního projekčního systému do UTM zone 33 N (EPSG: 32633). UTM jsme zvolili kvůli délkovým jednotkám v decimální soustavě (metr) na rozdíl od populárního WGS 84, kde jsou délky definované v soustavě hexadecimální (stupně), což by mohlo působit problémy při uživatelské definici velikosti obalových zón či rozsahu predikce.

Jméno vrstvy	Popis
CanopyHeightDMP_DMRcl	Model výšky vegetace. Rozdíl digitálního modelu povrchu a digitálního modelu reliéfu. Vrstva má rozlišení 10 m a je v systému UTM. Vrstva byla dále reklasifikována tak, že všechna vegetace nižší než 0.1 m dostala hodnotu 0 a vyšší než 80 m hodnotu právě 80.
ConvergenceIndex50m	Vrstva popisující hydrologické poměry dané buňky na základě buněk okolních v okolí 50 m. Zjednodušeně řečeno index popisuje, zda do konkrétní buňky voda z okolí spíše přitéká, nebo z ní spíše odtéká.
DiurnalAnisotropicHeating	Vrstva popisující potenciální příkon slunečního záření skrze výpočet nerovnoměrného zahřívání povrchu v závislosti na orientaci vůči světovým stranám. Oproti prosté orientaci jde o proměnnou, která není definována cirkulárně, a proto je zde vhodnější než orientace samotná.
elevation_10m	Nadmořská výška dané buňky.
ForestTreeClass	Kategoriální proměnná, klasifikované lesní dřeviny na základě satelitních snímků. ÚHÚL. Vrstva Lesní Dřeviny 2017. Obsahuje třídy: (Smrk, Nelesní půda, Buk, Borovice, Ostatní listnaté, Dub, Těžba, Směs, Kultura, Kleč)
Kves10m	Kategoriální proměnná, Konsolidovaná vrstva ekosystémů. AOPK 2020. Převedená do rastrové reprezentace. 36 kategorií krajinného pokryvu. Např: (Alpínské louky, Aluviální a vlhké louky, Bažina, močál Bučiny, Dopravní síť, Doubravy a dubohabřiny, Hospodářské lesy jehličnaté, Hospodářské lesy listnaté, Orná půda, Rašeliniště a prameniště, Rybníky a nádrže...)
MassBalanceIndex	Vrstva popisující terén z hlediska erozního potenciálu. Odlišuje místa s potenciálem pro odnášení a ukládání materiálu či místa s rovnováhou v potenciální sedimentaci a erozi.
TopographicPositionIndex250m	Relativní výška nad terénem v okolí 250 m
TopographicPositionIndex50-500m	Relativní výška nad terénem počítaná iterativně v deseti prstencích v okolí od 50 do 500 m.
TopoWetInd	Topografický vlhkostní index. Potenciální topografická vlhkost dané lokality vypočtená na základě velikosti povodí nad danou buňkou a sklonu svahu.
VectorRuggednessMeasure	Vrstva popisuje „vrásnitost“ či heterogenitu terénu v okolí daného bodu. Jedná se o dekompozici sklonu a orientace svahu do trojrozměrných vektorů a odtud výpočet velikosti těchto vektorů.
VertDistToDibavod	Vrstva aproximuje hloubku podzemní vody v daném místě za pomoci rozdílu topografického modelu z toků (DIBAVOD) a modelu reliéfu.

Tabulka 3 Seznam environmentálních vrstev použitých pro tvorbu predikčního modelu a výslednou mapu potenciálního výskytu habitatu.

Jedním z výstupů modelu jsou i diagramy distribuce hodnot (histogramy) environmentálních parametrů v oblasti predikce (požadové hodnoty) a v místech ověřeného výskytu druhů. Pro popis požadových hodnot environmentálních parametrů v oblasti predikce se náhodně generuje 10 x tolik náhodných bodů jako je vstupních výskytových dat (prezence + absence), pokud je desetinásobek výskytových dat menší než 500, pak je vygenerováno právě 500 náhodných požadových bodů.

Pomocí histogramů je možné dále analyzovat pokrytí environmentálního gradientu či sledovat environmentální preference cílového druhu v místech jeho skutečného výskytu v porovnání s distribucí environmentálních podmínek v celém území reprezentovaném minimálně pěti sty náhodnými body. Podrobněji viz kapitolu 3.6.3. sekci v.) této metodiky.

4.3 Oblast predikce

Pro oba typy predikčních úkolů, ať už jsou vstupní data z databáze DaLiBor nebo z tabulky CSV, má uživatel možnost částečně ovlivnit tvar oblasti výsledné predikce. Oblast predikce je vždy definovaná vstupními daty (prezence z DaLiBor nebo prezence + absence z CSV). Kolem těchto bodů se generuje polygon, který určuje, jaké pokrytí bude mít výsledná predikční mapa. V případě, že se generují pseudo-absence, pak má volba oblasti predikce vliv i na výsledky modelu. Pseudo-absence se generují právě v oblasti predikce, stejně jako pozadové body pro reprezentaci environmentálních proměnných. Viz kapitolu 3.6.3

Create new prediction task

Email

Upload coordinates table
☐ In the table, true absence coordinates are provided.

Pseudo absence count

EPSG

Spatial envelope typ

Buffer size [m]

Taxon from Dalibor

Email

Spatial envelope typ

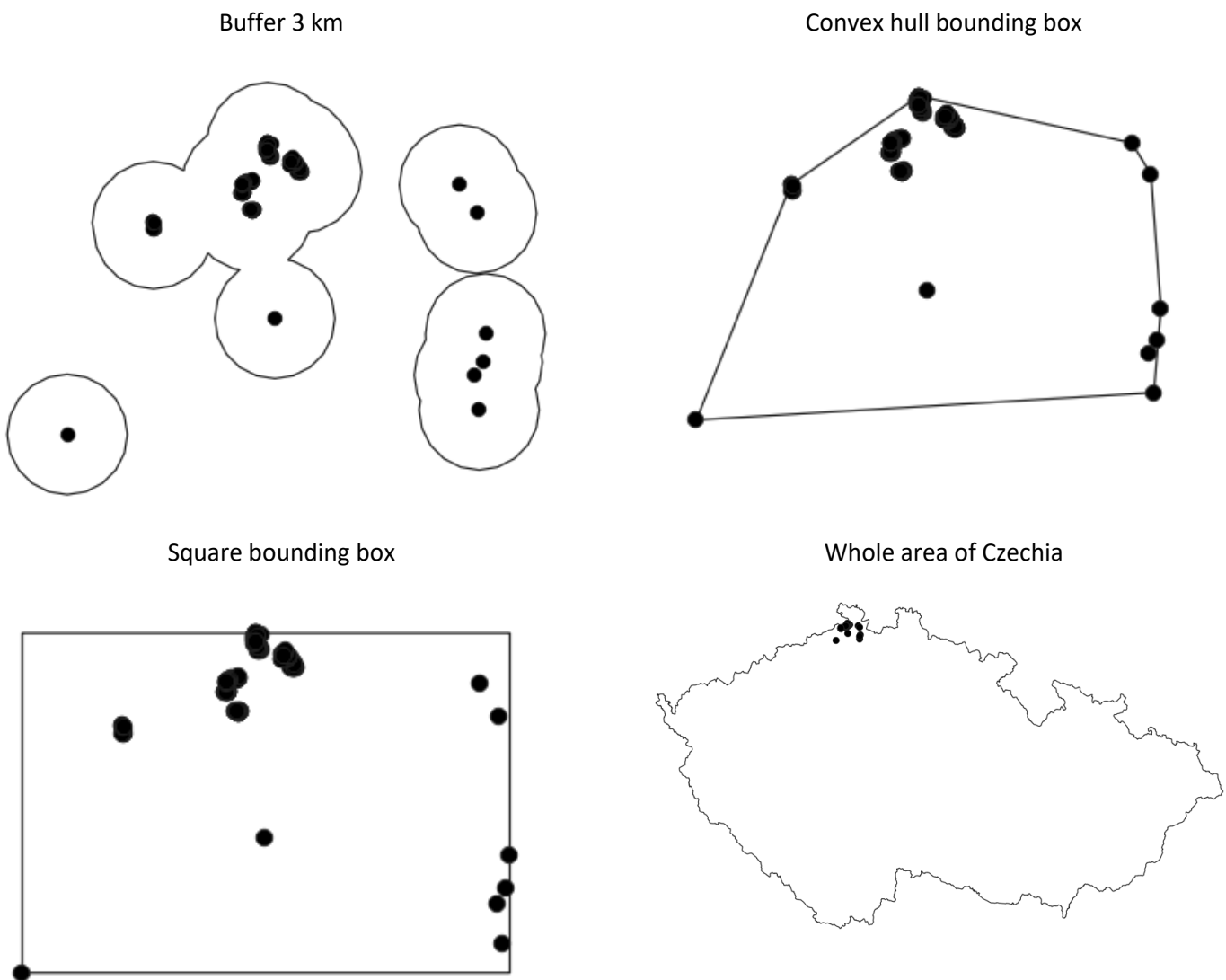
Buffer size [m]

Obrázek 23 Definice oblasti predikce v obou případech predikce (DaLiBor, CSV)

Defaultně je nastavena oblast predikce buffer, 3 km. To znamená, že výsledná mapa pokrývá okolí vstupních výskytových dat v okruhu 3 km. V tomto okruhu se generují také pseudo-absence a pozadové hodnoty.

K dispozici jsou tyto volby:

Buffer	Obalová zóna definované velikosti. Velikost se udává v metrech.
Convex hull bounding box	Nejmenší celistvý nepravidelný polygon opisující všechny body.
Square bounding box	Nejmenší čtyřúhelníkový polygon opisující všechny body
Whole area of Czechia	Celá ČR bez ohledu na distribuci bodů



Obrázek 24 Příklady konstrukce oblasti predikce pro všechny dostupné volby v predikčním software. Stejná data, pro ČR rozdílné měřítko.

Doporučujeme používat oblasti predikce definované vstupními body (buffer, bounding box), zejména v případě, že nás zajímá konkrétní region a vstupní body jsou prostorově klastrované, jak je to na obrázku 24. Pokud je oblast predikce nastavena regionálně, i vygenerované pseudo-absence a pozadové hodnoty tento region lépe charakterizují v porovnání s případem, kdy je oblast predikce celá ČR. Pokud uživatel poskytne i pravé absence, je výsledná predikce spolehlivější a je zde větší prostor pro predikci na území celé ČR.

4.4 Interakce uživatele s predikční aplikací

Pro modelování s použitím našeho predikčního software (<https://predikce.ibot.cas.cz/>) zadává uživatel také svůj email. Ten je využit výhradně pro potřeby automatizované interakce software a uživatele. Modelování, ale zejména predikce, může být výpočetně velice náročná v závislosti na požadované oblasti predikce a počtu trénovacích dat. Predikce pro celou ČR trvá například cca 8 hodin, a to má ke konci roku 2020 výpočetní server 16 CPU a 100 GB RAM. Proto je zde emailová komunikace, kdy server odešle uživateli 3 informační emaily. Emaily jsou odesílány z adresy: predikce@post.cz

1. Při vytvoření predikce

Hello,
the task was created; you can find it on this [link](#).
You will be informed about progress during the computation by email again.

2. Když je spočítaný predikční model a je připravený grafický report modelu (na základě reportu modelu vyžaduje aplikace rozhodnutí uživatele, na základě kterého algoritmu se má vypočítat predikce (Random Forest (RF) nebo Generalised Linear Model (GLM), nebo oba).

Hello,
the model for your task was successfully computed; please follow this [link](#) and decide, which type of prediction you want to compute.

Hello,
the prediction based on formerly computed model for your task was successfully finished; please follow this [link](#) and enjoy your results.

3. Když je výsledná mapa připravená ke stažení.

Link v emailech vede na výpis informací o daném úkolu, který se s postupem svého dokončení dynamicky rozšiřuje o další informace, volby a odkazy ke stažení konkrétních výstupů či výsledků.

1.	Nový task <pre> id: prediction-5f8571501f189 email: man.matej@gmail.com inputData: CSV taskCreated: 2020-10-13 11:20:16 taskModelStarted: 2020-10-13 09:20:40 includesAbsence: 1 pseudoabsenceCount: 161 epsg: 4326 areaType: buffer areaBuffer: 3000 </pre>	Možnost stažení vstupních dat, nebo odkaz na výpis z DaLiBor.
2.	Nový task <pre> id: prediction-5f8571501f189 email: man.matej@gmail.com inputData: CSV taskCreated: 2020-10-13 11:20:16 taskModelStarted: 2020-10-13 09:20:40 taskModelFinished: 2020-10-13 09:22:32 Model: link includesAbsence: 1 pseudoabsenceCount: 161 epsg: 4326 areaType: buffer areaBuffer: 3000 </pre> Pick type of prediction according to model results only GLM Confirm	Zobrazení grafického reportu, kde je evaluace modelu a výběr typu predikční mapy na jeho základě.
3.	Nový task <pre> id: prediction-5f7f186850fec email: man.matej@gmail.com inputData: CSV taskCreated: 2020-10-08 15:47:20 taskModelStarted: 2020-10-08 13:47:40 taskModelFinished: 2020-10-08 13:49:30 Model: link taskPredictionStarted: 2020-10-08 13:51:10 taskPredictionFinished: 2020-10-08 13:52:14 Prediction: link includesAbsence: 1 pseudoabsenceCount: 161 epsg: 4326 areaType: buffer areaBuffer: 3000 predictionType: both </pre>	Stažení predikce jako GIS GeoTif vrstvu.

Tabulka 4 Postupný dynamický vývoj online karty predikčního úkolu, na kterou se uživatel dostane pomocí odkazu v emailu vygenerovaného predikční aplikací.

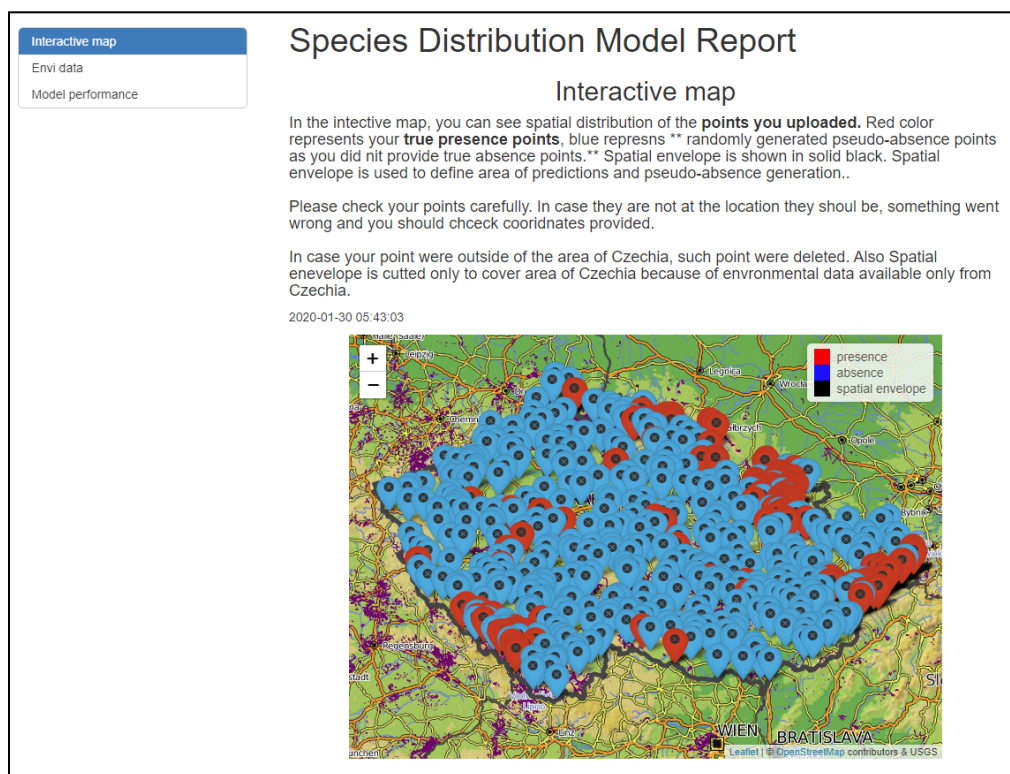
4.5 Model a jeho výstup

Jádro predikčního software, na kterém tato metodika staví, je napsán v prostředí R (R Core Team 2016) s využitím balíčku caret (Kuhn 2008). Grafický report modelů je uživateli zpřístupněn prostřednictvím html generovaného v balíčku rmarkdown (Xie et al. 2018). V software jsou implementované 2 základní modely GLM a RF. GLM ze základního balíčku rstats a RF z balíčku ranger (Wright & Ziegler 2017). Protože je jádro predikční aplikace napsané v prostředí R, je možné R script uživatelsky modifikovat a použít lokálně, mimo predikční aplikaci v prostředí R, což dává uživateli možnost větší kontroly nad nastavením modelu a implementaci dalších modelovacích algoritmů či funkcionalit. Zde je však nezbytné, aby uživatel dodal vlastní ekologická data, která není možné v současné době zpřístupnit kvůli licenčním podmínkám poskytovatelů. Kód je veřejně přístupný přes repositář GitHub https://github.com/manmatej/DaLiBor_predict2019

Standartním výstupem modelu je html grafický report, ke kterému se uživatel dostane prostřednictvím odkazu na kartě predikčního úkolu z emailu, který aplikace uživateli zašle. Na základě informací v reportu může uživatel rozhodnout, zda si přeje vypočítat mapu potenciální vhodnosti habitatu daného druhu a který ze dvou nabízených modelů se k tomu má využít. Uživatel je vrácen na grafické rozhraní, kde vybírá a potvrzuje zadání výpočtu (Tabulka 4, řádek 2).

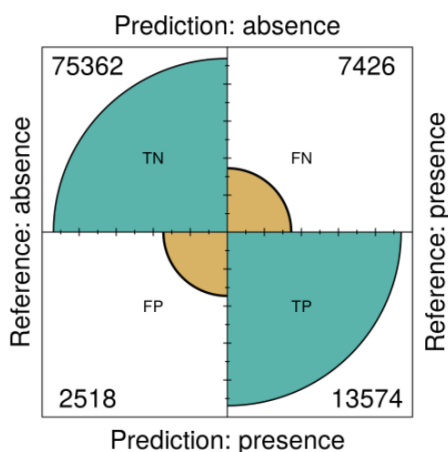
Součástí reportu je interaktivní mapa (obrázek 25), která ukazuje distribuci prezencí a absencí případně pseudo-absencí, zobrazena je také hranice oblasti predikce. Tato mapa slouží ke kontrole

- Zda leží vstupní výskytové body na očekávaném místě
- Zda jsou korektně odlišeny prezence a absence
- Zda jsou korektně vygenerované pseudo-absence a zda je uživatel spokojen s jejich počtem a prostorovou distribucí
- Zda je uživatel spokojen s oblastí predikce, kterou bude pokrývat výsledná mapa



Obrázek 25 Úvodní část grafického reportu modelu s interaktivní mapou pro kontrolu prostorové distribuce prezencí, absencí či pseudo-absencí.

Jednou z nejdůležitějších částí grafického reportu je sekce Model Performance, která poskytuje několik ukazatelů k posouzení kvality (úspěšnosti) modelu odlišit lokality s výskytem potenciálně vhodného habitatu od těch méně vhodných až zcela nevhodných. V této sekci reportu se nachází vždy párově jednotlivé metriky pro generalizovaný lineární model (GLM) a random forest (RF). Ukazatele spolehlivosti modelů se mohou v rámci páru modelů výrazně odlišovat. Zpravidla jsou ukazatele vychýlené v prospěch algoritmu random forest, který dokáže lépe predikovat nelineární vztahy, interakce i kategoriální proměnné.



Obrázek 26 Confusion matrix - matice ukazující poměr predikcí, kdy model na validačním datasetu správně predikoval prezenci (TP = true positive) a absenci (TN = true negative) ku situacím, kdy model predikoval falešně (FP = false positive, false negative). Sloupce rozlišují stavy pozorovaných hodnot, řádky rozlišují stavy predikovaných hodnot..

Jedním z vodítek, které nám napoví, jak úspěšný byl model v odhalování lokalit s potenciálním výskytem, je Confusion matrix - neboli matice záměn (**Chyba! Nenalezen zdroj odkazů.**). Ta nám ukazuje, v kolika případech model predikoval správně a chybně prezenci či absenci ve validačním datasetu. Čím větší je proporce TN (True Negative) a TP (True Positive) ku FN (False Negtive) a FP (False Positive), tím úspěšnější model byl v předpovídání prezence či absence potenciální vhodnosti habitatu daného druhu měřeno na validačním datasetu. Jinými slovy čím větší část koláče je zelená ku červené, tím lépe a tím větší šance, že model bude schopen efektivně odhalit lokality potenciálně vhodné pro výskyt cílového druhu.

Dalšími ukazateli úspěšnosti modelu, které grafický report přináší, jsou:

- **Sensitivita:** pravděpodobnost že prezence je klasifikována správně
- **Specificita:** pravděpodobnost že absence je klasifikována správně
- **Prevalence:** pouze popisný ukazatel podílu prezencí ve vstupním datasetu. Neříká nic o kvalitě modelu, ale je informativní při kritickém hodnocení ostatních ukazatelů. Prevalence je podíl pozorovaných prezencí druhu ku všem pozorováním. V případě generování pseudo-absencí je prevalence nastavena defaultně jako $\text{prevalence} = \frac{\text{počet prezencí}}{\text{počet prezencí} * 5 + \text{počet prezencí}}$
- **Kappa (Cohen's Kappa):** ukazatel spolehlivosti modelu korigovaný na výsledek, který bychom dostali pouze dílem náhody, tedy o kolik je model lepší, či horší než náhodná distribuce. Hodnoty kappa se pohybují v rozmezí od -1 do +1, kde +1 značí velice spolehlivý model, hodnoty kolem 0 ukazují, že výsledky predikčního modelu jsou blízké náhodě. Hodnoty kolem 0 a menší značí málo spolehlivý model.
- **True Skill Statistic (TSS):** Ukazatel spolehlivosti modelu, který odečte počet správně klasifikovaných případů způsobených pouze dílem náhody od těch skutečně správně klasifikovaných a tuto hodnotu porovná s hypoteticky nejlepším výsledkem spolehlivosti modelu. Výpočetně je $\text{TSS} = \text{Sensitivita} + \text{Specificita} - 1$. Podobně jako Kappa TSS nabývá hodnot od -1 do +1, kde +1 značí velice spolehlivý model, hodnoty kolem 0 a nižší ukazují, že výsledky predikčního modelu jsou blízké náhodě. Na rozdíl od Kappa, TSS není ovlivněna prevalencí ani velikostí testovacího datasetu.

Podrobněji se věnuje validačním metrikám kapitola 3.6.6 této metodiky nebo práce (Allouche et al. 2006; Liu et al. 2019)

Pokud má model vysoký podíl nesprávně klasifikovaných prezencí a absencí k těm správně klasifikovaným (confusion matrix) a ukazatele spolehlivosti modelu Kappa a TSS nabývají nízkých hodnot (blízkých nule či záporných), pak výsledný model není spolehlivý.

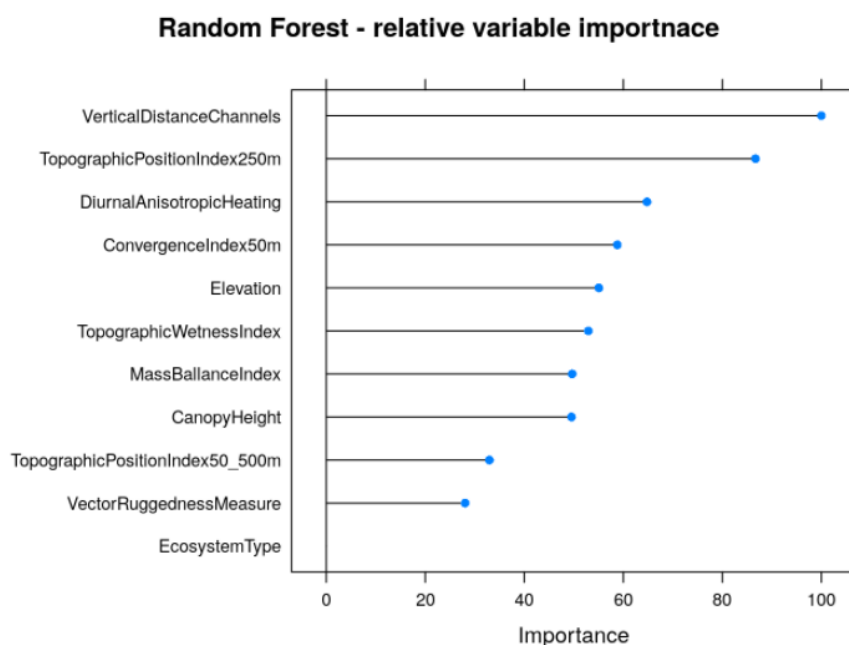
Nespolehlivý model můžeme dostat z mnoha důvodů, jak je podrobněji popsáno v teoretickém úvodu této metodiky. Někdy je možné spolehlivost modelu zvýšit, například zvýšením počtu trénovacích dat, nebo jejich filtrací, například odstraněním záznamů s nejvyšší prostorovou chybou lokalizace, odstraněním či naopak přidáním historických údajů o výskytu. Proto je dobré v případě nespolehlivého modelu zkusit jeho stabilitu právě pomocí filtrace vstupních dat.

Pokud jsou vstupní data důvěryhodná, recentní a přesně lokalizovaná, a model není spolehlivý, pak není vhodné jej používat ke generování predikční mapy a vyhledávání lokalit potenciálního výskytu. Použití takového podkladu pro vyhledávání nových lokalit výskytu by mohlo mít stejný efekt jako náhodná návštěva míst v oblasti predikce. Nespolehlivý model můžeme totiž dostat také z důvodu nedostatečnosti vstupních environmentálních dat nebo jejich nevhodnému rozlišení. Naše predikční aplikace například ke konci roku 2020 ještě neobsahuje geologická ani pedologická data, což může být pro mnohé druhy mechorostů a zejména lišejníků značně limitující. Podobně například údaje o výskytu dřevin, které mohou jako substrát výrazně ovlivňovat výskyt epifytických druhů.

Na základě grafického reportu se pak uživatel rozhodne zda:

- i) zopakuje proces tvorby modelu s obohacenými vstupními daty,
- ii) nechá si vygenerovat predikci za základě GLM
- iii) nechá si vygenerovat predikci za základě RF
- iv) nechá si vygenerovat obě predikce

Poslední informací, kterou grafický report přináší, je relativní důležitost jednotlivých proměnných (Obrázek 27), která nám může napovědět, jaké vstupní environmentální proměnné determinují nejsilněji modelovanou niku daného organismu. Zde je však třeba mít na paměti, že je to informace neúplná, protože vstupní sada environmentálních parametrů nemusí obsahovat ty skutečně relevantní proměnné, jak je diskutováno výše v této kapitole. Dále je třeba mít stále na paměti úspěšnost, s jakou se modelu podařilo predikovat skutečné prezence, což odráží například Kappa a TSS. I pokud je model zcela neúspěšný a jeho predikce se rovnají náhodné distribuci, stejně vygeneruje relativní pořadí důležitosti environmentálních parametrů, které ale v takovém případě nebude biologicky vůbec relevantní. Dalším omezením důvěryhodnosti pořadí důležitosti environmentálních proměnných determinujících distribuci druhu je počet vstupních dat. V případě malého počtu vstupních trénovacích prezencí, například méně než 70 bude relativní pořadí environmentálních proměnných nedůvěryhodné (Smith & Santos 2020).



Obrázek 27 Příklad relativního pořadí důležitosti jednotlivých environmentálních proměnných z grafického reportu modelu.

4.6 Mapa potenciální vhodnosti habitatu

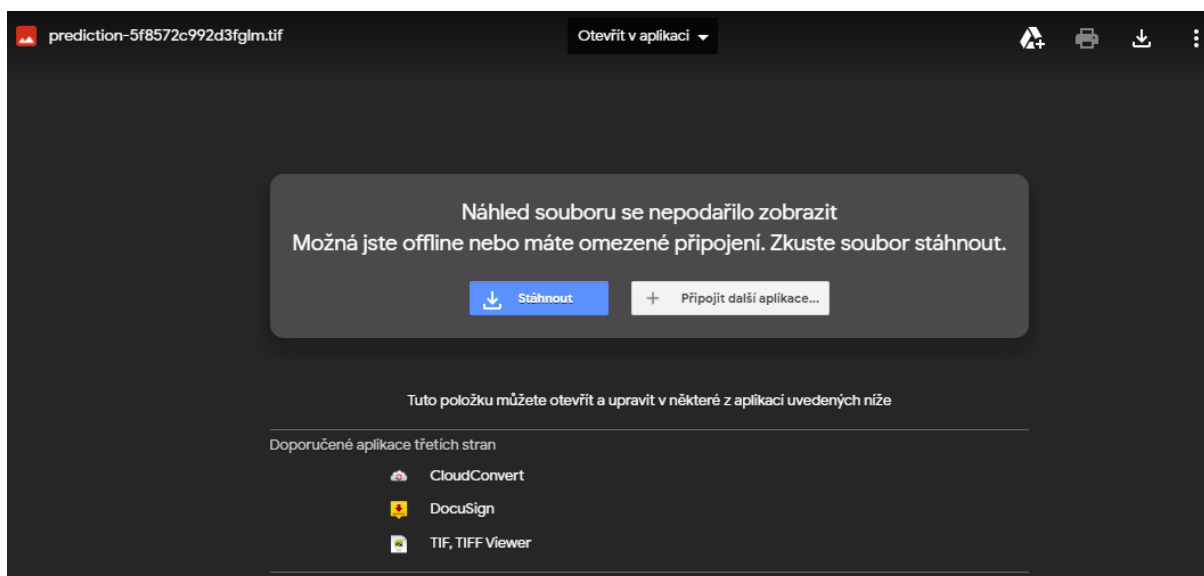
Poté co dá uživatel predikční aplikaci pokyn k vygenerování mapy predikce (Tabulka 4, řádek 2), server zařadí úkol do fronty a v závislosti na vytíženosti a výpočetní náročnosti úkolu vrátí email s odkazem na stažení výsledku predikce (Obrázek 28). To může trvat několik minut, ale v případě celé ČR také několik hodin. V závislosti na tom, zda si uživatel nechal spočítat mapu potenciálního výskytu na podkladu pouze jednoho či obou algoritmů, vrátí se mu odkaz na stažení jedné či dvou map. Výsledné GIS GeoTif mapy jsou zpřístupněny prostřednictvím služby Google Drive (Obrázek 28), kde si



Obrázek 28 Rozhraní pro přístup k výsledným predikčním mapám.

je uživatel stáhne.

Predikční mapy zobrazují pravděpodobnost výskytu vhodného habitatu na škále od jedné (největší pravděpodobnost) do nuly (nejmenší pravděpodobnost). Zde je však nutné mít na paměti, že relativní pravděpodobnost výskytu v mapě je relativní ku celkové úspěšnosti predikce modelu, nikoli absolutní pravděpodobnost.



Obrázek 29 Google Drive rozhraní pro získání mapy.

Pokud nám například grafický výstup modelu říká, že $Kappa = 0.012$ a $TSS = 0.008$ ukazuje to na model s velice slabou predikční schopností. V případě, že si uživatel i přesto nechá spočítat predikční mapu potenciálního výskytu vhodných habitatů, hodnoty v mapě budou přeskálované přibližně od jedné k nule a budou reprezentovat tento málo úspěšný model. V mapě pak můžeme vidět buňky s hodnotami relativní pravděpodobnosti kolem 0.9, je to však hodnota relativní, vztažená k málo úspěšnému modelu, a nikoli hodnota absolutní. Taková místa v mapě pak nemusí reálně nasvědčovat nic o vysoké pravděpodobnosti výskytu habitatu cílového druhu.

4.7 GIS a terénní práce

Finálním výstupem predikčního software je georeferencovaná mapa, rastr ve formátu GeoTiff. Rastr je možné vizualizovat v grafických editorech, ale doporučený způsob je další zpracování v geografickém informačním software GIS. Metadata k výsledným mapovým podkladům pro vyhledávání nových lokalit jsou uvedena v tabulce 5.

CRS	+proj=utm +zone=33 +datum=WGS84 +units=m +no_defs EPSG: 32633 (WGS 84 / UTM zone 33N)
Jednotky	metry
Velikost buňky	10 x 10
Prostorový rozsah	Záleží na volbě uživatele
Datový typ	Float32
NoData	-9999
Formát	GeoTiff
Minimum	0
Maximum	1

Tabulka 5 Metadata výsledných GeoTiff predikčních map

Predikční aplikace vrací mapový výstup v rastrovém formátu této podoby záměrně proto, aby bylo možné výstupy dále zpracovat do formy, která je koncovému uživateli příjemná a na kterou je zvyklý i s ohledem na technické vybavení a zaběhnuté postupy. Výstupy je tak možné kombinovat navzájem (viz kapitola 4.7.3) nebo kombinovat s jinými GIS vrstvami, které má koncový uživatel k dispozici.

Pro skutečně efektivní využití výstupu z predikční aplikace a jeho praktickou aplikaci pro zacílení terénního průzkumu doporučujeme jeho další zpracování v Geografickém Informačním Software (GIS). K tomu je možné využít komerčních nástrojů, jako například rozšířených produktů Esri, ale také volně dostupného software, jako například QGIS⁵, SAGA GIS⁶, GRASS GIS⁷ či rozšiřujících balíčků sf⁸ a raster⁹ pro platformu R¹⁰. V případě dalšího zpracování dat ve skriptovací aplikaci R je velice užitečným materiálem volně dostupná kniha *Geocomputation with R* autorů Robina Lovelace, Jakuba Nowosada a Jannese Muenchowa <https://geocompr.robinlovelace.net/>

Podrobněji se dále budeme věnovat variantě zpracování v QGIS, který je z výše jmenovaných volně dostupných nástrojů. Pro QGIS jsou také na internetu volně dostupné kvalitní výukové materiály i v češtině. Mnohé České univerzity také QGIS vyučují. Níže popsáný postup je uváděn pro QGIS verze 3.14, což je ke konci roku 2020 poslední dostupnou verzí. Postup je ale koncepčně analogický pro jakýkoliv zvolený software.

⁵ <https://qgis.org/en/site/>

⁶ <http://www.saga-gis.org/en/index.html>

⁷ <https://grass.osgeo.org/>

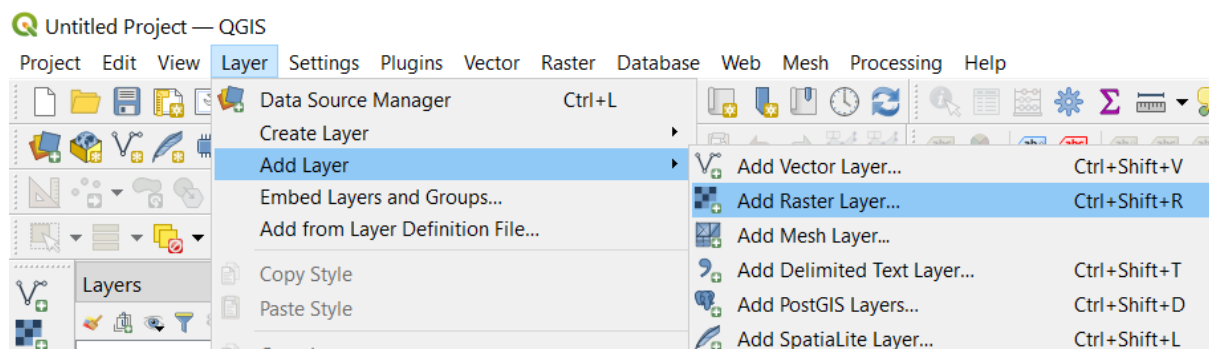
⁸ <https://r-spatial.github.io/sf/>

⁹ <https://r-spatial.org/raster/>

¹⁰ <https://www.r-project.org/>

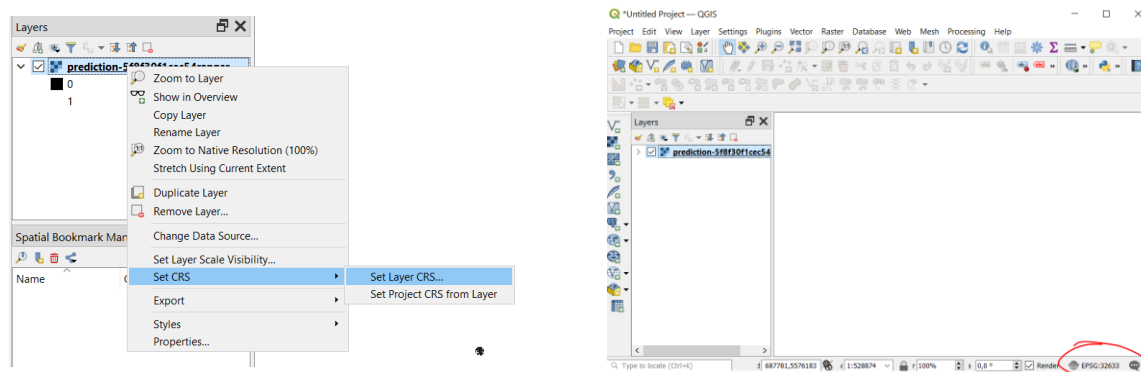
4.7.1 Vizualizace

Nejprve je třeba výstup modelování (predikční mapu) nahrát do zvoleného software. V případě QGIS v záložce **Layer > Add Layer > Add Raster Layer**. Obrázek 30. Pak stačí vybrat staženou mapu, která bude mít název dle ID predikčního úkolu s příznakem zvoleného modelu, na základě kterého byla vypočtena. Například: *prediction-5f8f30f1cec54ranger.tif* či *prediction-5f8f30f1cec54glm.tif* pro Random Forest či GLM.



Obrázek 30 Postup zobrazení predikční mapy v QGIS.

Pak je třeba nastavit správný souřadnicový referenční systém. Jak je uvedeno v tabulce 5, predikční mapa má referenční systém WGS 84 / UTM zone 33N (EPSG: 32633). GIS by měl tento systém sám rozeznat a načtené vrstvě přiřadit, ale je třeba to zkontrolovat. Chybné přiřazení referenčního systému povede k chybné lokalizaci celé mapy, a tak i k chybnému vytipování míst k vyhledávání cílových druhů. V QGIS je dobré mít nastavený shodný referenční systém u predikční mapy i mapového okna aplikace. Obrázek 31. To zaručuje rychlejší a přesnější práci s daty.

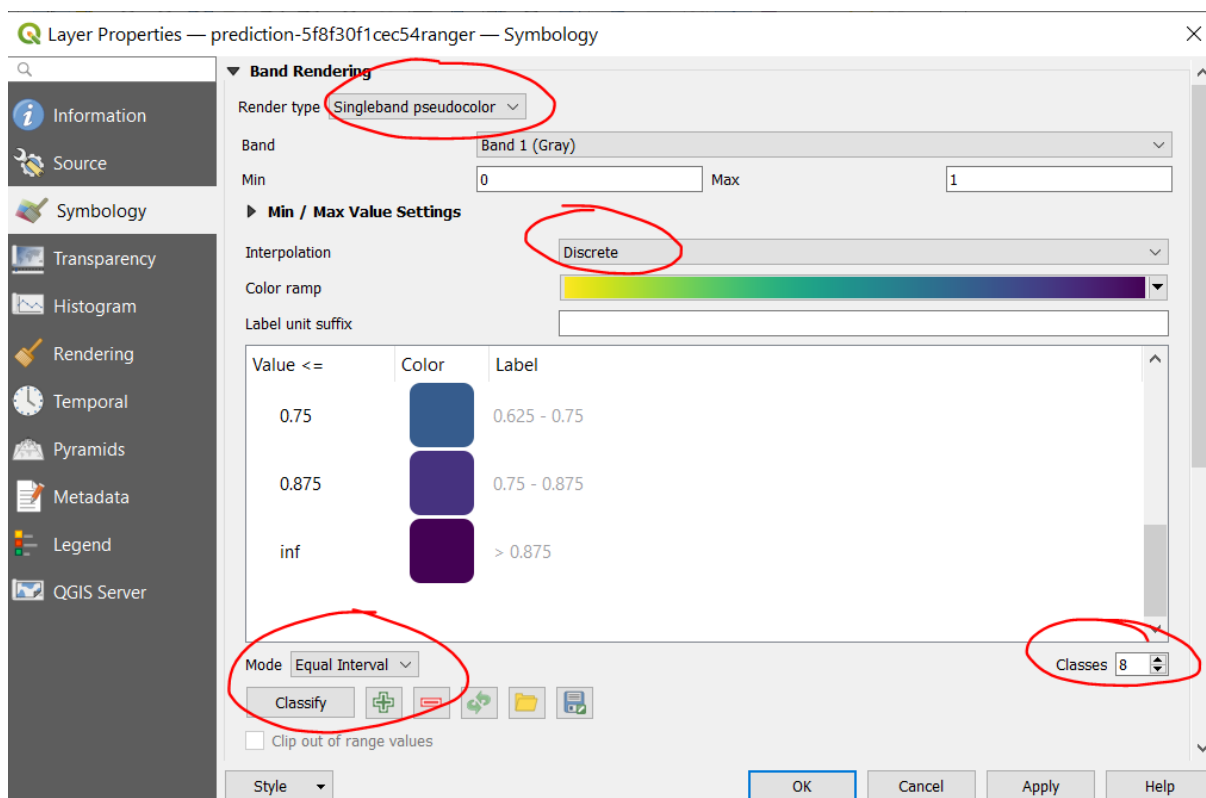


Obrázek 31 Nastavení (kontrola) referenčního systému v aplikaci QGIS. Vlevo pro mapu, vpravo pro celé mapové okno.

QGIS zobrazuje ve výchozím nastavení všechny přidané rastry na škále stupňů šedi, kdy černá zpravidla reprezentuje nízké hodnoty a bílá vysoké. To není příliš vhodné, protože plná bílá značí zároveň místa, kde nejsou žádná data. Proto je vhodné minimálně barevnou škálu invertovat, nebo ještě lépe nastavit tak, abychom dobře odlišili místa s vysokou pravděpodobností vhodného habitatu. To můžeme udělat, pokud na vrstvu klikneme pravým tlačítkem myši a zvolíme **Properties > Symbology**.

Jak je zachyceno na Obrázek 32, doporučujeme použít **Render Type = Singleband pseudocolor**, což umožňuje nastavit jinou barevnou škálu než stupně šedi. **Mode = Equal Interval** umožňuje nastavit počet tříd (**Classes**) do kolika má aplikace raster klasifikovat, kde je možné podle potřeby přidávat či ubírat. Doporučujeme použít **Interpolation = Discrete**. Tato volba zajistí, že na mapě skutečně uvidíme barvy, které odpovídají nastavené legendě, což je v tomto případě důležité. Předem nastavená hodnota *linear* zde není vhodná, protože výsledná mapa je sice na pohled líbivější, ale informačně je zavádějící. Symbolem plus a minus pod tabulkou legendy lze upravovat počet tříd a mezních hodnot. **Legenda se zpravidla objeví až po stisknutí tlačítka Classify**.

Pro lepší orientaci v geografickém prostoru je vhodné použít podkladovou mapu, a to buď s využitím populárního pluginu Quickmap Services¹¹ nebo přidáním vhodné WMS například základní mapy (ZM10 – ZM200) či Ortofoto poskytované Českým úřadem zeměměřickým a katastrálním¹². Příklad takové vizualizace s podkladovou mapou WMS ZM 25 je na Obrázek 33.



Obrázek 32 Doporučené nastavení vizualizace pro predikční mapy v aplikaci QGIS.

Příprava mapy v QGIS pro tisk se provádí pomocí nástroje Print Layout. Ten je dostupný v záložce **Project > Print layout**. Vyžaduje po uživateli jméno nového Layout a pak se otevře samostatné okno pro tvorbu mapy, které umožňuje vkládání legendy, měřítka, atributů, poznámek či přehledových map. Více o tvorbě mapy pro tisk například na informačně nabitém, českém webu kolektivu Gismentors:

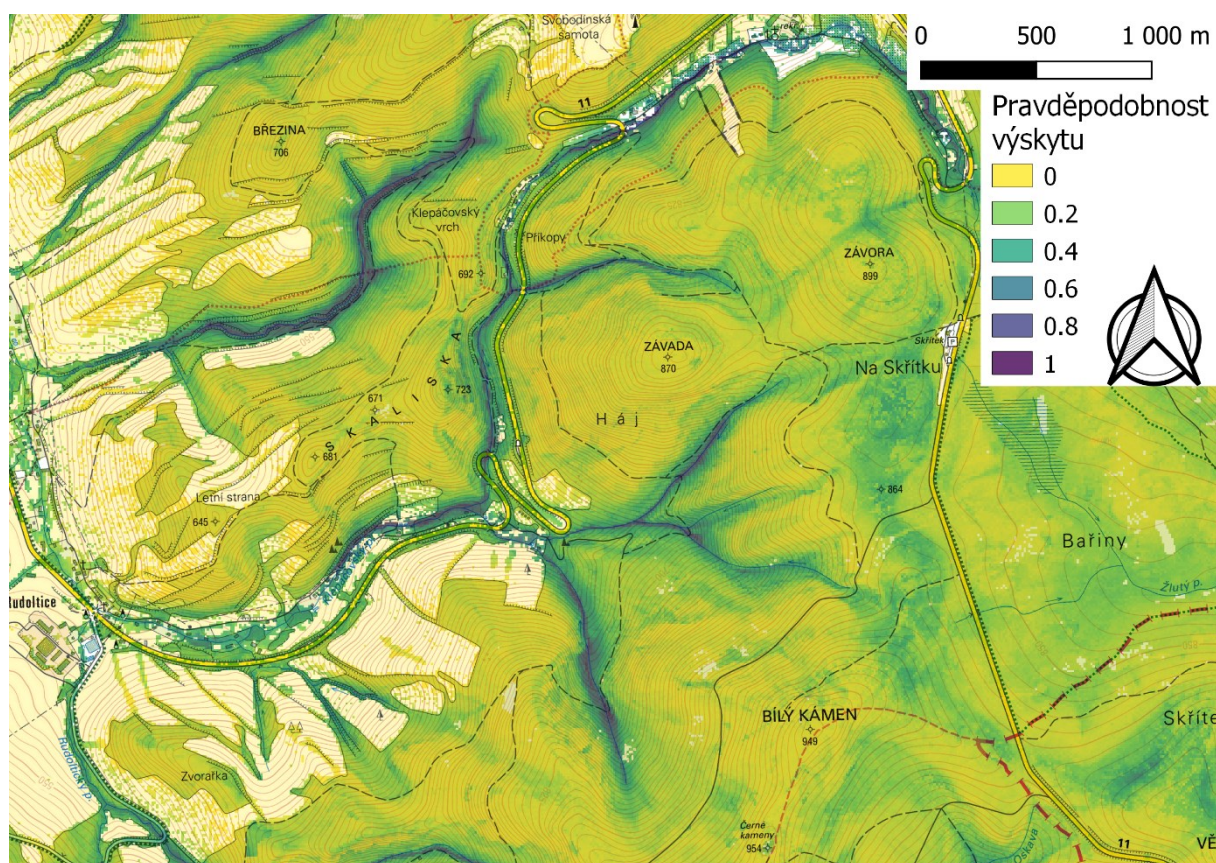
https://training.gismentors.eu/qgis-zacatecnik/mapovy_vystup/tvurce_map_vystupu.html#

¹¹ <https://qms.nextgis.com/about>

¹² [https://geoportal.cuzk.cz/\(S\(eclzpmgpt5jpm5gs3vk1lykw\)\)/Default.aspx?mode=TextMeta&side=wms.verejne&text=WMS.verejne.uvod&head_tab=sekce-03-gp&menu=311](https://geoportal.cuzk.cz/(S(eclzpmgpt5jpm5gs3vk1lykw))/Default.aspx?mode=TextMeta&side=wms.verejne&text=WMS.verejne.uvod&head_tab=sekce-03-gp&menu=311)

4.7.2 Plánování terénního průzkumu

Pro potřeby terénního průzkumu je možné si vystačit s prostou vizualizací a tiskem map či jejich exportem do chytrého zařízení. Samotný QGIS, pokud je nainstalovaný na zařízení s interní či externí GPS, umožňuje zobrazit v reálném čase pozici zařízení v mapovém okně, což může být přímo použitelné pro tablet či notebook v terénu. Umožňuje to panel GPS Information dostupný z **View > Panels > GPS Information**¹³. Rastrové mapy je také možné zobrazovat v různých aplikacích pro chytré telefony a tablety. Dobré zkušenosti máme například se zdarma dostupnou aplikací Locus Map Free¹⁴, je možné přímo nahrávat rastrová data v podporovaném formátu¹⁵. Ideální je například formát MBT (MBTiles), do kterého můžeme rastrová data jednoduše exportovat s pomocí QGIS¹⁶. Tím se vyhneme další práci v GIS s převodem rastrových hodnot na body. Při exportu rastru pro externí aplikace je potřeba změnit CRS, nastavit jej na WGS 84 (EPSG 4326), což je podporovaný referenční systém externích zařízení a aplikací.



Obrázek 33 Vizualizace pravděpodobnosti vhodného habitatu druhu *Buxbaumia viridis* nedaleko Velkých Losin. Podkladová mapa: WMS ZM25, ČÚZK. Výstup z QGIS, barvená škála viridis, renderování vrstev: multiply.

¹³ https://docs.qgis.org/3.10/en/docs/user_manual/working_with_gps/live_GPS_tracking.html?highlight=live%20tracking

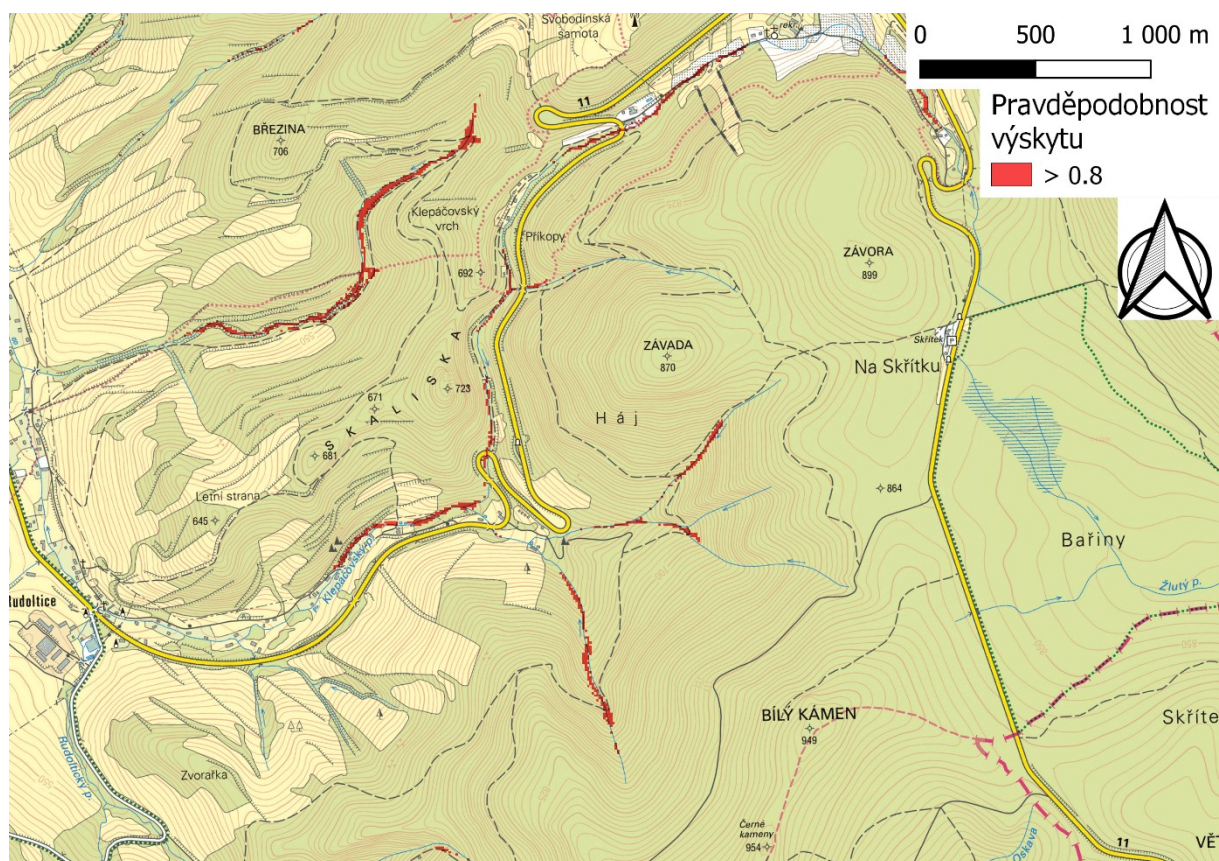
¹⁴ <https://www.locusmap.eu/cz/>

¹⁵ https://docs.locusmap.eu/doku.php?id=manual%3Auser_guide%3Amaps_external&do=

¹⁶ <https://www.fulcrumapp.com/blog/mbtiles-from-qgis>

Oproti prostému tisku či exportu rastrových map doporučujeme metodicky robustnější, technicky však náročnější postup přípravy elektronických podkladů využitelných v terénních zařízeních (GPS, tablet, chytrý telefon). Nejuniverzálnější forma akceptovaná většinou zařízení jsou bodová data. Jejich příprava vyžaduje jistou energii vloženou do práce s GIS, nicméně dle našich zkušeností se to vyplatí, zejména v případě, kdy nové lokality bude vyhledávat jiná osoba, než která připravovala podklady. Následuje několik kroků, jak odvodit z mapy pravděpodobnosti vhodnosti habitatu vektory – body pro použití v terénním zařízení. Níže popsany postup nemusí být jediný možný, ke stejnému nebo podobnému cíli může vést více cest. Prezентujeme zde dle našeho názoru výpočetně efektivní, a datově optimalizovaný postup.

Nejprve je vhodné vytipovat prahovou hodnotu pravděpodobnosti, která je pro účely vyhledávání nových lokalit relevantní. Na základě této prahové hodnoty pak vyfiltrovat buňky rastru pouze nad touto hodnotou a ty pak převést na body nebo polygony. Prahová hodnota pravděpodobnosti bude blízká jedné, ale její konkrétní stanovení závisí na celkové kvalitě modelu (Kappa, TSS viz kapitola 3.6.6 a 4.5 této metodiky), ale také na závěrech z vizuálního hodnocení mapového výstupu. Obrázek 34 ilustruje příklad vizualizace vytipovaných lokalit na základě prahové hodnoty relativní pravděpodobnosti vyšší než 0.8 (Model RF; Kappa = 0.67, TSS = 0.61) pro druh *Buxbaumia viridis*.



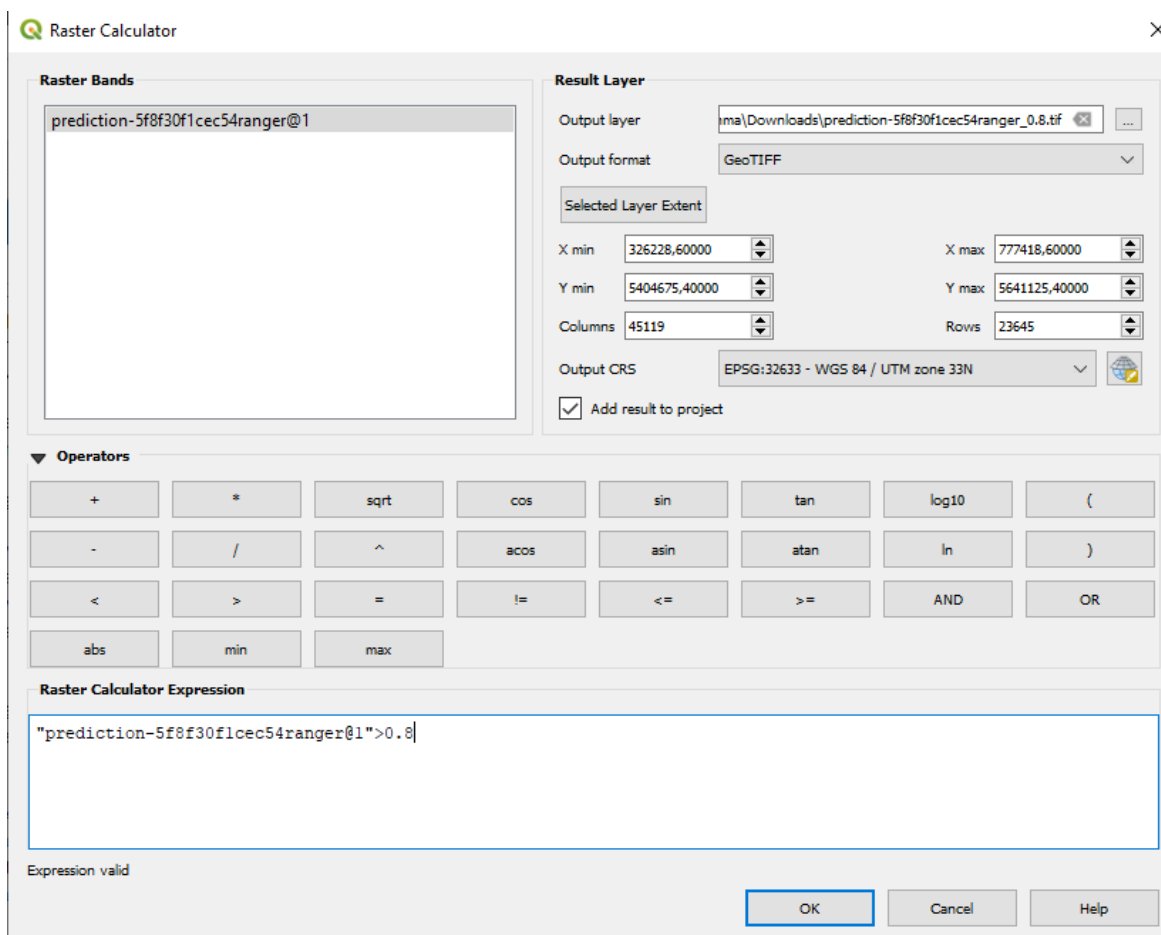
Obrázek 34 Vizualizace pravděpodobnosti vhodného habitatu druhu *Buxbaumia viridis* nedaleko Velkých Losin. Podkladová mapa: WMS ZM25, ČÚZK. Výstup z QGIS. Prahová hodnota relativní pravděpodobnosti vyšší než 0.8 (Model RF; Kappa = 0.67, TSS = 0.61)

Výpočetně optimalizovaný postup generování bodů pro terénní průzkum:

1. Reklasifikace rastru (Obrázek 35)

Nejrychlejší a nejspolehlivější cestou, jak v QGIS filtrovat buňky rastru nad nebo pod určitou prahovou hodnotou, je využití rastr kalkulátoru ze záložky **Raster > Raster Calculator**. Tam stačí zadat jednoduchou formulku cílový rastr > než prahová hodnota.

Např: "prediction-5f8f30f1cec54ranger@1">0.8



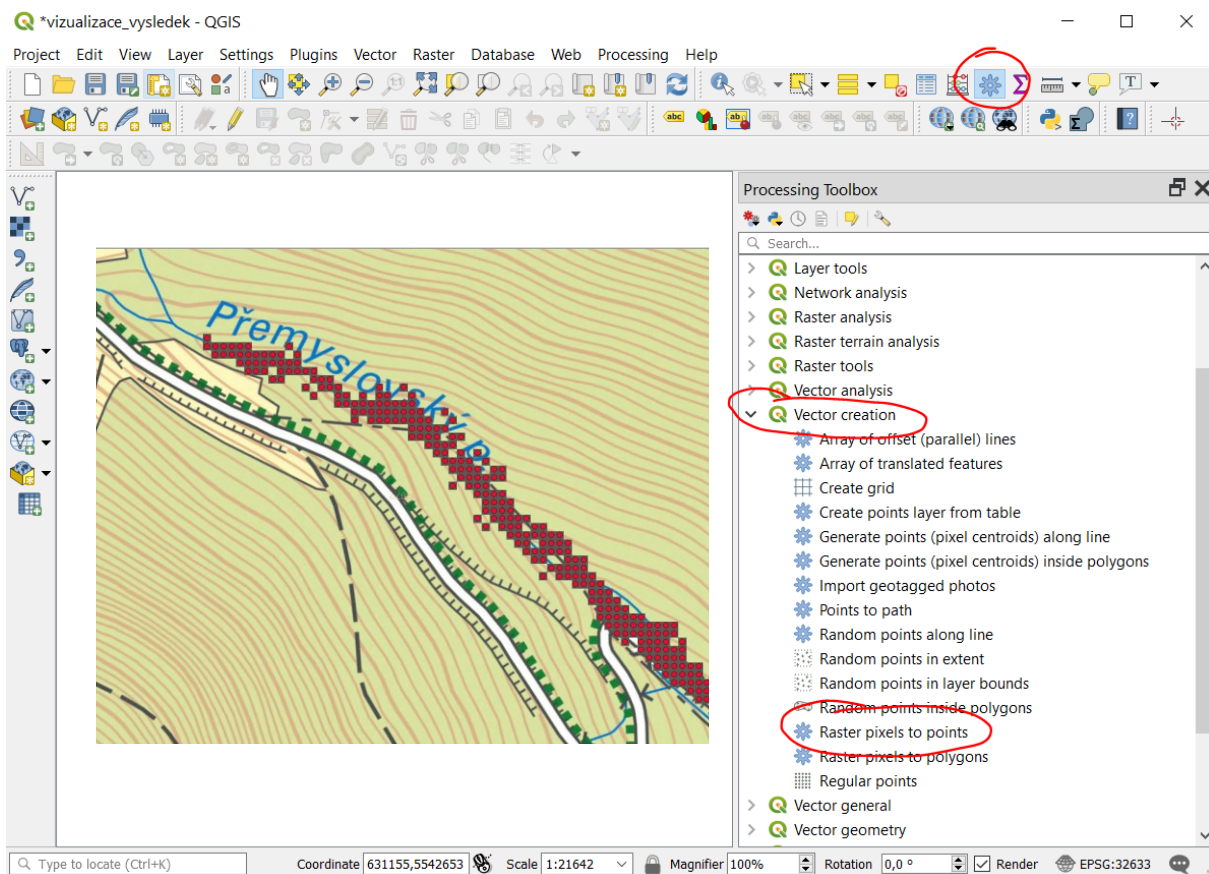
Obrázek 35 Reklasifikace rastru pomocí Raster Calculator v QGIS. Příklad prahové hodnoty 0.8

2. Nastavení nuly jako NoData

Výsledkem reklasifikace (kroku 1) je rastr s hodnotami 0 / 1 (Pravda / Nepravda) podle zadaného kritéria. Hodnoty nula jsou pro nás nepotřebné a zatěžovaly by dále výpočetní kapacitu při další práci, můžeme se jich zbavit tak, že je nastavíme jako NoData, což jde v QGIS elegantně. **Properties > Transparency > No Data Value > Additional no data value**, do kolonky vepíšeme 0.

3. Převod reklasifikovaného rastru na centroidy pixelů = body (Obrázek 36)

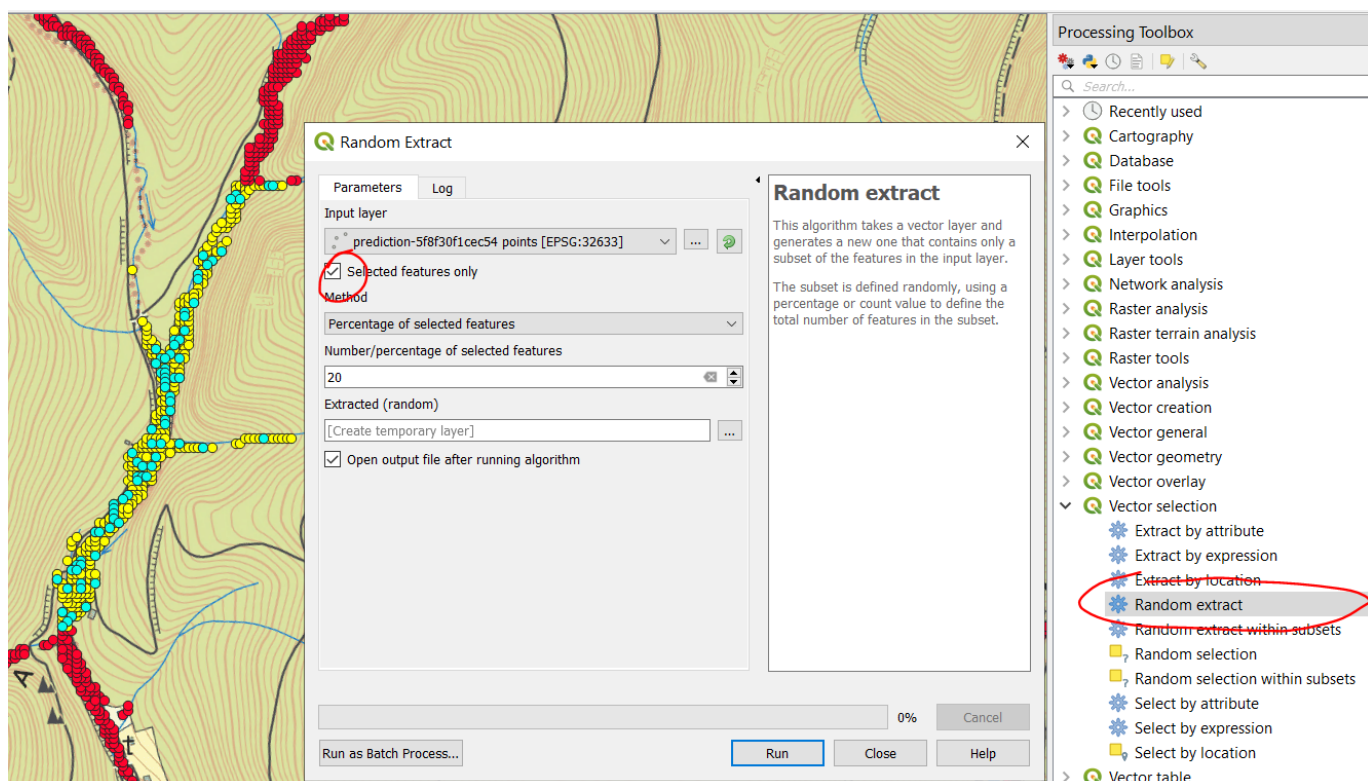
Převod buněk rastru na body (centroidy) je v QGIS velice výpočetně efektivní, ale vygeneruje zpravidla nepoužitelně velké množství bodů, které je pro praktický terénní průzkum potřeba dále zredukovat. Použijeme **Processing Toolbox > Vector creation > Raster pixels to points**. Processing toolbox spustíme ikonou ozubeného kola v horním panelu nástrojů. Pokud tam ozubené kolo není, je třeba Processing toolbox aktivovat **Plugins > Manage and install plugins > Installed**, a tam aktivovat plugin **Processing**.



Obrázek 36 Generování centroidů z pixelů reklasifikovaného rastru, QGIS se zobrazením výsledku operace. Šedivě původní rastr, červeně centroidy.

4. Výběr pouze některých bodů pro terénní průzkum (Obrázek 37)

Protože předchozí krok vygeneroval pravděpodobně příliš velké množství bodů, než abychom je přímo nahráli do terénního zařízení (GPS, chytrého telefonu, tabletu...), je vhodné jejich počet zredukovat na náhodný pod výběr. To můžeme udělat buď manuálně nebo použít nástroj **Processing toolbox > Vector selection > Random extract**. Ten vytvoří novou bodovou vrstvu na základě specifikace počtu či proporce všech centroidů. Počet extrahovaných bodů závisí na rozhodnutí uživatele. Zkušenost ukazuje, že 20 % bodů stále ještě dobře reprezentuje vybranou oblast a je to pro terénní zařízení únosné technicky i s ohledem na zobrazení. Pokud si předem označíme pouze část centroidů, například v menším zájmové území, můžeme aktivovat volbu **Selected features only**, QGIS pak provede výběr pouze z označených centroidů.



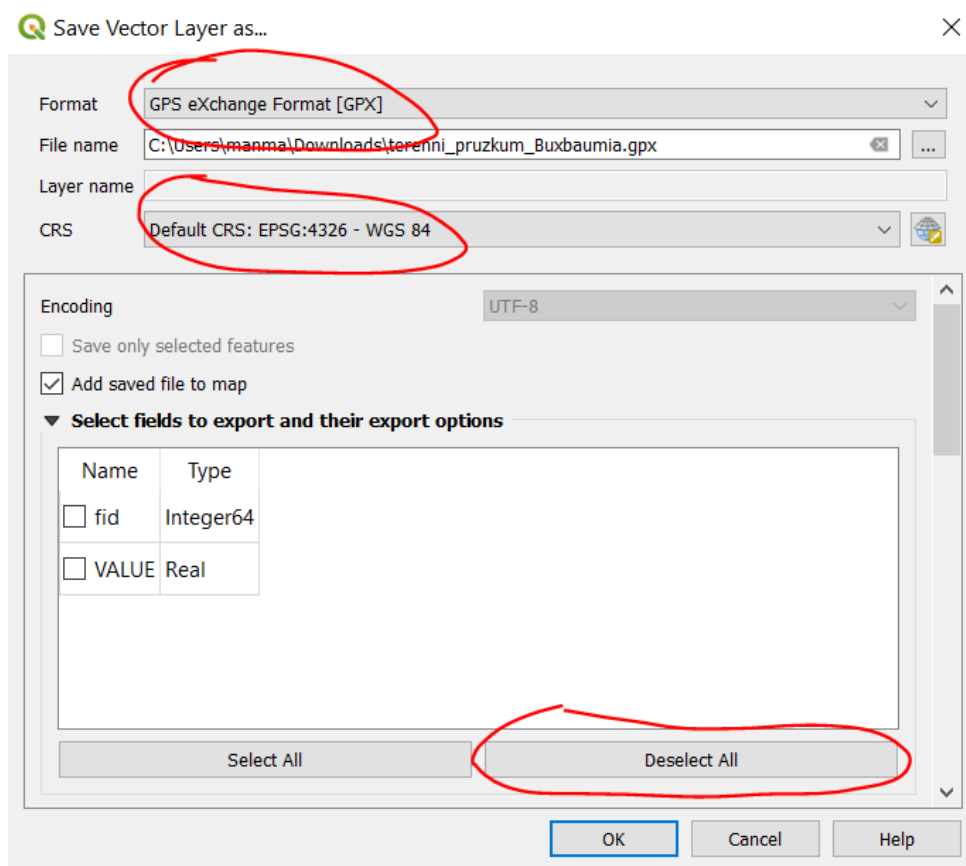
Obrázek 37 Extrakce 20 % náhodně vybraných bodů. Červeně centroidy pixelů rastru, žlutě označené body, tyrkysově náhodně vybrané body z označených. QGIS.

5. Export výsledných bodů pro použití v terénním zařízení (Obrázek 38)

Pokud je uživatel s výsledkem filtrování spokojen, může jej uložit do výměnného formátu GPX, nebo pokud má k dispozici, nahrát na geoserver např. komerční ArcGIS online. Pro export použijeme pravý klik na výslednou vrstvu a odtud **Export > Save feature as**. Při exportu GPX je třeba deaktivovat export atributů (Deselect all), protože GPX formát umožňuje export pouze předepsaných atributů pevné struktury, kterou QGIS automaticky vytvoří. Je také potřeba změnit CRS, nastavit jej na WGS 84 (EPSG 4326), což je podporovaný referenční systém externích zařízení a aplikací.

GPX soubor je možné nahrát přímo do turistických GPS přijímačů i do většiny mapových aplikací pro chytré telefony a tablety (bohužel na konci roku 2020 nelze do mapy.cz ani google maps). Doporučujeme zmiňovanou aplikaci Locus Map Free¹⁷.

¹⁷ <https://www.locusmap.eu/cz/>



Obrázek 38 Nastavení exportu výsledné vrstvy ve výměnném formátu GPX pro externí aplikace. QGIS.

4.7.3 Metoda biodiversity hotspots

Výsledné mapy pravděpodobné vhodnosti habitatu, které vrátí predikční aplikace, je v GIS možné dále zpracovat i jiným způsobem než jen vygenerovat body pro průzkum lokalit jediného druhu, ale také provést jejich kombinaci pro více druhů. Tím můžeme potenciálně odhalit neprobádaná místa s vyšší druhovou bohatostí. Největší smysl dává tato metoda v případě kombinace druhů, kde předpokládáme překryv ekologické niky. Například odděleně pro lesní druhy, horské, vlhkomilné, nebo naopak druhy exponovaných stanovišť atp. Díky tomu, že jsou mapy přeškolované na relativní pravděpodobnost a nabývají vždy hodnot v rozmezí nula až jedna, je možné je kombinovat prostým součtem, průměrem či váženým průměrem. I zde je třeba mít opět na paměti, že i velice špatný model vyprodukuje mapu s relativní pravděpodobností škálovanou od nuly do jedné.

V QGIS je možné kombinaci vrstev provádět pomocí nástroje **Raster > Raster Calculator** viz obrázek 36.

4.8 Případová studie NP České Švýcarsko

Kromě technického ověření funkčnosti této metodiky proběhlo také ověření její konceptuální a praktické stránky na případové studii vyhledávání nových lokalit dvou místně vzácných druhů mechorostů (*Dicranum majus* a *Polytrichastrum alpinum*) na území národního parku České Švýcarsko. Ten je známý velkou diverzitou geomorfologických a mikroklimatických podmínek, stejně jako celkově velkou biomasou i diverzitou lesních mechorostů.

a)



b)



Obrázek 39 Cílové druhy případové studie z NP České Švýcarsko a) *Dicranum majus* - foto Martina Procházková, b) *Polytrichastrum alpinum* - foto Štěpán Koval

Nejcennější část parku je tvořena soustavou hlubokých pískovcových roklí, nad nimiž se tyčí skalní věže. Gradient chladných roklí a osluněných skalních věží je zde k nalezení na poměrně malém a těžko přístupném území. Právě nelehká prostupnost tamního terénu příliš nepřeje zevrubným terénním botanickým průzkumům, tím méně pak bryologickému či lichenologickému. Z území Českého Švýcarska jsou hlášené výskyty lokálně zajímavých i národně vzácných druhů mechorostů a lišejníků. Botanici z tamní správy se domnívali, že se na území parku některé vzácné druhy mohou vyskytovat mnohem hojněji, než ukazují do té doby známé lokality výskytu.

A tak v rámci diplomové práce (Procházková 2019) vedené autory této metodiky při katedře botaniky na Přírodovědecké fakultě Univerzity Karlovy vznikla případová studie, kdy úspěšně proběhlo nalezení dosud neznámých lokalit cílových druhů. Nejprve byl na základě environmentálních a trénovacích dat o distribuci cílových druhů pro každý zvlášť sestaven model potenciální vhodnosti habitatu. Po kritické evaluaci modelu byly vytvořeny predikční mapy potenciální vhodnosti habitatu. Mapy byly dále vizualizovány a zpracovány v GIS a na jejich základě bylo provedeno terénní šetření. Celkem bylo navštíveno 30 vytipovaných lokalit. 10 lokalit s vysokou pravděpodobností vhodnosti pro *Dicranum*, 10 pro *Polytrichastrum* a pro kontrolu i 10 lokalit, kde byl pro oba druhy predikovaný zcela nevhodný habitat.

Návštěva 20 potenciálně vhodných lokalit přinesla objevení šesti nových lokalit. Čtyř pro *Dicranum majus* a dvou pro *Polytrichastrum alpinum*. Na potenciálně nevhodných lokalitách nebyl žádný z druhů nalezen. Podrobnější popis metodiky, použitých environmentálních vrstev a diskuse výsledků viz volně dostupná práce (Procházková 2019).

5 Novost metodiky

Tato metodika je první svého druhu v ČR, která cílí na podporu vyhledávání nových lokalit výskytu vzácných či ohrožených druhů mechorostů a lišejníků. Cílené mapování nových lokalit těchto skupin organismů nemá na našem území příliš velkou podporu a je omezeno zejména na evropsky významné druhy. Novost této metodiky spočívá v praktické aplikaci distribučního modelování výskytu potenciálně vhodných habitatů k zacílení terénního průzkumu. To je praxe, která se začíná využívat v zahraničí, v ČR však zatím jen velmi okrajově, a nikoliv pro mechorosty ani lišejníky.

Skrze metodiku je uživatelům zcela nově zpřístupněn autorský predikční software. Kde jsou k dispozici unikátní environmentální data připravená autorským kolektivem metodiky s pokrytím celé ČR ve velmi podrobném rozlišení 10 m. Zcela nová je také možnost predikovat potenciální vhodnost habitatů s využitím trénovacích dat přímo z národní výskytové databáze mechorostů a lišejníků.

Metodika pojímá nově a moderně i přípravu dat pro praktické využití terénními pracovníky při místních průzkumech s využitím chytrých zařízení a moderních aplikací. Metodika implementuje dílčí kroky a postupy, tak aby na sebe procesně navazovaly a tvořili kompaktní, dobře dokumentovaný celek, přístupný na jednom místě. Metodika je zpracovaná specificky pro použití v ČR – obecné postupy a pravidla jsou, pokud to bylo možné, implementované lokálně a nastaveny pro místní podmínky.

6 Popis uplatnění metodiky

Metodika slouží jako nástroj pro efektivní zacílení terénního výzkumu, jehož cílem je nalezení nových lokalit s výskytem vzácných či ohrožených organismů. Cílovou skupinou této metodiky jsou zejména pracovníci a pracovnice v ochraně přírody, mající na starosti biologický průzkum. Metodika je zaměřená primárně na mechorosty a lišejníky, pro které je dostupná i nadstavba v podobě možnosti vytváření predikcí přímo z národní nálezové databáze. Metodika je však použitelná pro jakýkoliv organismus, u kterého je možné předpokládat výskyt alespoň částečně determinovaný environmentálními faktory, které tato metodika nabízí, viz kapitola 4.2. Metodiku je možné uplatnit na Agentuře ochrany přírody a krajiny, na správách chráněných území, krajských či městských úřadech i na univerzitách. Metodiku je možné uplatnit při obecné snaze o lepší poznání rozšíření druhů, ale i pro odhalování potenciální kolize se stavebními záměry a výskytem vzácných druhů či jako podklad pro studentské a absolventské práce v biologických oborech, kde jde o terénní průzkumy zaměřené na konkrétní druhy organismů.

Účelem této metodiky je doplnit a podpořit expertní přístup k biologickým průzkumům, nikoliv jej nahradit či mu oponovat. Tato metodiky musí být využívána kriticky a informovaně minimálně se znalostí prezentovanou v teoretické části metodiky.

7 Ekonomické aspekty

Tato metodika není zaměřena na ekonomické aktivity. Jejím cílem není podpořit generování zisku, na druhou stranu tato metodika ani nezavádí postupy, které by vyžadovaly vstupní ani průběžné investice od svých uživatelů. Tato metodika může pomoci ušetřit finanční prostředky vynaložené na biologický průzkum. Na druhou stranu může tato metodika vést ke zvýšení finančních prostředků alokovaných na biologický průzkum v případě, kdy uživatele motivuje k zahájení monitorovacích či vyhledávacích aktivit, které by bez této metodiky vůbec neprobíhaly.

Jedním z cílů této metodiky je zefektivnit vyhledávání nových lokalit vzácných či ohrožených druhů mechorostů a lišejníků. Zefektivněním zde rozumíme úsporu finančních i lidských zdrojů při porovnání terénního průzkumu prováděného bez využití této metodiky a s jejím využitím. Ani zde není ekonomický dopad jasně vyčíslitelný.

- 1.) Pokud bychom si představili situaci naivního pracovníka, který se vydá náhodně do terénu za účelem nalezení nové lokality vzácného druhu, pak by finanční i časová úspora generovaná aplikací této metodiky byla obrovská. Příklad naivního pracovníka a náhodného vyhledávání však patrně není příklad z reálného světa biologických průzkumů.
- 2.) Reálnější je situace informovaného pracovníka s biologickým vzděláním, který dokáže díky své zkušenosti a znalosti ekologie daných druhů průzkum zacílit, a to minimálně do oblastí geograficky blízké známému výskytu cílových druhů. I v tomto případě si troufáme tvrdit, že může předkládaná metodika představovat ekonomickou i časovou úsporu terénních pracovníků prostřednictvím lepšího zacílení zkoumaného regionu. Metodika zde nabízí podporu například v regionech ČR, se kterými nemá odborný pracovník zkušenost a mohly tak jeho pozornosti uniknout.
- 3.) A konečně je tu případ, kdy terénní průzkum provádí profesionální bryolog či lichenolog, který bude patrně efektivnější než předkládaná metodika. I zde je však prostor pro její aplikaci opět na jemu málo známá území či přesnější zacílení v rámci předpokládaného místa výskytu. Půjdeme-li zde do finančního detailu, pak se nabízí porovnání času a platu, který na takový průzkum či jeho plánování spotřebuje profesionální bryolog či lichenolog a zcela minimální náklady na provedení kroků v této metodice, jejíž výstup může profesionální bryolog či lichenolog pouze validovat a předat méně kvalifikovaným pracovníkům k provedení samotného terénního šetření.

Věříme, že tato metodika bude mít motivační efekt a v konečném důsledku povede ke zvýšení financí alokovaných do terénního bryologického a lichenologického průzkumu, který ale díky metodice bude efektivnější než bez ní.

8 Citovaná literatura

- Allouche, O., Tsoar, A., & Kadmon, R. 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43: 1223–1232.
- Austin, M.P., & Smith, T.M. 1989. A new model for the continuum concept. *Vegetatio* 83: 35–47.
- Barbé, M., Fenton, N.J., & Bergeron, Y. 2016. So close and yet so far away: long-distance dispersal events govern bryophyte metacommunity reassembly (P. Veski, Ed.). *Journal of Ecology* 104: 1707–1719.
- Barbet-Massin, M., Jiguet, F., Albert, C.H., & Thuiller, W. 2012. Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution* 3: 327–338.
- 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: 2793–2804.
- Callaghan, D.A., & Ashton, P.A. 2008. Knowledge gaps in bryophyte distribution and prediction of species-richness. *Journal of Bryology* 30: 147–158.
- Cerrejón, C., Valeria, O., Mansuy, N., Barbé, M., & Fenton, N.J. 2020. Predictive mapping of bryophyte richness patterns in boreal forests using species distribution models and remote sensing data. *Ecological Indicators* 119.
- Chaine, I., & Beaubien, E.G. 2001. Phenology is a major determinant of tree species range. *Ecology Letters*.
- Coudun, C., & Gégout, J. 2007. Quantitative prediction of the distribution and abundance of *Vaccinium myrtillus* with climatic and edaphic factors. *Journal of Vegetation Science*.
- Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., & E. Zimmermann, N. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.
- Fick, S.E., & Hijmans, R.J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37: 4302–4315.
- Fourcade, Y., Besnard, A.G., & Secondi, J. 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography* 27: 245–256.
- Gábor, L., Moudrý, V., Barták, V., & Lecours, V. 2020. How do species and data characteristics affect species distribution models and when to use environmental filtering? *International Journal of Geographical Information Science* 34: 1567–1584.
- Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A., & Zimmermann, N.E. 2006. Using niche-based models to improve the sampling of rare species. *Conservation Biology*.
- Guisan, A., & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147–186.
- Hastie, T., Tibshirani, R., & Friedman, J. 2009. *The Elements of Statistical Learning The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition*.

- Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. 2017. SoilGrids250m: Global gridded soil information based on machine learning (B. Bond-Lamberty, Ed.). *PLOS ONE* 12
- Hirzel, A., & Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling* 157: 331–341.
- Holá, E., & Turoňová, D. 2014. Hodnotící zprávy o stavu mechorostů a cévnatých rostlin – podruhé. *Ochrana přírody* 22–25.
- Hollings, T., Robinson, A., van Andel, M., Jewell, C., & Burgman, M. 2017. Species distribution models: A comparison of statistical approaches for livestock and disease epidemics (L. Kumar, Ed.). *PLOS ONE* 12
- Horydská, E., Krása, A., Neuwirthová, H., & Tomášková, L. 2011. K aktualizaci seznamu zvláště chráněných druhů. *Ochrana přírody* 14–17.
- Hutchinson, G.E. 1957. Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22: 415–427.
- Jandová, J., & Holá, E. 2012. Vzácný mech dvouhrotec zelený. *Ochrana přírody* 22–23.
- Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., & Kessler, M. 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4:.
- Kremen, C., Cameron, A., Moilanen, A., Phillips, S.J., Thomas, C.D., Beentje, H., Dransfield, J., Fisher, B.L., Glaw, F., Good, T.C., Harper, G.J., Hijmans, R.J., Lees, D.C., Louis, E., Nussbaum, R.A., Raxworthy, C.J., Razafimpahanana, A., Schatz, G.E., Vences, M., Vieites, D.R., Wright, P.C., & Zjhra, M.L. 2008. Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science* 320: 222–226.
- Kučera, J., Váňa, J., & Hradílek, Z. 2012. Bryoflóra České republiky: aktualizace seznamu a červeného seznamu a stručná analýza. *Preslia* 84: 813–850.
- Kuhn, M. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software* 28: 1–26.
- Lawson, C.R., Hodgson, J.A., Wilson, R.J., & Richards, S.A. 2014. Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution* 5: 54–64.
- Leroy, B., Delsol, R., Hugué, B., Meynard, C.N., Barhoumi, C., Barbet-Massin, M., & Bellard, C. 2018. Without quality presence-absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography* 45: 1994–2002.
- Liška, J., & Palice, Z. 2010. Red List of lichens of the Czech Republic (version 1.1). *Příroda* 29: 68.
- Liu, C., Newell, G., & White, M. 2019. The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. *Ecography* 42: 535–548.
- Liu, C., White, M., Newell, G., & Griffioen, P. 2013. Species distribution modelling for conservation planning in Victoria, Australia. *Ecological Modelling* 249: 68–74.
- Lobo, J.M., Jimenez-Valverde, A., & Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17: 145–151.

- Malíček, J., Palice, Z., & Vondrák, J. 2018. Additions and Corrections to the Lichen Biota of the Czech Republic Additions and corrections to the lichen biota of the Czech Republic. *Herzogia* 31: 453–475.
- Meynard, C.N., & Kaplan, D.M. 2012. The effect of a gradual response to the environment on species distribution modeling performance. *Ecography* 35: 499–509.
- Mi, C., Huettmann, F., Guo, Y., Han, X., & Wen, L. 2017. Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ*.
- Mod, H.K., Scherrer, D., Luoto, M., & Guisan, A. 2016. What we use is not what we know: environmental predictors in plant distribution models. *Journal of Vegetation Science* 27: 1308–1322.
- Patiño, J., & Vanderpoorten, A. 2018. Bryophyte Biogeography. *Critical Reviews in Plant Sciences*.
- Pecchi, M., Marchi, M., Burton, V., Giannetti, F., Moriondo, M., Bernetti, I., Bindi, M., & Chirici, G. 2019. Species distribution modelling to support forest management. A literature review. *Ecological Modelling* 411.
- Phillips, S.J., Anderson, R.P., & Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*.
- Polechová, J., & Storch, D. 2018. Ecological niche. In *Encyclopedia of Ecology*, pp. 72–80. Elsevier.
- Procházková, M. 2019. Predikční modelování potenciálního výskytu vybraných druhů mechorostů na území Národního parku České Švýcarsko. diplomová práce. školitel: Man, M.
- R Core Team. 2016. R: A Language and Environment for Statistical Computing.
- Reese, G.C., Wilson, K.R., Hoeting, J.A., & Flather, C.H. 2005. Factors affecting species distribution predictions: A simulation modeling experiment. *Ecological Applications* 15: 554–564.
- Sérgio, C., Figueira, R., Draper, D., Menezes, R., & Sousa, A.J. 2007. Modelling bryophyte distribution based on ecological information for extent of occurrence assessment. *Biological Conservation* 135: 341–351.
- Skowronek, S., Ewald, M., Isermann, M., Van De Kerchove, R., Lenoir, J., Aerts, R., Warrie, J., Hattab, T., Honnay, O., Schmidtlein, S., Rocchini, D., Somers, B., & Feilhauer, H. 2017. Mapping an invasive bryophyte species using hyperspectral remote sensing data. *Biological Invasions* 19: 239–254.
- Smith, A.B., & Santos, M.J. 2020. Testing the ability of species distribution models to infer variable importance. *Ecography*.
- Spitale, D., & Mair, P. 2015. 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*.
- Štechová, T., & Kučera, J. 2007. The requirements of the rare moss, *Hamatocaulis vernicosus* (Calliergonaceae, Musci), in the Czech Republic in relation to vegetation, water chemistry and management. *Biological Conservation* 135: 443–449.
- Stockwell, D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International journal of geographical information science* 13: 143–158.
- Wisn, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., Elith, J., Dudík, M., Ferrier, S., Huettmann, F., Leathwick, J.R., Lehmann, A., Lohmann, L., Loiselle, B.A., Manion, G., Moritz, C.,

- Nakamura, M., Nakazawa, Y., Overton, J.M.C., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S.E., & Zimmermann, N.E. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14: 763–773.
- Wright, M.N., & Ziegler, A. 2017. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* 77: 1–17.
- Xie, Y., Allaire, J.J., & Golemund, G. 2018. *R Markdown: The Definitive Guide*. Chapman and Hall/CRC, Boca Raton, Florida.
- Zhu, D., & Liu, Y. 2018. Modelling spatial patterns using graph convolutional networks. In *Leibniz International Proceedings in Informatics, LIPIcs*, Schloss Dagstuhl- Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing.
- Zmrhalová, M., & Holá, E. 2010. Nové poznatky o výskytu mechu šikouška zeleného. *Ochrana přírody* 24–26.

9. Použité zkratky

AIC	Akaike's Information Criterion
ANN	Artificial Neural Networks
AOPK	Agentura ochrany přírody a krajiny
AUC	Area Under Curve
BRNM	Herbář Moravského Muzea
BRT	Boosted Regression Trees
CRS	Geografický referenční systém
CSV	Comma separated values
CBFS	Herbář katedry botaniky BF Jihočeské univerzity
ČNFD	Česká národní fytocenologická databáze
ČÚZK	Český ústav zeměměřický a katastrální
DaLiBor	Database of Lichens and Bryophytes
DIBAVOD	Digitální báze vodohospodářských dat
DPZ	Dálkový průzkum Země
ENFA	Ecological niche factor analysis
GAM	Generalized additive models
GARP	Genetic algorithms for rule production
GBIF	Global Biodiversity Information Facility
GIS	Geografický informační systém
GLM	Generalised Linear Model, generalizovaný lineární model
GPS	Global Positioning System
GUI	Graphic User Interface
IUCN	International Union for Conservation of Nature
MARS	Multivariate adaptive regression splines
MZCHU	Maloplošné zvláště chráněné území
NDOP	Nálezová databáze ochrany přírody
RF	Random Forests
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
SDM	Species distribution modelling
SQL	Structured Query Language
SVM	Support Vector Machines
TSS	True Skill Statistic
UAV	Unmanned aerial vehicle
UTM	Univerzální transversální Mercatorův systém souřadnic
ÚHÚL	Ústav pro hospodářskou úpravu lesů
WGS	World Geodetic System, Světový geodetický systém
WMS	Web Map Service
ZM10 - ZM200	Základní mapa 1: 10 000 - Základní mapa 1: 200 000