

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE
FAKULTA ŽIVOTNÍHO PROSTŘEDÍ

KATEDRA EKOLOGIE



Modelování prostorové distribuce
chytridiomykózy pomocí software MaxEnt

BAKALÁŘSKÁ PRÁCE

Vedoucí práce: Ing. Jiří Vojar, Ph.D.

Bakalant: Martin Sedlák

2013

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Katedra ekologie
Fakulta životního prostředí

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Sedlák Martin

Aplikovaná ekologie

Název práce

Modelování prostorové distribuce chytridiomykózy pomocí software MaxEnt

Anglický název

Chytridiomycosis spatial distribution modelling using software MaxEnt

Cíle práce

Chytridiomykóza je závažné onemocnění obojživelníků zapříčiněné houbou *Batrachochytrium dendrobatidis* (Bd). Tato choroba, dosud objevená u nejméně 350 druhů (v Evropě u 29 druhů ze 14 zemí), představuje ohrožení biodiverzity obojživelníků v globálním měřítku. V České republice (ČR) se její přítomnost začala zjišťovat teprve v roce 2008. Nemoc zde byla prokázána u 4 druhů na 6 lokalitách, zejména u rodu *Bombina*, který patrně funguje jako přenašeč Bd do nových oblastí. V současné době se jako nástroj pro predikci výskytu nemoci používá program MaxEnt, pomocí kterého lze lépe pochopit rozšíření Bd a účinněji zamezit jejímu šíření do nových oblastí.

Konkrétními cíli práce jsou: (i) vypracování stručné literární rešerše věnované modelování prostorové distribuce obecně (programové možnosti s příklady využití u konkrétních taxonů, rozsah 5–8 stran); (ii) podrobnější popis prostředí a práce s nástrojem MaxEnt a jeho využití při modelování distribuce Bd u obojživelníků (rozsah 10–15 stran); (iii) kritické zhodnocení možností programu, jeho výhody, nevýhody, omezení atd.; (iv) aktualizace modelu rozšíření Bd u rodu *Bombina* v ČR, definice oblastí s významnějším výskytem nemoci.

Metodika

Metodika spočívá ve standardní práci s literárními zdroji (WOS atd.) v rámci tvorby literární rešerše a dále zvládnutí programu MaxEnt (propojením klimatických vrstev s digitálním modelem terénu a vrstvou s výskyty Bd v aplikaci ArcGIS + sestavení modelu potenciální distribuce chytridiomykózy u nás).

Harmonogram zpracování

09–12/2012 – tvorba rešerše, zvládnutí programu MaxEnt

01–03/2013 – odevzdání rešerše, tvorba modelu, dokončení pracovní verze BP

04/2013 – dokončení a odevzdání BP

Rozsah textové části

20–30 stran textu, přílohy dle potřeby

Klíčová slova

obojživelníci, Amphibia, nemoci obojživelníků, Batrachochytrium dendrobatidis, GIS

Doporučené zdroje informací

- Bridges V. E., Akkina J., Grannis J., Johnson C., Johnson R., Tuszynski C. 2007: A qualitative assessment tool for the potential of infectious disease emergence and spread. *Preventive Veterinary Medicine*. 81: 80–91.
- Elith J., Phillips S. J., Hastie T., Dudík M., Yung En Chee & Yates C. J., 2011: A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17: 43–57.
- Fisher M. C., Garner T. W. J., Walker S. F. 2009: Global Emergence of *Batrachochytrium dendrobatidis* and Amphibian Chytridiomycosis in Space, Time, and Host. *The Annual Review of Microbiology*. 63: 291–310.
- Garner T. W. J., Walker S., Bosch J., Hyatt A. D., Cunningham A. A., Fisher M. C. 2005: Chytrid fungus in Europe. *Emerging Infectious Diseases*, 11(10): 1639–1641.
- Jones K. E., Patel N. G., Levy M. A., Storeygard A., Balk D., Gittleman J. L. et al. 2008: Global trends in emerging infectious diseases. *Nature*. 451: 990–993.
- Kilpatrick A. M., Briggs C. J., Daszak P. 2010: The ecology and impact of chytridiomycosis: an emerging disease of amphibians. *Trends in Ecology & Evolution*, 25(2): 109–118.
- Murray A. G. 2009: Using simple models to review the application and implications of different approaches used to simulate transmission of pathogens among aquatic animals. *Preventive Veterinary Medicine*. 88: 167–177.
- Nokes D. J., Anderson R. M. 1988: The use of mathematical models in the epidemiological study of infectious diseases and in the design of mass immunization programmes. *Epidemiology and Infection*. 101: 1–20.
- Phillips S. J., Anderson R. P., Schapire R. E. 2006: Maximum entropy modeling of species geographic distributions. *Ecological Modelling*. 190: 231–259.
- Phillips S. J., Dudík M. 2008: Modelling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*. 31: 161–175.
- Phillips S. J., Dudík M., Schapire R. E. 2004: A maximum entropy approach to species distribution modeling. *Proceedings of the Twenty-First International Conference on Machine Learning*. 655–662.
- Rogers D. J., Randolph S. E. 2003: Studying the global distribution of infectious diseases using GIS and RS. *Nature Reviews. Microbiology*. 1: 231–237.
- Thrush M. A., Murray A. G., Brun E., Wallace S., Peeler E. J. 2011: The application of risk and disease modelling to emerging-freshwater diseases in wild aquatic animals. *Freshwater biology*. 56(4): 658–675.

Vedoucí práce

Vojar Jiří, Ing., Ph.D.

Konzultant práce

Ing. Petr Civiš


prof. RNDr. Vladimír Bejček, CSc.

Vedoucí katedry




prof. Ing. Petr Sklenička, CSc.

Děkan fakulty

V Praze dne 21.3.2013

Prohlášení

Prohlašuji, že jsem tuto práci vypracoval samostatně, na základě odborné literatury a pramenů uvedených v seznamu literatury a všechny použité zdroje jsem v textu řádně odcitoval.

V Praze, dne 15. 4. 2013

.....

podpis

Poděkování

Tímto bych chtěl poděkovat mému vedoucímu bakalářské práce Ing. Jiřímu Vojarovi, Ph.D., za cenné rady, podporu a za poskytnutí spousty vědomostí pro napsání bakalářské práce. Dále Ing. Petru Civišovi za poskytnutí dat, a pomoc při vytváření modelu. V neposlední řadě bych chtěl poděkovat rodině a přátelům za morální podporu.

Abstrakt

Chytridiomykóza je houbové onemocnění, které v současné době ohrožuje populace obojživelníků na celém světě. Tato nemoc je způsobena patogenem *Batrachochytrium dendrobatidis*, a je jediným známým chytridiomycetním onemocněním, které postihuje i skupinu obratlovců. Šíření této nemoci se stává globálním problémem a omezení její další expanze se stává prioritou ochránců přírody v mnoha ohrožených oblastech. K plánování ochrany populací obojživelníků před tímto patogenem je využíváno nástrojů geografických informačních systémů v kombinaci s metodami sloužícími k vytváření modelů prostorové distribuce. Za použití těchto nástrojů je možné vymezit oblasti výskytu onemocnění a zastavit její další šíření z těchto lokalit. Tato práce je zaměřena na představení základních metod využívaných k modelování prostorové distribuce a seznámení s programem MaxEnt. V rámci práce s programem MaxEnt byla vytvořena mapa současného rozšíření chytridiomykózy v České republice.

Klíčová slova:

obojživelníci, *Amphibia*, nemoci obojživelníků, *Batrachochytrium dendrobatidis*, GIS, SDM

Abstract

Chytridiomycosis is a disease of fungus origin, which endangers populations of amphibians all over the world. The disease is caused by chytrid fungus *Batrachochytrium dendrobatidis*, and it is the only chytrid fungus known to attack a group of vertebrates. Spreading of this disease becomes a global problem and stopping this spread becomes a priority of conservators in many endangered areas. Geographic information system tools in combination with species distribution modelling tools are used for planning the amphibian protection against this pathogen. These tools are able to define areas with high occurrence probability of the fungus and they can be used to stop the spreading of the disease from these areas. This work aims to introduce the basics of using the species distribution modeling and to show the usage of MaxEnt modelling software. A map analyzing the current distribution of chytridiomycosis in the Czech Republic was created within the work with this software.

Key Words:

amphibians, *Amphibia*, diseases of amphibians, *Batrachochytrium dendrobatidis*, GIS, SDM

Obsah

1. Úvod	9
1.1 Cíle práce	11
2. Literární rešerše	12
2.1 Rozšíření chytridiomykózy	12
2.1.1 Rozšíření ve světě	12
2.1.2 Rozšíření v Evropě	13
2.1.3 Chytridiomykóza v ČR.....	13
2.2 Modelování prostorového rozšíření organismů	14
2.2.1 Základy modelování prostorové distribuce.....	14
2.2.2 Typy dat a práce s nimi.....	16
2.2.3 Metody modelování a software	17
2.2.4 Programy pro modelování prostorové distribuce.....	23
2.3 MaxEnt.....	25
2.3.1 Princip maximální entropie	25
2.3.2 Software MaxEnt.....	25
3. Metodika práce s programem MaxEnt	26
3.1 Příprava dat pro MaxEnt.....	26
3.2 Import dat a základní nastavení modelu	28
3.3 Zvolené nastavení pro predikci rozšíření <i>Bd</i>	31
4. Výsledky práce	32
5. Diskuze	36
6. Závěr.....	37
7. Použitá literatura:	38
7.1. Literární zdroje.....	38
7.2. Internetové zdroje:.....	40

8. Přílohy	41
------------------	----

Seznam zkratek:

AUC – Area Under Curve

Bd – *Batrachochytrium dendrobatidis*

CENIA – Česká informační agentura životního prostředí

ČÚZK – Český úřad zeměměřický a katastrální

ČZU – Česká zemědělská univerzita v Praze

FŽP – Fakulta životního prostředí

GIS – Geographical Information System

GLM – Generalized Linear Model

ROC – Receiver Operating Characteristic

SDM – Species Distribution Modelling

S-JTSK – Systém jednotné trigonometrické sítě katastrální

VFU – Veterinární a farmaceutická univerzita v Brně

WGS 84 – World Geodetic System 1984

1. Úvod

Obojživelníci jsou prastarou a v současnosti spolu s plazy nejvíce ohroženou skupinou obratlovců, a to vlivem mnoha faktorů včetně lidské činnosti (Vié et al. 2008). V posledních desetiletích se tlak na životní prostředí vlivem člověka stále zvyšuje. Tomuto tlaku jsou vystaveni i obojživelníci. Jedním z nebezpečí, kterým tato skupina čelí, jsou znečišťování, změna či úplná likvidace vhodných biotopů. K těmto jevům dochází především díky změně zemědělského a lesnického hospodaření či výstavbě sídel a dopravní infrastruktury (Vojar et al. 2012). Ke globálně působícím faktorům patří kontaminace prostředí, zvýšené UV záření a změny klimatu (Gibson & Freeman 1997). Tyto jevy často působí synergicky, přičemž ne vždy je možno pravou příčinu ohrožení definovat (Keisecker et al. 2001). Významným ohrožením jsou rovněž nemoci, zejména pak v nedávné době objevené houbové onemocnění chytridiomykóza (Berger et al. 1998, Longcore et al. 1999).

Chytridiomykóza je houbové onemocnění kůže obojživelníků způsobené houbou *Batrachochytrium dendrobatidis* (dále jen *Bd*). Toto onemocnění vede k odlupování či rohovatění pokožky a má následný dopad na kožní dýchání a osmoregulaci (Berger et al. 1998, Voyles et al. 2009). U nakažených jedinců může docházet ke zhoršení jejich kondice, změnám chování včetně ztrát antipredačních reflexů a v některých případech tato nemoc způsobuje i přímé uhynutí. Existence *Bd* byla zjištěna v roce 1998 (Berger et al. 1998), ale její výskyt byl zpětně zaznamenán na musejních vzorcích drápatek z Jižní Afriky až z roku 1938 (Weldon 2004). Chytridiomykóza se v průběhu 20. století rozšířila na všechny kontinenty kromě Antarktidy (Fischer et al. 2009) a v současné době představuje globální ohrožení obojživelníků, o čemž svědčí i její zapsání na Seznam nemocí volně žijících zvířat (Wildlife Disease List) Světovou organizací pro zdraví zvířat (The World Organization for Animal Health) v roce 2001 (Johnson & Speare 2003).

V České republice byla nemoc poprvé zaznamenána v roce 2008 (Civiš et al. 2010). První rozsáhlejší průzkum týkající se přítomnosti *Bd* v České republice, byl proveden Balážem, Civišem a Vojarem v roce 2009; poměrně intenzivní monitoring výskytu chytridiomykózy ve spolupráci FŽP ČZU Praha a VFU Brno pokračuje až do současnosti (Civiš et al. 2010, Vojar os. sdělení).

Vzhledem k nebezpečí jaké chytridiomykóza představuje pro populace obojživelníků, jsou po celém světě prováděna mapování jejího výskytu a výsledky bývají zveřejňovány. Příkladem jsou stránky <http://www.bd-maps.net/>, kde jsou umístěny mapové aplikace s podrobnými záznamy o výskytu *Bd*. Data získaná při monitorování jsou velmi cennými zdroji informací, které mohou být dále využívány při zkoumání této choroby. Jedním z mnoha využití těchto dat je jejich aplikace v prostorových analýzách. Současné možnosti nástrojů geografických informačních systémů (GIS) nám umožňují vytvářet nejrůznější analýzy využitelné v řadě vědních oborů. S využitím informací o výskytu chytridiomykózy a environmentálních dat je možné modelovat (předvídat) rozšíření nemoci a definovat vlastnosti prostředí, které výskyt onemocnění podporují. Jedním z programů, používaných v prostorovém modelování, je software MaxEnt (Franklin & Miller 2009), který využívá principu maximální entropie a je díky své schopnosti vytváření predikcí na poměrně malém vzorku dat hojně využíván právě k modelování rozšíření *Bd* (např. Puschendorf 2009, Röder et al. 2009, Measey 2012). První a poslední predikcí rozšíření patogenu v ČR je model vytvořený Civišem (2010), který obsahuje data shromážděná před rokem 2010. Vzhledem k tomu, že od té doby je k dispozici značné množství nových záznamů o přítomnosti *Bd* na území ČR, by vytvoření aktuálního modelu bylo velmi užitečnou pomůckou při ochraně obojživelníků před tímto patogenem. Jedním z cílů této práce tedy je takový model vytvořit. Predikce rozšíření *Bd*, následně může sloužit k vymezení kritických oblastí se zvýšeným výskytem patogenu, které mají být podrobeny monitoringu za účelem zamezení rozšiřování nemoci do okolních oblastí. Puschendorf et al. (2009) naopak navrhuje použít predikci výskytu k lokalizaci oblastí s nejmenší pravděpodobností výskytu patogenu a jejich následnému využití jako „záchranných center“ pro druhy ohrožené chytridiomykózou.

1.1 Cíle práce

Hlavním cílem této práce je vytvořit aktualizovaný model rozšíření *Bd* v České republice a přiblížit práci s prostředím MaxEnt. Dílčími cíly jsou:

- Vytvoření stručné rešerše věnované tématu modelování prostorové distribuce.
- Podrobný popis prostředí MaxEnt a práce s tímto programem včetně jeho využití při modelování distribuce *Bd*.
- Zhodnocení možností programu, jeho výhody, nevýhody a omezení.
- Aktualizace modelu rozšíření *Bd* v České republice a definice oblastí s významnějším výskytem nemoci.

2. Literární rešerše

2.1 Rozšíření chytridiomykózy

2.1.1 Rozšíření ve světě

K rozšíření chytridiomykózy došlo pravděpodobně činností člověka, a to při chovu afrických drápatek rodu *Xenopus*, které přišly do styku s volně žijícími obojživelníky (Weldon et al. 2004). Samotný rod *Xenopus* má vůči chytridiomykóze silnou rezistenci, což z nich činí skvělého přenašeče. Vzhledem ke své nenáročnosti se drápatky staly oblíbené v zájmovém i laboratorním chovu a často byly, a stále jsou, využívány ve výzkumu a lékařství (Stanley 1949, Asashima et al. 2009). Tato zvířata přicházela během chovu do styku s domácími druhy a docházelo tak k přímému přenosu onemocnění. Další cestu mohly poskytovat odpadní vody, ve kterých končila uhynulá zvířata nesoucí na sobě zárodky *Bd*. Další možný způsob přenosu mohlo představovat vypouštění chovaných jedinců do volné přírody. V neposlední řadě svojí roli sehrála i introdukce drápatek ve Spojených státech, Chile a ve Spojeném království. Dle Weldon et al. (2004) byl další výskyt *Bd* mimo africký kontinent zaznamenán v roce 1961, a to v Severní Americe u populace *Rana clamitans*, obývající východní pobřeží Spojených států. Díky mezinárodnímu obchodu s drápatkami a jinými druhy chovanými v zajetí (např. skokan volský, *Lithobates catesbeianus*) se nemoc postupně rozšířila i do dalších částí světa. Postup choroby mezi kontinenty pokračoval dále do Austrálie (1978), Střední Ameriky (1983), Jižní Ameriky (1986), Evropy (1997) i Oceánie (1999), (Weldon et al. 2004). Současné šíření *Bd* představuje tedy globální problém, který je zvláště nebezpečný hlavně v klimaticky teplých a vlhkých oblastech, kde je její výskyt velmi hojný v důsledku příznivějších podmínek pro přežívání houbových zárodků (Berger et al. 1998). Zvláště velkou hrozbu představuje chytridiomykóza pro endemické druhy deštných lesů, které díky svému malému areálu a nízké početnosti mohou být snadno vyhubeny. Jak uvádí Fischer et al. (2009), *Bd* způsobila vyhynutí již nejméně 34 druhů obojživelníků a její výskyt byl prokázán u více než 350 druhů po celém světě.

2.1.2 Rozšíření v Evropě

První výskyt *Bd* na území Evropy byl hlášen z Německa, a to v roce 2000, kde byla zjištěna přítomnost patogenu na dovezených obojživelnících z Jižní Ameriky. Stejný rok byla přítomnost *Bd* hlášena na jedincích odchovávaných v zajetí, opět z Německa a také z Belgie (Mutschmann et al. 2000). K prvnímu zaznamenanému úhynu jedinců v důsledku chytridiomykózy došlo ve Španělsku v Národním parku Peñalara. V roce 2001 zde byl pozorován rapidní úbytek početnosti zdejších populací obojživelníků (Bosch & Martínez-Solano 2001). Díky zpětnému testování na přítomnost *Bd* bylo z archivovaných vzorků zjištěno, že první výskyt *Bd* v Evropě byl nevědomky zaznamenán již v roce 1997, ve Španělsku (Weldon et al. 2004, Garner 2005). V současné době je výskyt chytridiomykózy evidován v 15 státech Evropy (Česká republika, Estonsko, Francie, Chorvatsko, Itálie, Lotyšsko, Lucembursko, Maďarsko, Německo, Portugalsko, Slovensko, Slovinsko, Španělsko, Švýcarsko, Velká Británie) a je jím postižena jedna třetina druhů (23) evropského kontinentu. Nicméně vzhledem k nedostatečné pozornosti, může být její přítomnost v některých dalších státech zatím nezjištěna (Civiš et al. 2010).

2.1.3 Chytridiomykóza v ČR

Jak již bylo zmíněno v úvodu, v České republice byla přítomnost *Bd* poprvé zaznamenána v roce 2008 (Civiš et al. 2010). Od roku 2009 je na našem území prováděn pravidelný monitoring, na kterém se podílejí Civiš, Vojar a Baláz z pracovišť ČZU Praha a VFU Brno (Civiš et al. 2010). Výskyt patogenu v ČR byl zatím prokázán u sedmi druhů, nicméně ve střední Evropě byla zjištěna jeho přítomnost u 19 z 21 druhů vyskytujících se v ČR. Zasaženými druhy na našem území jsou: *Bufo bufo* (ropucha obecná), *Ichthyosaura alpestris* (čolek horský), *Bombina bombina* (kuňka obecná), *Bombina variegata* (kuňka žlutobřichá) a celý rod *Pelophylax* (rod zelených skokanů). Z uvedených druhů je nejvíce zasažen rod *Bombina*. Prozatím není rozšíření *Bd* v ČR nijak alarmující (9 % vzorkovaných jedinců) a na osmi lokalitách s prokázaným výskytem nebyl zaznamenán žádný úhyn jedinců (Civiš et al. 2012). Tento zatím malý projev chytridiomykózy v ČR je patrně způsoben zdejším klimatem, jenž je

pro vývoj a přežívání zárodků *Bd* méně příznivý, v porovnání s klimaticky teplejšími a vlhčími oblastmi jakými jsou tropy a subtropy (Berger et al. 1998).

2.2 Modelování prostorového rozšíření organismů

2.2.1 Základy modelování prostorové distribuce

Modelování distribuce jedinců/populací/druhů v prostoru, (v angličtině species distribution modelling čili SDM, či mapping species distribution) se zakládá na statistických modelech týkajících se rozmístění dat v prostoru a ekologických procesech v přírodě (Elith & Leathwick 2009). Tyto predikční modely využívají korelace mezi výskytem jedinců a podmínkami prostředí, v němž se jedinci vyskytují (Franklin & Miller 2009). Mohou přitom brát v úvahu buď pouhou přítomnost jedinců na stanovišti, nebo mohou být založeny na míře „využitosti“ stanoviště (jeho upřednostňování oproti jiným stanovištím) danými jedinci. Metoda porovnávající důležitost faktoru dostupnosti stanoviště versus jeho využitosti je v angličtině nazývána resource selection function zkráceně RFS (Elith & Leathwick 2009).

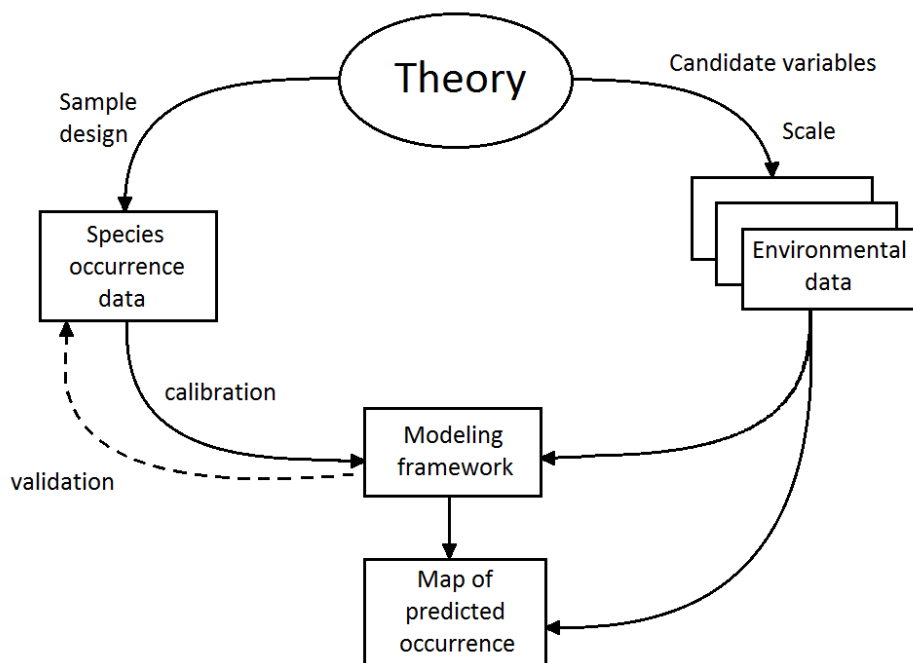
Prostorové modelování je dle Austina (2002) založeno na třech klíčových částech celkového modelu, které mohou být brány jako samostatné modely. Tyto tři části představují ekologický model, datový model a statistický model. Ekologický model sestává z ekologických poznatků a z teorie, která je v dané studii použita nebo testovaná. Datový model představuje souhrn rozhodnutí, která byla utvořena s ohledem na sběr dat a na to zda byla data změřena či odhadnuta. Statistický model zahrnuje výběr metody statistického testování, míru chyby a signifikance zvoleného testu. Tyto základní předpoklady pro prostorové modelování se dají rozvést do následujících kroků (Franklin & Miller 2009):

- Výstavba teoretického modelu abiotických a biotických faktorů ovlivňujících rozšíření druhů v prostoru a čase, který je posouzen z několika různých pohledů v závislosti na zkoumané problematice.

- Posouzení přesnosti dat o rozmístění druhů v prostoru (geografické souřadnice) a jejich charakteru (prezence nebo nepřítomnost, využívanost biotopu, početnost, atd.).
- Zajištění mapových podkladů reprezentujících vlastnosti biotopu, nebo které korelují s jeho kvalitami. Tyto jsou odvozeny od modelů představujících procesy probíhající v životním prostředí nebo jsou jejich přímým výsledkem. Zdrojem těchto mapových podkladů jsou různé geoportály sloužící nástrojům GIS.
- Sestavení modelu, který spojuje nároky druhu na prostředí, nebo jeho využívanost druhem, s vlastnostmi které může toto prostředí danému druhu nabídnout. Model může být statistický, deskriptivní, logický či jiný (Burgman et al. 2005).
- Použití nástrojů pro převedení modelu do mapové podoby zobrazující výsledné hodnoty modelu ve viditelné snadněji srozumitelné škále (barvy, stínování), především s využitím nástrojů GIS.
- Zhodnocení výstupních dat a kritérií ověřujících korektnost predikce, popřípadě zjištění správné interpretace chyb a nepřesností provedené analýzy.

Provázanost těchto jednotlivých kroků dobře znázorňuje diagram v knize Mapping Species Distribution (Franklin & Miller 2009)

Obr. č. 1: Diagram znázorňující komponenty prostorového modelování .



©Franklin & Miller 2009

2.2.2 Typy dat a práce s nimi

Modelování prostorové distribuce je od svého počátku spojené s použitím výpočetní techniky. Díky rozmachu tohoto nástroje bylo vyvinuto množství softwarů, včetně statistických programů, které vytváření predikčních modelů značně usnadňují. Výběr vhodného programu je podmíněn typem dostupných dat a metodou jejich hodnocení (Austin 2002).

Povaha dat zpracovávaných při modelování může mít různý charakter, a to v případě vysvětlované i vysvětlujících proměnných. Závislá (vysvětlovaná) proměnná může nabývat jak spojitých hodnot (hustota, počet) tak i hodnot kategoriálních (přítomnost/absence). Totéž platí i v případě vysvětlovaných proměnných, které mohou být charakterizované jako kvantitativní data na poměrové (např. délka, váha jedinců), intervalové (stupně Celsia),

ordinální stupnici (např. klasifikační stupně, Lepš 1996) či jako data kategoriální (např. typy půd či hornin, Franklin 1995).

Jak uvádí ve své práci Hastie et al. (2009), každé modelování distribuce by se mělo zakládat na „učení se z dat“ („learning from data“), což znamená, že každý jednotlivý model má být založen na vlastní jedinečné modelovací metodě, která by měla odpovídat charakteru dat. Tyto data by měla být statisticky testována a analyzována. Z výsledků, které tímto získáme, by měla být navržena nová metoda modelování, pouze a jen pro tato data. Před samotnou analýzou dat a před hodnocením zvolené či navržené metody modelování je proto důležité zajistit další sadu dat podobného charakteru. Druhá sada dat poslouží jako tréninková data („training data“) zatím co „ostrá data“, která jsou objektem zkoumání, jsou označována jako testovaná data („testing data“). Původ tréninkových dat může být zcela odlišný od našich testovaných dat, přičemž by měl být dodržen (ale není to striktním pravidlem) jejich podobný charakter, nebo se tréninková data získávají oddělením části testovaných dat. V současné době, kdy jsou data v široké škále vědních oborů založená na mnoha pozorováních a mnoha proměnných, je obzvláště důležité stavět predikční modely na poznacích získaných z pozorování charakteru sebraných dat a postup „učení se z dat“ je téměř nezbytný ke správnému vytvoření a kalibraci výsledného predikčního modelu (Franklin & Miller 2009).

2.2.3 Metody modelování a software

Metody predikčního modelování lze rozdělit na tři základní skupiny, a to na statistické metody, metody strojového učení („machine-learning methods“) a metody založené pouze na informaci o prezenci druhu („presence-only data“). Toto rozdělení je převzato od Franklin & Miller (2009), ale není to jediný způsob, podle kterého lze modelovací metody dělit.

2.2.3.1 Statistické metody

Do skupiny statistických metod jsou řazeny lineární modely na bázi regrese, obecné lineární modely (GLM) založené na moderních regresních metodách, obecné aditivní metody (GAM) a mnohorozměrné adaptivní regresní křivky (multivariate adaptive regression splines, MARS) a další (Crawley 2007). Tyto metody jsou v oblasti modelování distribuce používány od osmdesátých let dvacátého století a patří k nejhojněji používaným metodám v prostorovém modelování a modelování obecně. Jejich princip je založen na tzv. „učení s učitelem“ (supervised learning), které lze charakterizovat jako pozorování proměnných a následné vybrání modelu, který bude použit pro testování tréninkových dat. Na základě analýz tréninkových dat je zhodnocena přesnost predikce vybraného modelu a ten je v případě vyhovujících parametrů použit na „ostrá data“. Tento postup je proto založen na zkušenostech člověka, který modelování provádí (Franklin & Miller 2009).

a) Lineární a zobecněné lineární modely (GLM)

Lineární modely pracují s daty o normálním rozdělení vysvětlované proměnné. Vytvářejí o struktuře dat a měření stabilní, ale pravděpodobně nepřesné predikce, vzhledem k lineární povaze vztahu závislosti mezi závislou a nezávislou proměnnou (Hastie et al. 2009). V případě kvantitativních dat je závislá proměnná (obvykle značena Y) vysvětlována nezávislou proměnnou (X), přičemž jejich vztah lze znázornit přímkou, na které se nacházejí tzv. odhady či predikované hodnoty. Díky tomu jsme schopni odhadnout i neměřené hodnoty závislé proměnné Y , nacházející se v rámci měřených hodnot (interpolace) či mimo jejich rozsah (extrapolace), protože se s určitou pravděpodobností budou nacházet na regresní přímce (Crawley 2007).

Nicméně jeden z předpokladů pro použití lineárních modelů, tedy normalita rozdělení dat vysvětlované proměnné, resp. normalita jejich residuálů (odchylek skutečných hodnot od predikovaných, Pekár & Brabec 2009), je velmi často narušen. Rozšířením lineárních modelů robustních vůči těmto předpokladům (tj. tyto předpoklady nemusí být splněny) jsou zobecněné lineární modely, které dokážou pracovat s vysvětlovanými proměnnými s rozdělením odlišným od normálního (Venables & Ripley 1994). GLM poskytují alternativu

k postupu nutné transformace dat a následnému použití lineárního modelu. GLM pracují s normálním, Poissonovým, binomickým, negativně binomickým a gamma rozdělením (Crawley 2007). Záleží tedy na vysvětlované proměnné, kterým rozdělením bude aproximována. Jinými slovy, pokud bude mít vysvětlovaná proměnná binomický charakter (např.: živý/mrtvý, přítomný/nepřítomný), bude tato proměnná aproximována binomickým rozdělením. Díky této toleranci k charakteru dat jsou GLM jednou z nejvíce používaných metod využívaných k prostorovému modelování distribuce a dle Franklin & Miller (2009) je užití GLM předmětem mnoha vědeckých prací zabývajících se širokou škálou živočišných, ale i rostlinných druhů.

b) Obecné aditivní modely (GAM)

Obecné aditivní modely představují neparametrické rozšíření obecných lineárních modelů, od kterých se liší schopností určit a popsat nelineární vztah mezi vysvětlující a závislou proměnnou a též schopností charakterizovat povahu závislé proměnné (Yee & Mitchell 1991). Pro GAM platí, stejně jako pro GLM, že závislá proměnná a její rozdělení náleží do skupiny stejné jako u GLM (normální, Poissonovo, binomické, negativně binomické nebo gamma rozdělení). GAM byly původně navrženy jako grafická metoda pro zjišťování a popsání nelineárních závislých funkcí v datech, na základě jejichž výsledků byl následně vystaven zobecněný lineární model. Touto cestou mohou být GAM využity v mnohonásobném regresním modelování (Franklin & Miller 2009).

2.2.3.2 Metody strojového učení

Tyto metody jsou založené na užití algoritmů, které jsou schopné přímo z tréninkových dat určit, který model bude mít nejlepší výsledky. Metody strojového učení mají tedy větší flexibilitu a nejsou závislé na zkušenostech a vědomostech člověka provádějícího analýzu (Gahegan et al. 2003). Mezi tyto metody řadíme užití tzv. umělých neuronových sítí (neural network), rozhodovacích stromů (decision trees, tree-based methods) či maximální entropie (maximum entropy či MaxEnt). Používání machine learning metod v ekologii je méně časté než použití regresních modelů. Je to dáno všeobecně menšími znalostmi biologů v oblasti programování algoritmů potřebných k vytváření těchto modelů. Proto je tato metoda převážně doménou lidí zabývajících se statistikou (Elith et al. 2008)

a) Klasifikační stromy

Franklin & Miller (2009) označuje užití metod klasifikačních stromů jako velmi užitečné v případech, kdy vysvětlující proměnné nabývají hodnot, jak spojitých, tak kategoriálních, a zvláště pak v případech, kdy kategoriální proměnné nabývají více jak dvou hodnot. Základním principem klasifikačních stromů je rozdělení dat do homogenních podskupin, kde proměnné nabývají podobných hodnot (spojité proměnné) nebo náleží do stejné kategorie (kategoriální proměnné). Rozdělení do podskupin lze názorně ukázat v Tab. č. 1.

Podskupiny jsou následně užití ve třech stupních, a to v tzv. výstavbě stromu (tree building či growing), zastavení výstavby stromu (tree stopping) a v následném klestění stromu (tree pruning nebo též optimal tree selection), (Olden et al. 2008). Struktura stromu je rozdělena na patra dle podskupin, jež binárně dělí dle kritérií, která se předem stanoví. Struktura se takto větví až ke konečnému výsledku. Tento popis je jen ilustrativní, podrobněji lze problematiku nastudovat z Olden et al.(2008), Franklin & Miller (2009) nebo Hastie et al. (2009).

Tab. č. 1: Rozdělení dat do podskupin v principu klasifikačních stromů.

Y (P/A)	X ₁ (půda)	X ₂ (nadmořská výška)
0	A	930
0	A	1100
1	A	650

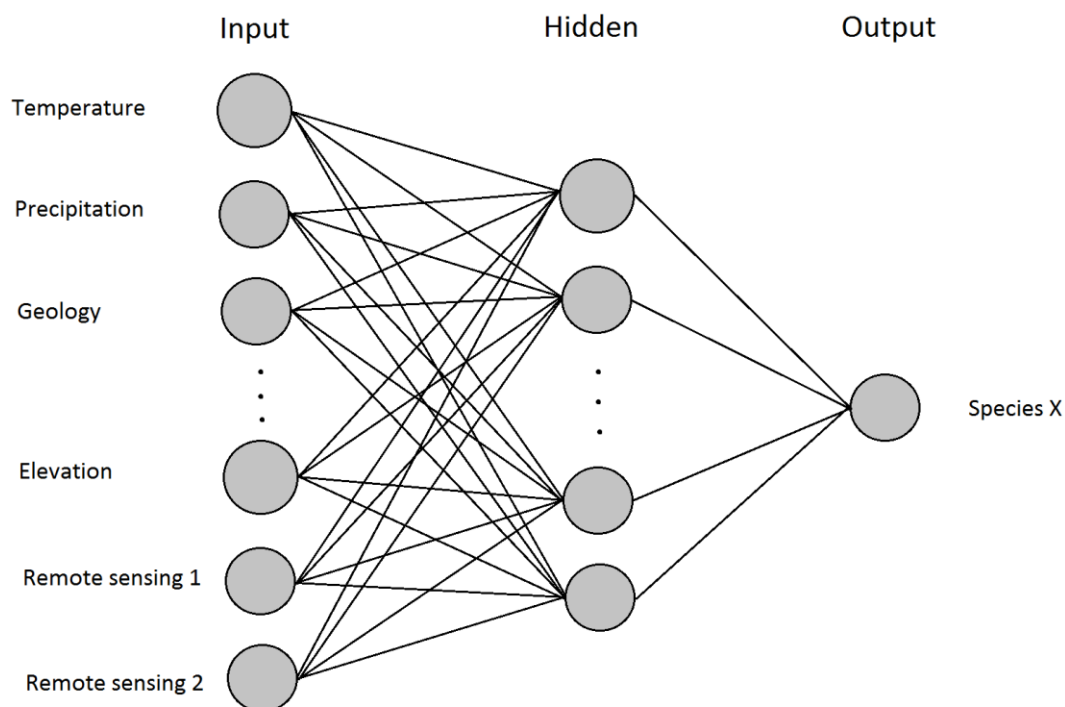
1	B	880
1	B	700
0	B	590
1	B	630

Závislá proměnná Y vypovídá o přítomnosti (1) či absenci (0) druhu. V tomto případě byly hodnoty rozděleny do dvou skupin dle kategoriální vysvětlující proměnné X₁ a následně se porovnává směrodatná odchylka proměnné Y před rozdělením a po něm což nám ukáže tzv. „míru čistoty“ (purity of class membership) rozdělení. Tento postup se provádí následně i dle proměnné X₂. Dle „míry čistoty“ se pak vybere vhodnější způsob rozčlenění. Takto by vypadalo rozdělení dat do jednoho stupně ve struktuře stromu (©Franklin & Miller 2009).

b) Umělé neuronové sítě

Pojmenování této metody je dáno původním účelem, pro který byla vyvinuta, a to pro modelování pochodů probíhajících v lidském mozku. Gahegan (2003) ve své práci předkládá, že rozdělení používaná ve statistickém modelování distribuce neposkytují dostatečně kvalitní výsledky, zejména v případech mnoha-rozměrných problémů. Částečně je to způsobeno „prokletím dimenzionality“ (curse of dimensionality). Většina environmentálního prostoru je prázdná, proto je obtížné aplikovat parametrickou závislou funkci na mnohorozměrný prostor. Neuronové sítě i klasifikační stromy rozdělují data do podskupin, a soustřeďují se na jejich informativní podstatu, přičemž ignorují prázdná místa v prostoru, a díky tomu pracují v těchto podmínkách lépe než metody zmíněné výše. Základním postupem této metody je vytvoření skryté proměnné (hidden) na základě lineárních vztahů mezi vysvětlujícími proměnnými, z kterých je následně modelován výstup, jako výsledek nelineárních či lineárních funkcí. Jednoduchý neuronový model je zobrazen na Obr. č. 2.

Obr. č. 2: Příklad jednoduché neuronové sítě.



Vytvoření skryté (hidden) proměnné z lineárního vztahu mezi vloženými (input) vysvětlujícími proměnnými a následné vypočítání výstupních hodnot závislé proměnné, představujících lineární funkci proměnných hidden. Převzato z Franklin & Miller (©2009).

c) Maximální entropie (Maximum Entropy Modelling)

Maximální entropie (MaxEnt) je také klasifikována jako metoda strojového učení, přičemž je používána i v tak odlišných vědních odvětvích jako jsou astronomie a ekonomika (Franklin & Miller 2009). Vzhledem k povaze dat, se kterými MaxEnt pracuje (data o přítomnosti – presence-only data), bude tato metoda podobněji popsána v následující podkapitole a zejména pak v samostatné části (Kap. 2.3).

d) Ostatní metody

Mezi metody strojového učení, které se v oboru prostorového modelování začaly využívat v posledních desetiletích, patří i další metody, které v této práci nejsou podrobněji zmiňovány. Kromě klasifikačních stromů, umělých neuronových sítí a MaxEnt sem lze zařadit

metody založené na genetických algoritmech, podpůrných vektorových mechanismech, předpovědních souborech či konsenzové metody (Franklin & Miller 2009).

2.2.3.3 Metody využívající pouze data o přítomnosti

S pomocí výše uvedených metod se převážně vypracovávají predikce pro vysvětlovanou proměnnou binárního charakteru, která vypovídá o přítomnosti či absenci druhu na stanovišti. Metody, které budou popsány v následující podkapitole, pracují s daty, které udávají pouze informaci o přítomnosti druhu (presence-only data), a nikoliv o jeho absenci. Franklin & Miller (2009) ve své práci rozděluje tyto metody na tři skupiny:

- Modely porovnávající vlastnosti prostředí nových stanovišť a stanovišť, kde byl druh pozorován. Jako nejpoužívanější zde bude uvedena metoda „environmentálních obálek“ (environmental envelope methods, EE).
- Modely založené na presenci druhů na stanovišti v porovnání s jeho dostupností, která je stanovena na základě porovnání s jinými stanovišti či podobnou krajinou. Zde je možné použití nástrojů GIS.
- Modely tzv. „vhodného habitatu“ na základě zkušeností člověka provádějícího analýzu a následné definování mapovacích funkcí pro environmentální prediktory.

2.2.4 Programy pro modelování prostorové distribuce

V současné době existuje mnoho programů, které slouží k modelování prostorové distribuce druhů. Tyto programy používají různé metody, z nichž několik základních je uvedeno v předchozí části práce. Stručný přehled programů používaných k modelování prostorové distribuce spolu s metodami, které využívají je uveden v Tab. č. 2.

Tab. č. 2: Příklad softwarů užívaných v modelování prostorové distribuce.

Program	Metoda	Odkaz
BIOCLIM	Environmentální obálky	http://www.worldclim.org/bioclim
GARP	Genetické algoritmy	http://www.nhm.ku.edu/desktopgarp/
MARS	Mnohorozměrné adaptivní regresní křivky	http://www.salford-systems.com/products/mars
MaxEnt	Maximální entropie	www.cs.princeton.edu/~schapire/maxent
PRESENCE	Logistická regrese	http://www.mbr-pwrc.usgs.gov/software/presence.html

Franklin & Miller 2009

2.3 MaxEnt

2.3.1 Princip maximální entropie

Maximální entropie (MaxEnt) je princip, který pochází ze statistické mechaniky a informační teorie (Jaynes 1957). Princip maximální entropie nám říká, že nejlepším způsobem aproximace neznámého rozdělení je rozdělení s maximální entropií (nejvíce rozprostřené či nejbližší uniformnímu) podléhající známým omezením. Tato omezení jsou definována očekávanou hodnotou distribuce, která je odhadnuta z dat o pozorovaném výskytu (Franklin & Miller 2009). Jak již bylo řečeno je to metoda strojového učení, která byla vyvinuta na práci s daty založenými pouze na přítomnosti druhu (presence-only data), (Dudík et al. 2007).

2.3.2 Software MaxEnt

MaxEnt je volně stažitelný software (na www.cs.princeton.edu/~schapire/maxent), který je schopen provádět výpočty „Maximální entropie geografické distribuce druhů“ ("Maximum entropy modelling of species geographic distributions") a byl vytvořen následujícími autory: Steven Phillips, Miro Dudík a Rob Schapire, s podporou AT&T Labs-Research, Princeton University a Center for Biodiversity & Conservation, American Museum of Natural History. Klíčová úloha tohoto programu spočívá v tom, že počítá predikce geografické distribuce druhů jen za použití záznamů o přítomnosti. Jak dobře vysvětluje Civiš (2011) ve své práci, „Software MaxEnt používá metody optimalizace tak, aby našel nejvhodnější model, který předpovídá pravděpodobnostní distribuci druhu napříč geografickým prostorem jako funkci environmentálních proměnných spojených s tímto prostorem“.

3. Metodika práce s programem MaxEnt

Vytvoření jednoduchého predikčního modelu je pomocí softwaru MaxEnt vcelku snadné, a to díky předem definovaným algoritmům pro výpočet modelu, za užití principu maximální entropie. V následující části práce bude popsán postup pro vytvoření jednoduchého modelu prostorové distribuce v programu MaxEnt. Důležité pojmy v popisu práce s programem budou kvůli přehlednosti tučně zvýrazněné.

3.1 Příprava dat pro MaxEnt

Data zpracovávaná v prostorovém modelování distribuce mohou mít různou podobu. Záleží na metodice užití k jejich sběru a také na rozsahu sesbíraných informací, které vypovídají o jednotlivých vzorcích (v angličtině samples). V dnešní době je elektronická podoba dat naprostou samozřejmostí a data z průzkumu či monitoringu jsou po roztřídění a zpracování zaznamenána ve formě tabulek nejčastěji pomocí aplikace Excel z balíčku Microsoft Office. Takto zpracovaná data obsahují informace o vysvětlované proměnné a ve většině případů udávají přítomnost, či nepřítomnost zkoumaného organismu a jeho geografickou polohu (Franklin & Miller 2009). Totéž platí i v případě programu MaxEnt, který pracuje pouze s údaji o přítomnosti zkoumaného objektu a s jeho zeměpisnou polohou (Dudík 2007). Tabulku s daty je nutné mít ve formátu .csv a ve stanovené struktuře. Ukázka podoby uspořádání dat vysvětlované proměnné v souboru s koncovkou .csv zobrazuje Obr. č. 3.

Obr. č. 3: Podoba zápisu vysvětlované proměnné pro MaxEnt.

	A	
1	species,dd long,dd lat	
2	microryzomys_minutus,-64.7,-17.85	
3	microryzomys_minutus,-67.8333,-16.3333	
4	microryzomys_minutus,-67.8833,-16.3	
5	microryzomys_minutus,-67.8,-16.2667	

Data sestávají ze tří údajů: druh (species), zeměpisná délka (longitude), zeměpisná šířka (latitude). V souboru jsou zapsána v jednom sloupci a údaje jsou oddělené čárkou (<http://www.cs.princeton.edu/~schapire/maxent/>, Tutorial data).

Vysvětlující proměnné jsou zastoupeny souborem mapových vrstev (v angličtině Environmental layers), které nám udávají vlastnosti prostředí (např. průměrná teplota, srážky, nadmořská výška, přirozená vegetace či typy půd), (Franklin & Miller 2009). Získání těchto mapových podkladů je možné na některých geoportálech či internetových stránkách nabízejících mapové podklady k vědeckým účelům. Příkladem může být geoportál agentury CENIA (<http://www1.cenia.cz/www/>), geoportál ČÚZK (<http://geoportal.cuzk.cz/>) či celosvětová databáze WorldClim (<http://www.worldclim.org/>). Získání některých konkrétních mapových podkladů však může představovat problém a pro jejich obstarání je nutná jejich koupě např. na stránkách ČÚZK. Software MaxEnt klade několik podmínek na podobu a vlastnosti mapových podkladů (Phillips, MaxEnt Tutorial):

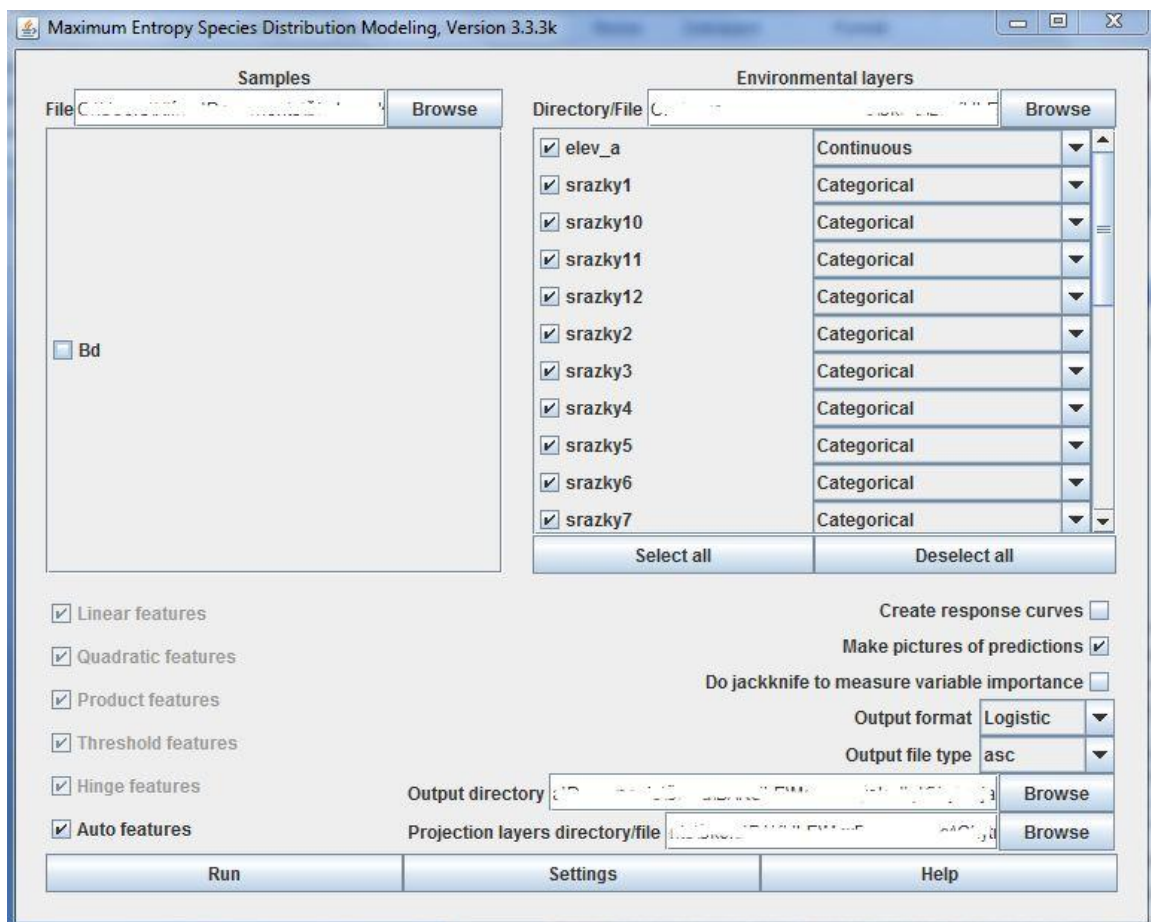
- Veškeré mapové vrstvy vkládané do MaxEnt musejí mít jednotný geografický systém souřadnic; v případě ČR je nejvýhodnější S-JTSK, pro zbytek světa je nepoužívanějším souřadnicovým systémem WGS 84. Souřadnicový systém by měl být shodný nejen mezi mapovými vrstvami, ale i mezi vrstvami a údaji o poloze vysvětlovaných proměnných.
- Veškeré mapové podklady musejí být jednotné v rozlišení a velikosti, je tedy výhodné vytvořit si „masku“, podle které jsou mapové vrstvy jednotně ořezávané (např. hranice státu, kontinentu, hranice pevniny). K těmto úpravám je nejvhodnější použít nástroje ArcGIS, a to funkce Extract by Mask (maska má rasterový formát)
- Formát mapových podkladů pro software MaxEnt je .asc (ASCII). Je tedy nutné převést vrstvy na tento formát pomocí Raster to ASCII, a v případě vektorových map je potřeba před tímto krokem užít funkci Topo to Raster (např. převod vrstevnic na rastrový obraz).
- Pro import mapových vrstev do programu je potřebné, mít tyto vrstvy umístěné v jedné složce.

Pro model vytvořený v rámci této práce byla data o vysvětlované proměnné uložena v souboru s názvem **Bd_MX.csv**, obsahující 17 lokalizací výskytu *Bd*. Jako údaj „**species**“ bylo použito **Bd** (*Batrachochytrium dendrobatidis*). Vysvětlující proměnné byly zastoupeny 12 mapami průměrných měsíčních srážek, 12 mapami nejnižších teplot v měsíci a mapou reliéfu, která byla vytvořena z vektorové mapy vrstevnic. Veškeré mapové podklady se týkají území České republiky a použitý souřadnicový systém je S-JTSK.

3.2 Import dat a základní nastavení modelu

Po spuštění programu MaxEnt se otevře základní okno programu (Obr. č. 4), kde prvním krokem je nainportování dat, které budou použity k modelování.

Obr. č. 4: Úvodní okno programu MaxEnt po importu dat.



S pomocí záložky „**Samples**“ je nutno vyhledat tabulku ve formátu .csv, která obsahuje informace o závislé proměnné (viz Obr. č. 4) a nahrát ji do MaxEnt. V poli pod lištou Samples se objeví jméno druhu, které je zadané v .csv souboru (v případě Obr. č. 3 by se jednalo o microryzomys_minutus, v případě této práce Bd, Obr. č. 4).

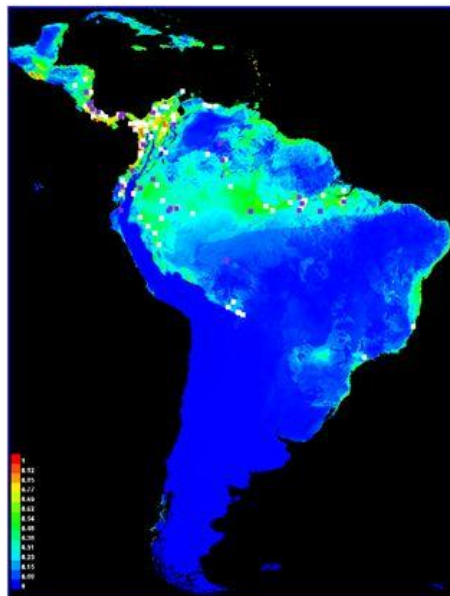
V záložce „**Environmental layers**“ vyhledáme a následně nahrajeme složku obsahující veškeré naše mapové podklady, které chceme použít v modelu jako vysvětlující proměnné. Tyto podklady mohou být kategoriální (mapy klimatu, teplot, srážek, typů půd) nebo mohou mít spojitý charakter (reliéf terénu, intenzita dopravy). Po nahrání dat do programu je nutné vybrat typ proměnné spojitá/kategoriální (Categorical/Continuous, viz Obr. č. 4). V případě modelu pro *Bd* jsou veškeré použité vrstvy kategoriálního charakteru, kromě vrstvy reliéfu.

Dále je třeba zvolit formát výstupních dat. Lištou „**Output format**“ se nastavuje typ modelu, který bude vytvořen na základě našich dat (Logistic, Raw, Cumulative). Základním typem modelu je Logistic, který je doporučován jako nejpřesnější (bez výrazného nadhodnocení či podhodnocení predikce). Dalším důvodem proč je Logistic doporučován jako základní typ je snadná interpretace výsledného modelu, jelikož pravděpodobnost je zaznamenána jako interval od 0 do 1. Na základě těchto předností, byl typ modelu Logistic vybrán i ke zpracování predikce výskytu *Bd*. Další lištou s názvem „**Output file type**“, je možné nastavit výstupní formát mapové vrstvy, která lze dále využívat v nástrojích GIS (.asc, .mxe, .grd a .bil).

Volba umístění výstupních dat se provádí v liště „**Output directory**“. Výstupními daty, která jsou pro zhotovovatele modelu nejdůležitější je složka „**plots**“, která obsahuje grafické výstupy, soubor ve formátu .html, nesoucí jméno shodné s údajem „**species**“ (v našem případě Bd.html) a stejně pojmenovaný soubor ve formátu, který lze dále využít v práci s nástroji GIS (viz Output file type). Dále je ve výstupních datech obsaženo několik dalších souborů, které se uplatňují v dalším testování a upravování modelu, ale v rámci této práce není potřeba s jejich bližším seznámením. K nastavení výstupních dat patří i tři volby nalézající se v pravé části základního okna, a to „**Create response curves**“, „**Make pictures of prediction**“ a „**Do jackknife to measure variable importance**“ (viz Obr. č. 4). Tyto tři volby nám umožňují vytvoření grafických výstupů modelu, které se ukládají do výstupní složky s názvem plots (viz výše). Volba Make pictures of prediction je po spuštění programu

nastavena základně a jejím výsledkem je mapka zobrazující výslednou predikci pravděpodobného výskytu. (viz Obr. č. 5). Funkce Create response curves nám vytvoří grafy, které zobrazují, jak předpovídaná pravděpodobnost výskytu závisí na hodnotách environmentálních proměnných. Poslední funkce s názvem Do jackknife to measure variable importance, vytvoří pruhový graf znázorňující míru důležitosti každé vysvětlující proměnné, je-li tato proměnná použita jako jediná vysvětlující nebo naopak je-li z modelu vynechána.

Obr. č. 5: Příklad grafického výstupu prostorové distribuce.



Větší pravděpodobnost výskytu je zobrazena „teplejšími“ barvami menší pravděpodobnost naopak „chladnějšími“ (©Phillips MaxEnt, Tutorial).

Po předchozích krocích je v základním okně možné dále vybrat typ funkce, která bude uplatněna jako vztah mezi environmentálními proměnnými a vysvětlovanou proměnnou při tvorbě modelu. Na výběr jsou: lineární, kvadratické, sčítací (product feature), prahové (threshold) a „pantové“ (hinge feature) funkce (viz Obr. č. 4). Z nich lze vybírat jednotlivě nebo v definovaných kombinacích. Po zapnutí programu je přednastavena automatická funkce (**Auto feature**), která sama na základě počtu vzorků a environmentálních proměnných aplikuje nejvhodnější funkci pro vytvoření modelu. Tato automatická volba byla použita i pro vytvoření predikce rozšíření *Bd*. Podrobněji lze problematiku výběru použitého typu funkce nastudovat v nabídce Help (nápopěda v hlavním okně programu) či z literatury některého z autorů programu (Phillips 2006, Phillips 2008, Elith 2011).

Kromě nastavení v základním okně, byla provedena jediná úprava pro model výskytu *Bd*. V nabídce „**Settings**“ byla v první záložce „**Basic**“ změněna hodnota v kolonce „**Random test percentage**“. Nastavením této hodnoty se určí, kolik procent z údajů vložených do Samples bude použito jako testovacích bodů, které budou sloužit k vypočtení hodnoty AUC, na jejímž základě je hodnocena přesnost modelu.

Po nastavení těchto základních hodnot a funkcí je program připravený k vytvoření modelu a zahájení výpočtů se spustí pomocí tlačítka „**Run**“, po kterém lze sledovat průběh vytváření modelu v informačním okně a práce programu je dokončena ve chvíli, kdy informační okno zmizí z obrazovky. Následně stačí otevřít složku, která byla nastavena jako cílová v liště Output directory a zhodnotit výstupy programu.

3.3 Zvolené nastavení pro predikci rozšíření *Bd*

Charakter proměnných a nastavení programu, které bylo zvoleno pro vytvoření modelu výskytu *Bd* v ČR, lze shrnout do následujících bodů:

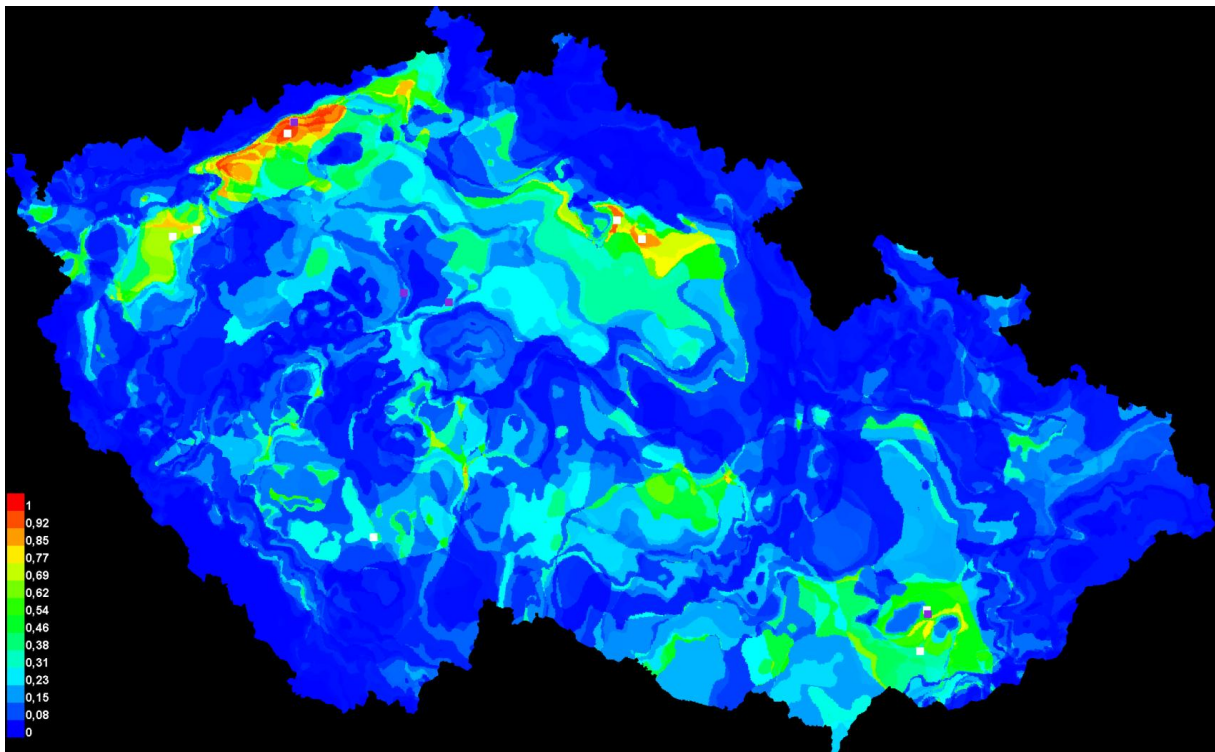
- Samples – 17 údajů o přítomnosti *Bd* v ČR
- Environmentální proměnné – 12 teplotních map, 12 srážkových map a 1 mapa reliéfu
- Použité funkce – Auto features
- Formát výstupního modelu – Logistic
- Formát výstupní mapové vrstvy – .asc (ASCII)
- Hodnota Random test percentage – 25 %

Další možnosti nastavení programu jsou k dispozici pod záložkou Settings v základním okně MaxEnt. Pro potřeby této práce zcela postačuje nastavení provedené v základní nabídce programu a výsledkem bude jednoduchý model predikce výskytu patogenu v ČR. Způsob zhodnocení přesnosti výsledného modelu bude popsán v části Výsledky (Kap. 4.)

4. Výsledky práce

Mapa predikce výskytu *Bd* na území ČR, která byla vytvořena programem MaxEnt, na základě poskytnutých dat, je zobrazena na Obr. č. 6. Tento obrázek je ilustrativní a jeho podrobnější podoba je připojena v přílohách práce.

Obr. č. 6: Mapa predikce výskytu *Bd* na území ČR.

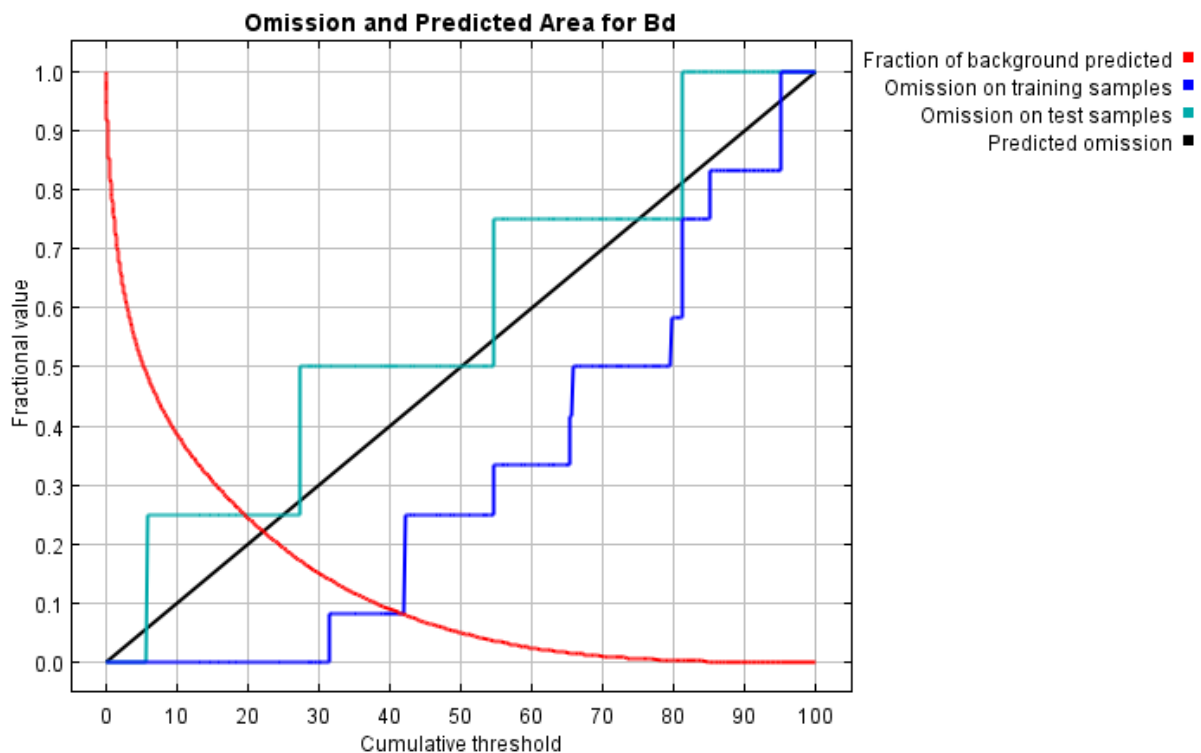


Barevná škála nám zobrazuje pravděpodobnost předpovídaného výskytu patogenu. „Teplejší“ odstíny barev zobrazují místa s větší pravděpodobností výskytu *Bd* a naopak místa s menší pravděpodobností jsou zobrazena „chladnějšími“ barvami. Vzorky použité k modelování jsou zobrazeny bílými a fialovými čtverečky (fialové zobrazují náhodně vybrané testové body, viz Random test percentage).

Na základě tohoto obrázku je možné odhadnout, kde jsou podmínky prostředí vhodné pro výskyt patogenu. Z mapky je patrné, že výsledná predikce koreluje s daty o výskytu *Bd*, vzhledem k vyšší zobrazované pravděpodobnosti výskytu v místech, kde byla hlášena přítomnost patogenu.

Dalšími důležitými výstupy programu MaxEnt jsou grafy sloužící ke statistické analýze a znázorňující kvalitu výsledného modelu. První graf (Obr. č. 7) znázorňuje míru chyby ze zanedbání (omission) pomocí třech křivek, a to křivky předpovídané chyby (černá, Predicted omission), chyby testovacích bodů (tyrkysová, Omission on training samples) a tréninkových bodů (tmavě modrá, Omission on training samples). Míra přesnosti je posuzována dle vzdálenosti testové a tréninkové křivky od přímky předpovídané chyby.

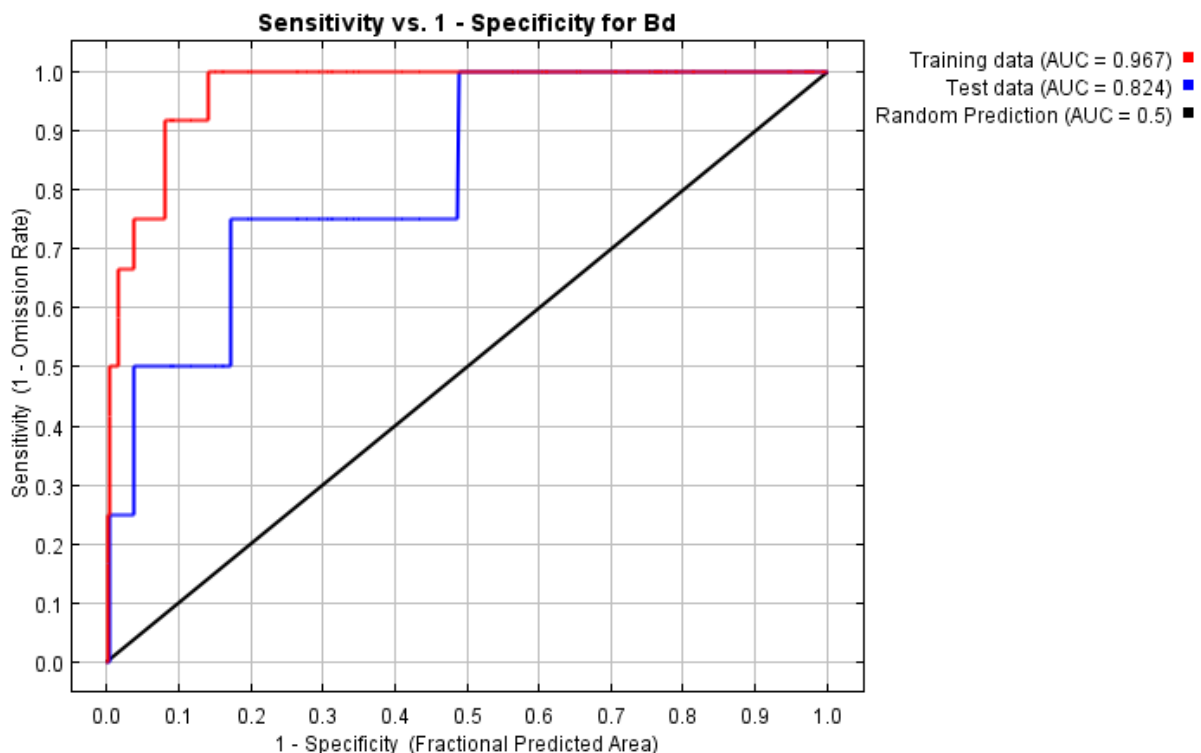
Obr. č. 7: Graf znázorňující míru chyby modelu.



Pozice tréninkové křivky se ve většině případů odchyluje směrem dolů od předpovídané chyby. V případě modelu pro *Bd* není velikost tohoto odchylení zcela uspokojivá, ale ani alarmující. Oproti tomu vzdálenost testové křivky od křivky předpovídané chyby není nijak velká, z čehož lze usoudit, že přesnost modelu je vcelku dobrá. Může nastat situace, že testová křivka klesne hluboko pod křivku předpovídané chyby, v tom případě by to znamenalo, že testové a tréninkové body nejsou nezávislé, což by značně ovlivnilo objektivnost výsledku predikce.

Druhý graf (Obr. č. 8) zobrazuje křivky ROC (Receiver Operating Curve), které udávají vztah mezi specificitou a senzitivitou diagnostické metody použitého modelu pro veškeré přípustné hodnoty prahu (threshold). Prostor pod křivkou ROC se nazývá AUC (Area Under Curve) a jeho velikost udává míru přesnosti modelu. Kvalita modelu se s rostoucí hodnotou AUC zvětšuje a naopak. Křivka pro testovací data (tmavě modrá) a její hodnota AUC, je hlavní pro určení přesnosti vytvořeného modelu.

Obr. č. 8: Graf zobrazující ROC s vypočtenými hodnotami AUC.



Vzhledem k hodnotě AUC, pro testová data modelu rozšíření *Bd*, která je 0,824 (viz Obr. č. 8) lze tento model hodnotit jako dobrý. Přičemž stupnice AUC, dle které je kvalita testu hodnocena je:

- 0,50 až 0,75 = oprávněný/použitelný
- 0,75 až 0,92 = dobrý
- 0,92 až 0,97 = velmi dobrý
- 0,97 až 1,00 = vynikající

Toto je základní zhodnocení výstupních hodnot modelu rozšíření *Bd* v České republice. Dále je možné, po zaškrtnutí možností *Create response curves*, a *Do jackknife to measure variable importace* v základním menu programu, vytvořit grafy hodnotící vliv jednotlivých vysvětlujících na rozmístění vysvětlovaných proměnných v prostoru. V případě této práce se analýzou vlivu environmentálních proměnných nebudeme dále zabývat. Pro vytvoření základního modelu výskytu a jeho zhodnocení postačují výše uvedená fakta.

5. Diskuze

Na základě práce Hernandez et al. (2006) lze říci, že software MaxEnt je v porovnání s programy Bioclim, Domain a GARP více výkonný v případě, kdy jsou dostupná data pouze o přítomnosti druhu. Dále bylo zjištěno, že přesnost těchto programů určených k modelování prostorové distribuce se zvětšuje, je-li sledován výskyt vzácných druhů, které jsou náročnější na kvalitu stanoviště. Jako další důležitý faktor hrající roli v přesnosti vytvářené predikce je správný výběr mapových podkladů. Tento fakt byl zdůrazněn i v případě prvního vytvořeného modelu výskytu *Bd* v ČR (Civiš os. sdělení).

V případě modelu vytvořeného v této práci bylo užito přesných environmentálních proměnných, ale většina z nich měla podobnou vypovídající hodnotu (např. použití 12 map měsíčních srážkových úhrnů). Pro vytváření predikce výskytu by mělo být využito co možná nejvíce druhů environmentálních proměnných, aby bylo docíleno co možná nejobjektivnějších výsledků. Nicméně obstarání vhodných mapových podkladů je dosti obtížné a v některých případech časově i finančně nákladné. Tento fakt hrál velkou roli i v případě předložené práce. Je tedy možné, že výsledná hodnota AUC, dle které jsou hodnoceny výstupy z MaxEnt, by byla vyšší v případě užití jiných (různorodějších) environmentálních proměnných, čímž by byla i zvýšena přesnost predikce. Nicméně úloha hodnoty AUC jako prostředku užívaného k analýze přesnosti modelu je v současnosti předmětem vědeckých diskuzí.

Používání softwarů, které vytvářejí predikce na základě naprogramovaných algoritmů, je často zpochybňováno z důvodu netransparentnosti procesů probíhajících v těchto statistických aplikacích. Někteří autoři ve svých pracích (např. Breiman 2001) označují tyto softwary jako černé skříňky (black boxes), jejichž práci na vytvářeném modelu nelze zcela analyzovat či upravit. Nicméně v případě nedostatečně obsáhlých vědomostí v oblasti statistiky a modelování prostorové distribuce představují tyto programy velké usnadnění při tvorbě predikčních modelů.

6. Závěr

Cílem této práce bylo sepsat literární rešerši pojednávající o modelování prostorové distribuce druhů a následné seznámení s modelovacím programem MaxEnt, s jehož pomocí měl být vytvořen aktualizovaný model rozšíření *Bd* v ČR.

Základní seznámení s vytvořením predikčního modelu v programu MaxEnt bylo provedeno formou popisu práce na modelu rozšíření *Bd* v České republice. S ohledem na složitost problematiky SDM bylo použito jen základních nastavení program, které nicméně postačuje k vytvoření poměrně kvalitního modelu.

Mapa pravděpodobného výskytu chytridiomykózy by mohla být využita k ochraně obojživelníků před touto nebezpečnou nemocí, která je v současnosti velkým problémem v celosvětovém měřítku. Vymezení rizikových míst a jejich monitoring by mohl přispět k prevenci rozšíření patogenu z infikovaných lokalit do dalších částí naší republiky.

Model vypracovaný v rámci této práce by se po svém zpřesnění a za využití většího počtu environmentálních proměnných dal k tomuto účelu dobře uplatnit.

7. Použitá literatura:

7.1. Literární zdroje

Asashima M., Ito Y., Chan T., Michiue T., Nakanishi M., Suzuki K., Hitachi K., Okabayashi K., Kondow A. & Ariizumi T., 2009: In vitro organogenesis from undifferentiated cells in *Xenopus*. *Developmental Dynamics* 238: 1309–1320

Austin M. P., 2002: Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157: 101–118.

Berger L., Speare R., Daszak P., Green D. E., Cunningham A. A., Goggin C. L., Slocombe R., Ragan M. A., Hyatt A. D., McDonald K. R., Hines H. B., Lips K. R., Marantelli G. & Parkes H., 1998: Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. *Proceedings of the National Academy of Sciences, USA* 95: 9031–9036.

Bielby J., Bovero S., Sotgiu G., Tessa G., Favelli M., Angelini C., Doglio S., Clare F. C., Gazzaniga E., Lapietra F. & Garner T. W. J., 2009: Fatal Chytridiomycosis in the Tyrrhenian Painted Frog. *EcoHealth* 6: 27–32.

Bosch J. & Martínez-Solano I., 2005: Chytrid fungus infection related to unusual mortalities of *Salamandra salamandra* and *Bufo bufo* in the Peñalara Natural Park (Central Spain). *Biological Conservation* 97: 331–337.

Breiman L., 2001: Statistical Modeling: The Two Cultures. *Statistical Science* 16(3): 199–215.

Burgman M. A., Lindenmayer D. B. & Elith J., 2005: Managing landscapes for conservation under uncertainty. *Ecology* 86: 2007–2017 .

Civiš P., Vojar J. & Baláž V., 2010: Chytridiomykóza, hrozba pro naše obojživelníky? *Ochrana přírody* 65(4): 18–20.

Civiš P., 2011: Výskyt a potenciální dopady chytridiomykózy v České republice. Teze k disertační práci (nepublikováno). Katedra ekologie, Fakulta životního prostředí, Česká zemědělská univerzita v Praze.

Civiš P., Vojar J., Literák I. & Baláž V., 2012: Current State of *Bd* Occurrence in the Czech Republic. *Herpetological Review* 43(1): 150–159.

Crawley M. J., 2007: *The R Book*. John Wiley & Sons, Chichester, England.

Dudík M., Phillips S. J. & Schapire R. E., 2007: Maximum entropy density estimation with generalized regularization and an application to species distribution modelling. *Journal of Machine Learning Research* 8: 1217–1260.

Elith J., Leathwick J. R. & Hastie T., 2008: A working guide to booster regression trees. *Journal of Animal Ecology* 77: 802–813.

Elith J. & Laethwick J. R., 2009: Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology and Systematics* 40: 677–97.

Elith J., Phillips S. J., Hastie T., Dudík M., Yung En Chee & Yates C. J., 2011: A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17: 43–57.

Fischer C. F., Garner T. W. J. & Walker S. F., 2009: Global Emergence of *Batrachochytrium dendrobatidis* and Amphibian Chytridiomycosis in Space, Time, and Host. *Annual Review Microbiology* 63: 291–310.

Franklin J., 1995: Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19: 474–499.

Franklin J. & Miller J. A., 2009: Mapping Species Distribution. *Spatial Inference and Prediction*. Cambridge University Press, England.

Gahegan M., 2003: Is inductive machine learning just another wild goose (or might it lay the golden egg?). *International Journal of Geographical Informatic Science* 17: 69–92.

Garner T. W. J., Walker S., Bosch J., Hyatt A. D., Cunningham A. A. & Fisher M. C., 2005: Chytrid fungus in Europe. *Emerging Infectious Diseases* 11(10): 1639–1641.

Gibson R. C. & Freeman M., 1997: Conservation at home: Recovery programme for the agile frog *Rana dalmatina* in Jersey. *Dodo-Journal of the Wildlife Preservation Trusts* 33: 91–104.

Hastie T., Tibshirani R. J. & Friedman J., 2009: The Elements of Statistical Learning: Data Mining, Inference and Prediction. 2nd ed. Springer – Verlag, New York.

Hernandez P. A., Graham C. H., Master L. L. & Albert D. L., 2006: The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29: 773–785.

Johnson M. L. & Speare R., 2003: Survival of *Batrachochytrium dendrobatidis* in water: quarantine and control implications. *Emerging Infectious Diseases* 9(8): 922–925.

Keisecker J. M., Blaustein A. R. & Belden L. K., 2001: Complex causes of amphibian population declines. *Nature* 410: 681–684.

Lepš J. 1996: Biostatistika. Biologická fakulta JČU, České Budějovice.

Longcore J. E., Pessier A. P. & Nichols D. K., 1999: *Batrachochytrium dendrobatidis* gen et sp. nov., a chytrid pathogen to amphibians, *Mycologia* 91: 219–227.

Measey, G. J., Roedder, D., Green, S. L., Kobayashi R., Lillo F, Lobos G., Rebelo R. & Thirion J. M., 2012: Ongoing invasions of the African clawed frog, *Xenopus laevis*: a global review. *Biological Invasions* 14: 2255–2270.

Mutschmann F., Berger L., Zwart P. & Gaedicke C., 2000: Chytridiomycosis in amphibians - first report in Europe. *Berl Munch Tierarztl Wochenschr* 113: 380–3.

Olden J. D., Lawler J. J. & Poff N. L., 2008: Machine learning methods without tears: A primer for ecologists. *Quarterly Review of Biology* 82: 171–193.

Pekár S. & Brabec M., 2009: Moderní analýza biologických dat. Zobecněné lineární modely v prostředí R. Scientia, Praha.

Puschendorf R., Carnaval A. C., VanDerWal J., Zumbado-Ulate H., Chaves G., Bolaños F. & Alford R. A., 2009: Distribution models for the amphibian chytrid *Batrachochytrium dendrobatidis* in Costa Rica: proposing climatic refuges as a conservation tool. *Diversity and Distributions* 15: 401–408.

Rödder D., Kielgast J., Bielby J., Schmidtlein S., Bosch J., Garner T. W. J., Veith M., Walker S., Fisher M. C. & Lötters S., 2009: Global Amphibian Extinction Risk Assessment for the Panzootic Chytrid Fungus. *Diversity* 1: 52–66.

Stanley S. P., 1949: The *Xenopus* pregnancy test. *Canadian Medical Association Journal* 60: 159–161.

Vié J. C., Hilton Taylor C. & STUART S. N., [eds] 2009: *Wildlife in a Changing World – An Analysis of the 2008 IUCN Red List of Threatened Species.* Gland, Switzerland, IUCN.

Venables W. M. & Ripley B. D., 1994: *Modern Applied Statistics with S-Plus.* Springer-Verlag, New York.

Vojar J., Zavadil V. & Musilová R., 2012: Biodiverzita obojživelníků. In: Machar I., Drobilová L. [eds]: *Ochrana přírody a krajiny v České republice: vybrané aktuální problémy a možnosti jejich řešení.* Univerzita Palackého, Olomouc pp: 653-659.

Voyles J., Young S., Berger L., Campbell C., Voyles W. F., Dinudom A., Cook D., Webb R., Alford R. A., Skerratt L. F. & Speare R., 2009: Pathogenesis of Chytridiomycosis, a Cause of Catastrophic Amphibian Declines. *Science*, 326: 582–585.

Weldon C., Preez L. H., Hyatt A. D., Muller R. & Speare R., 2004: Origin of Amphibian Chytrid Fungus. *Emerging Infectious Diseases* 10: 2100–2105.

Yee T. W., & Mitchell N. D., 1991: Generalized additive models in plant ecology. *Journal of Vegetation Science* 2: 587–602.

7.2. Internetové zdroje:

<http://www.bd-maps.net/>

<http://www1.cenia.cz/www/>

<http://geoportal.cuzk.cz/>

<http://www.cs.princeton.edu/~schapire/maxent/>

<http://www.worldclim.org/>

8. Přílohy

Příloha č. 1: Mapa rozšíření *Bd* v ČR

