

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE  
FAKULTA ŽIVOTNÍHO PROSTŘEDÍ



DIPLOMOVÁ PRÁCE

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE  
FAKULTA ŽIVOTNÍHO PROSTŘEDÍ  
KATEDRA APLIKOVANÉ GEOINFORMATIKY A  
ÚZEMNÍHO PLÁNOVÁNÍ

**Vliv rozdílnosti geodat reprezentujících vodní habitaty  
na modely druhové distribuce**

DIPLOMOVÁ PRÁCE

Vedoucí práce: Ing. Vítězslav Moudrý, Ph.D.

Diplomant: Bc. Marie Vosmíková

2017

# ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Fakulta životního prostředí

## ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Marie Vosmíková

Ochrana přírody

Název práce

**Vliv rozdílnosti geodat reprezentujících vodní habitaty na modely druhové distribuce**

Název anglicky

**Influence of water habitat geodata differences on species distribution models**

### Cíle práce

Cílem práce je posoudit vliv různých geodat reprezentujících vodní habitaty na kvalitu modelů druhové distribuce.

- Generování virtuálního druhu
- Vytvoření modelů druhové distribuce s různými geodaty
- Porovnání vytvořených modelů z hlediska predikované presence/absence

### Metodika

Modelování druhové distribuce (Species Distribution Modelling, SDM) patří mezi oblíbené nástroje v ekologii a biogeografii. Jejich cílem je obvykle najít environmentální proměnné, které mohou mít významný vliv na rozšíření druhu nebo rozšíření druhu predikovat. Nicméně data použitá pro kalibraci těchto modelů mohou mít zásadní vliv na jejich přesnost. Zatímco s využitím reálně sbíraných dat je obtížné tyto vlivy kvantifikovat, generováním tzv. virtuálního druhu lze snadno určit, jaký vliv mají vstupní data na výsledný model. Úkolem autorky je zpracovat literární rešerši na téma modelování druhové distribuce a vytvořit virtuální druh, na jehož příkladu otestuje vliv různých geodat reprezentujících vodní habitaty na kvalitu modelu.

**Doporučený rozsah práce**

40 – 50 stran textu

**Klíčová slova**

Formulace klíčových slov je úkolem autora.

---

**Doporučené zdroje informací**

- Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics*, 40, 677-697.
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological modelling*, 135(2), 147-186.
- LEROY, Boris, Christine N. MEYNARD, Céline BELLARD a Franck COURCHAMP. Virtualspecies, an R package to generate virtual species distributions. *Ecography*. 2015, , 001-009. DOI: 10.1111/ecog.01388. ISSN 09067590.
- MEYNARD, Christine N. a David M. KAPLAN. The effect of a gradual response to the environment on species distribution modeling performance. *Ecography*. 2012, 35(6), 499-509. DOI: 10.1111/j.1600-0587.2011.07157.x. ISSN 09067590
- MOUDRÝ, Vítězslav. Modelling species distributions with simulated virtual species. *Journal of Biogeography*. 2015,42(8). DOI: 10.1111/jbi.12552. ISBN 10.1111/jbi.12552.
- 

**Předběžný termín obhajoby**

2016/17 LS – FŽP

**Vedoucí práce**

Ing. Vítězslav Moudrý, Ph.D.

**Garantující pracoviště**

Katedra aplikované geoinformatiky a územního plánování

Elektronicky schváleno dne 17. 4. 2017

**doc. Ing. Petra Šimová, Ph.D.**

Vedoucí katedry

Elektronicky schváleno dne 18. 4. 2017

**prof. RNDr. Vladimír Bejček, CSc.**

Děkan

V Praze dne 18. 04. 2017

## **Prohlášení**

Prohlašuji, že jsem tuto diplomovou práci vytvořila samostatně pod vedením Ing. Vítězslava Moudrého, Ph.D. a že jsem uvedla veškeré literární prameny, ze kterých jsem čerpala.

Ve Slaném dne 10.4.2017

.....

## **Poděkování**

Děkuji tímto vedoucímu práce Ing. Vítězslavu Moudrému, Ph.D. za příkladnou pomoc a velkou trpělivost. Děkuji také celé své rodině za podporu během studia.

Ve Slaném dne 10.4.2017

.....

## **Abstrakt**

Modelování druhové distribuce je velmi častým nástrojem využívaným pro potřeby různých přírodovědných disciplín jako je například ochrana přírody či ekologie. Důležitým aspektem ovlivňujícím přesnost a vypovídací hodnotu těchto modelů je kvalita a relevance vstupních dat. V případech, kdy smyslem výzkumu není přímé vysvětlení výskytu konkrétního druhu v závislosti na podmínkách životního prostředí či mezidruhových vztazích, ale spíše pochopení fungování modelů či využití různých prediktorů, je vhodnou alternativou vytvoření tzv. virtuálního druhu. Výskyt virtuálního druhu je závislý pouze na zvolených faktorech a není ovlivněn dalšími podmínkami.

Tato diplomová práce zkoumá vliv několika různě podrobných datových sad zobrazujících vodní plochy v České republice a Evropě na přesnost modelů druhové distribuce. Výzkum byl proveden právě s využitím virtuálního druhu závislého na rozloze vodních ploch podle databáze CORINE Land Cover a teplotě v jarních měsících podle databáze WorldClim. Následně byla modelována distribuce virtuálního druhu s použitím rozlohy vodních ploch databází OpenStreetMap a Global Inland Water (pro Českou republiku byla ještě využita národní databáze DIBAVOD).

Výsledné modely byly porovnány podle ukazatele AUC (Area Under Curve). Na základě dosažených hodnot nelze potvrdit, že v rozlišení, v jakém bylo modelování provedeno, ovlivňuje přesnost použité datové sady přesnost modelu druhové distribuce. Problematika je ovšem značně rozsáhlá a zasloužila by další výzkum.

## **Klíčová slova**

virtuální druh, virtualspecies, GLM, modelování druhové distribuce, AUC, rozlišení

## **Abstract**

Species distribution modelling represents an important tool that is very often used for various natural sciences like nature protection, conservation biology or ecology. Quality and relevancy of input data are key elements affecting the accuracy and predicative value of these models. When the study of different aspects of modelling itself is the aim of the research, rather than the explanation of particular species occurrence in dependence on environmental conditions or interspecific relationships, then so-called virtual species can be a perfect alternative. Virtual species' occurrence is modelled based on selected predictors only and therefore is not affected by other circumstances.

This thesis investigates an effect of various datasets representing water bodies in the Czech Republic and Europe in different detail on the species distribution models. Virtual species which occurrence was dependent on the area of water bodies derived from CORINE Land Cover database and the mean spring temperature derived from WorldClim database was generated. Virtual species distribution models were created using the area of water bodies derived from OpenStreetMap and Global Inland Water datasets (national dataset DIBAVOD was also used for the Czech Republic).

Final models were compared as per the achieved value of AUC (Area Under Curve). Based on this indicator, the detail of dataset used for SDMs didn't seem to have a detectable effect on the accuracy of the model. The scope is quite complicated and would require further research.

## **Key words**

virtualspecies, GLM, species distribution modelling, AUC, scale



## Obsah

1.	ÚVOD .....	10
1.1	Cíle práce .....	11
2.	LITERÁRNÍ REŠERŠE.....	12
2.1	Modelování druhové distribuce .....	12
2.1.1	Vstupní data .....	12
2.1.2	Metody modelování.....	13
2.2	Rozlišení .....	14
2.2.1	Rozlišení území .....	14
2.2.2	Rozlišení použitých environmentálních proměnných .....	15
2.3	Využití virtuálních druhů při modelování druhové distribuce.....	15
3.	METODIKA.....	17
3.1	Charakteristika území .....	17
3.2	Vstupní data .....	18
3.2.1	Virtuální druh – data o výskytu .....	18
3.2.2	Environmentální proměnné.....	19
3.3	Analýza dat.....	23
3.3.1	Příprava dat.....	23
3.3.1.1.	Modelové území .....	23
3.3.1.2.	Vektorové databáze DIBAVOD, OSM, CLC .....	23
3.3.1.3.	Rastrové databáze GIW, WC .....	24
3.3.2	Výběr vhodné velikosti kvadrátů .....	27
3.3.3	Virtuální druh.....	28
3.3.4	Logistická regrese .....	32
3.3.5	Modelování .....	33
3.4	Zhodnocení modelů .....	34
4.	VÝSLEDKY.....	36
4.1	Srovnání modelů dle AIC .....	36
4.2	Srovnání modelů dle AUC .....	37
4.3	Srovnání modelů dle kalibračních grafů .....	38
5.	DISKUSE .....	41
6.	ZÁVĚR .....	44
7.	SEZNAM POUŽITÉ LITERATURY.....	45
8.	SEZNAM PŘÍLOH .....	49
9.	PŘÍLOHY .....	50

## SEZNAM POUŽITÝCH ZKRATEK

AUC	Area Under Curve
CLC	CORINE Land Cover
DIBAVOD	Digitální báze vodohospodářských dat
EEA	European Environment Agency
ENFA	Ecological Niche Factor Analysis
ETRS	European Terrestrial Reference System 1989
GCS	Geographic Coordinate System
GIW	Global Inland Water
GLM	General Linear Model
LAEA	Lambert Azimuthal Equal-Area
OSM	OpenStreetMap
PCA	Principal Components Analysis
ROC	Receiver Operating Characteristic Curve
S-JTSK	Systém Jednotné trigonometrické sítě katastrální
SDM	Species Distribution Models
UTM	Univerzální transverzální Mercatorův systém souřadnic
WC	WorldClim
WGS 84	World Geodetic System 1984

# 1. ÚVOD

Současná společnost svou činností významně ovlivňuje životní prostředí, a to především v negativním slova smyslu. Naštěstí tento trend nezůstává bez povšimnutí nejen odborné veřejnosti. Světové organizace, vlády jednotlivých zemí i menší organizace působící v oblasti životního prostředí se snaží tyto dopady usměrňovat a udržet životní prostředí v co nejlepší kondici pro budoucí generace. Bohužel se zatím daří nacházet spíše teoretická východiska, která jsou pak mnohem obtížněji aplikovatelná v praxi.

Jedním z častých témat je ochrana rostlinných i živočišných druhů, které jsou lidskou činností postiženy nejen přímo, ale také nepřímo v důsledku zániku vhodných habitatů či celých ekosystémů, v nichž se druhy vyskytují. Ekologie, ochrana přírody a obdobné disciplíny se snaží na základě dostupných informací o druzích a o podmínkách životního prostředí poskytnout podklady pro praktickou ochranu druhů a ekosystémů. Díky moderním technologiím lze data zpracovávat pomocí nejrůznějších specializovaných softwarů za využití celé řady metod. Velmi častou metodou je vytváření modelů druhové distribuce (species distribution models, SDM) (Elith et Leathwick, 2009). Principem těchto modelů je snaha o vysvětlení výskytu druhů pomocí relevantních a dostupných environmentálních proměnných (obvykle klimatických, geografických, geomorfologických, apod.) a to za pomoci různých statistických nástrojů (Hijmans et Elith, 2017).

První složkou, nutnou k vytvoření SDM, jsou data o výskytu druhů. Zde můžeme využít data, která byla pro konkrétní druhy nějakým způsobem nasbírána v terénu. Ta lze získat například z atlasů rozšíření jednotlivých druhů. Vznik takovýchto atlasů vyžaduje rozsáhlou práci velkých skupin odborníků či zainteresovaných laiků, kteří se mapování účastní. U nás takto vzniká například Atlas hnízdního rozšíření ptáků v České republice 2001 - 2003 (Šťastný et al., 2006) nebo Atlas rozšíření savců v České republice (Anděra, 2000 – 2007). Výsledky sledování jsou zaznamenány v přehledných mapách a umožňují nejen jednorázové využití, ale také porovnání a sledování změn výskytu druhů, popř. jejich početnosti apod. (například Jednotný program sčítání ptáků). Pro další zpracování je pak ideální, aby mapování probíhala opakovaně, vždy stejným způsobem a na stejném území (Reif et al., 2008, Robertson et al., 2010)

Nezabývá-li se studie přímo konkrétním druhem, ale spíše obecnějšími principy modelování druhové distribuce, nebo například zkoumá vhodnost různých dostupných dat, jež lze využít jako vysvětlujících proměnných pro modelování, nepracuje se většinou s daty o výskytu opravdového druhu. V takových případech se spíše přistupuje k variantě vytvoření tzv. virtuálních druhů (Bombi et D'Amen, 2012). Výstupem jsou presenčně-absenční data o výskytu virtuálního druhu, která jsou vymodelována na základě environmentálních proměnných, které autor sám zvolí za signifikantní a při modelování použije (Leroy, 2016).

Důležitým aspektem, který je potřeba vzít v úvahu jak při sběru druhových dat, tak při hledání vhodných environmentálních proměnných, je měřítko (Goodchild, 2011, Moudrý et Šimová, 2012). To bývá limitujícím faktorem především při získávání vstupních dat, protože vytvoření podrobné databáze je ekonomicky náročné (Guisan et Zimmermann, 2000). Environmentální proměnné často nejsou shromažďovány přímo měřením v terénu, ale mohou být vytvořeny pomocí moderních technologií – např. automatickou či vizuální interpretací satelitních snímků (Guisan et Zimmermann, 2000). Druhová data naopak vyžadují terénní monitoring, neboť jinak není možné získat relevantní výstupy.

Samotné modelování pak probíhá na základě dat v rastrové podobě, kdy je rastr rozdělen na jednotlivé buňky o stejné velikosti. V každé z těchto buněk je pak vypočtena jedna hodnota pro každou ze sledovaných proměnných. Právě vhodná velikost pixelů je další klíčovou otázkou, kterou je třeba pečlivě promyslet před zahájením studie (Robertson et al., 2010)

V této diplomové práci budou nejprve porovnány tři volně dostupné datové sady popisující výskyt vodních ploch a toků v Evropě (CORINE Land Cover, Global Inland Water a OpenStreetMap). Data budou pro porovnání a následné zpracování v dalších krocích převedena do rastrové podoby, přičemž bude využito tří různých velikostí pixelu, a to 50 km, 10 km a 5 km. Velikost pixelu, která bude vyhodnocena jako nejlepší, pak bude využita pro vytvoření virtuálního druhu, jehož výskyt bude záviset na rozloze vodních ploch dle databáze CORINE Land Cover a na průměrné teplotě v jarních měsících podle databáze WorldClim. Jarní měsíce byly vybrány proto, že pro řadu druhů představují období vhodné k reprodukci. Modelování druhové distribuce bude provedeno pomocí generalizovaného lineárního modelu (GLM) a jako vstupní data budou využita presenčně-absenční data vygenerovaného virtuálního druhu a data zbylých dvou databází, které nebyly využity pro vytvoření druhu. Celý postup bude ještě zvláště proveden pouze pro území České republiky, kde bude pixel o velikosti 50 km nahrazen pixelem o velikosti 1 km. Zde bude kromě databází CLC, GIW a OSM figurovat také národní databáze DIBAVOD. Zdrojem průměrné teploty v jarních měsících bude opět databáze WorldClim.

## 1.1 Cíle práce

Cíle práce lze tedy vymežit takto:

- vytvoření virtuálního druhu, jehož výskyt závisí na rozloze vodních ploch podle databáze CORINE Land Cover a na teplotě v jarních měsících podle databáze WorldClim,
- vymodelování distribuce virtuálního druhu s použitím teploty dle databáze WorldClim a rozlohy vodních ploch databází Global Inland Water a OpenStreetMap (pro Českou republiku navíc ještě DIBAVOD),
- stanovení velikosti pixelu, při níž přestane mít přesnost vysvětlujících proměnných vliv na modely druhové distribuce.

## 2. LITERÁRNÍ REŠERŠE

### 2.1 Modelování druhové distribuce

Modelování druhové distribuce (SDM) je hojně využívaným nástrojem současných studií zaměřených na otázky ochrany přírody, ekologie, atd. (např. Meynard et Kaplan, 2013; De Clercq et al., 2015). Jedná se o modely, které pracují s daty popisujícími výskyt druhů nebo jejich početnost a s environmentálními proměnnými. Smyslem SDM je nalezení vztahu mezi výskytem (resp. početností) druhu (tj. vysvětlovanou, závislou proměnnou) a podmínkami prostředí (vysvětlujícími, nezávislými proměnnými), které výskyt druhu přímo či nepřímo ovlivňují. V každé studii je třeba nejprve vymezit, jakých druhů se modelování týká, na jakých vysvětlujících proměnných bude provedeno a jaký je jeho cíl. Všechny tyto aspekty spolu musí ladit.

Studie bývají prováděny na druzích různých prostředí. Častými objekty jsou terestrické druhy rostlinné i živočišné (např. Fernández et al., 2012; Szymkowiak et al., 2014; De Clercq et al., 2015). V posledních letech se objevují i studie druhů vodního prostředí (např. Boscutti et al., 2015), v němž je poněkud komplikovanější získat vstupní data. Data o výskytu druhu bývají například chápána spíše jako pravděpodobnosti výskytu, protože vysoké procento druhů vodního prostředí je velmi mobilních (Elith et Leathwick, 2009). Dosud nejméně zkoumanou skupinou druhů jsou organismy žijící v půdě (Elith et Leathwick, 2009). Ideální model by měl klást stejný důraz na to, aby byl realistický, obecný a zároveň přesný. Udržet všechny tři parametry na téže úrovni však není zcela možné a dochází tak k jisté směně, kdy některý z parametrů musí být obětován na úkor zbylých dvou (Guisan et Zimmermann, 2000).

#### 2.1.1 Vstupní data

Kvalita vstupních dat je základním předpokladem pro přípravu SDM se silnou vypovídací hodnotou. Je třeba zohlednit elementární ekologické charakteristiky zkoumaného druhu (Elith et Leathwick, 2009). Guisan et Thuiller (2005) zmiňují, že pokud bude předmětem zájmu druh migrující, bylo by vhodné zohlednit i proměnné, které jeho migraci podmiňují. Naopak u druhů sedentárních by mohlo k sofistikovanějším výsledkům vést využití proměnných, které na druhy působí spíše v lokálním měřítku, ale v globálním měřítku by byly hůře sledovatelné (např. kompetice, predace,...). Využití těchto proměnných je ovšem značně limitováno jejich dostupností, neboť vyžadují jednak adekvátní data o dalším druhu a jednak předpokládají dostatečnou znalost mezidruhových vztahů.

Vstupní data lze získat dvěma základními způsoby:

- a) sběrem či měřením přímo v terénu,
- b) odvozením z map a satelitních snímků.

Terénní sběr je typický především při shromažďování dat o výskytu druhů. Sběr probíhá obvykle za pomoci skupiny dobrovolníků, kteří se pohybují ve vymezených, předem stanovených územích a dle vypracované jednotné metodiky tato území prozkoumávají. Zjištěné údaje pak předávají koordinátorům celé akce, kteří na jejich základě sestaví mapu výskytu druhu a to buď ve formě presenčně-absenčních dat, presenčních dat nebo jako údaje o početnosti (Robertson et al., 2010). Moudrý et al. (2017) navrhuje, že při tvorbě presenčně-absenčních dat by měly být brány v úvahu i další aspekty, jako například početnost či bližší informace o samotném průběhu mapování v jednotlivých územích. Mapy bývají vydávány ve formě atlasů.

Robertson et al. (2010) dále uvádí, že území, na kterém ke sběru dat dochází, by v ideálním případě mělo být rozděleno na co nejmenší jednotky. Pak by bylo možné s daty pracovat ve studiích o různých měřících, protože podrobná data lze jednoduše upravovat tak, aby vyhovovala i tam, kde je využito méně podrobného rozlišení. Opačný postup by byl ovšem značně nepřesný. Právě rozlišení dostupných vstupních dat tak bývá limitujícím faktorem při výběru měřítka pro celou studii.

Informace o environmentálních proměnných jsou obvykle získány počítačovým zpracováním map či satelitních snímků, často podpořených údaji z měřících stanic, rozprostřených po celém světě. V globálním měřítku je to jediný možný způsob, neboť prosté měření těchto charakteristik přímo v terénu by bylo problematické a nákladné (Guisan et Zimmermann, 2000). Terénní mapování je opět využíváno spíše při tvorbě databází na menších územích (např. DIBAVOD, která vzniká pouze pro území ČR). Několik způsobů možného odvození těchto dat z primárních zdrojů je uvedeno v kapitole 3.2.1 Environmentální proměnné.

### **2.1.2 Metody modelování**

Modelování provádíme pomocí metod regresní analýzy, které zkoumají závislost jedné náhodné veličiny na jiné veličině, popř. několika jiných veličinách. Vhodný druh analýzy pak volíme podle charakteru vstupních veličin, kdy především hledíme na to, zda se jedná o veličiny diskrétní či spojité, a na jejich rozdělení (Zvára 2008). Klíčovým kritériem pro výběr metody je bezpochyby také stanovení cíle studie (Guisan et Zimmermann, 2000). Mezi často využívané metody regresní analýzy patří především modely GAM (Generalized Additive Model) a GLM (General Linear Model) (Guisan et al., 2002).

GLM vychází z obecných lineárních modelů, ale na rozdíl od nich nevyžadují použití závisle proměnné kvantitativního charakteru s normálním rozdělením. Jedná se tedy vlastně o modely, které zobecňují lineární regresi, neboť pracují s celou řadou jiných rozdělení (binomické, Poissonovo, negativně binomické, Gamma, Gaussovo) (Šmilauer, 1998 – 2007). GLM pracují s tzv. link funkcí, která převádí hodnoty závisle proměnné na hodnoty srovnatelné s vysvětlujícími proměnnými, a předpovídají průměrnou hodnotu vysvětlované proměnné. V případě použití GLM pro diskrétní

závisle proměnnou s binomickým rozdělením je přiřazena kanonická link funkce logit, pro jiná rozdělení se pak využívají jiné funkce (Guisan et al., 2002).

## 2.2 Rozlišení

### 2.2.1 Rozlišení území

Velikost řešeného území zahrnuje dvě úrovně. Nejprve se zabýváme celkovou plochou, na které studie probíhá. Může to být například kontinent, stát nebo konkrétní lokalita (Whittaker et al., 2005). Vstupní data musí být pro potřeby statistického zpracování upravena do podoby rastru rozděleného na stejně velká území (kvadráty) (Hijmans et al., 2017). Právě tyto kvadráty představují druhou úroveň. Jsou vlastně územím, na kterém dochází k dílčí výzkumné práci (Whittaker et al., 2005).

Před zahájením práce je třeba se zamyslet nad tím, jaká velikost kvadrátu je pro danou studii vhodná a smysluplná. Volba správného rozlišení na druhé úrovni by měla určitým způsobem korelovat s velikostí první úrovně a měla by také zohledňovat výkonnost využitých výpočetních technik. Pokud provádíme globální studii na území rozděleném na pixely o velikosti stovek metrů čtverečních, může být následné počítačové zpracování dat velmi náročné. Naopak zabýváme-li se studií lokálního významu, může být podrobnější rozlišení přínosem (Robertson et al., 2010). V případě atlasových dat můžeme zmínit například publikace Atlas hnízdního rozšíření ptáků v České republice (Šťastný et al., 2006), který pracuje s kvadráty o rozměrech cca 12 x 11,2 km, nebo Ptáci Krkonoš – atlas hnízdního rozšíření 2012 – 2014 (Flousek et al., 2015), v němž mohlo být s ohledem na velikost zájmové lokality využito menších kvadrátů (cca 1,5 x 1,4 km).

Jak již bylo zmíněno v kapitole 2.1.1., při výzkumu vycházejícím z dat o výskytu druhu v kvadrátech o určité velikosti je tato velikost kvadrátů značně limitující. Atlasová data lze sice přizpůsobit podrobnějšímu měřítku, ale výsledek takového počínání nemusí být přesný a vytvoření SDM na základě takových dat může vést ke zkresleným a nereálným výsledkům (Araújo et al., 2005). Moudrý (2017) tento fakt demonstruje na příkladu dat o výskytu druhu nasbíraných v kvadrátech velikosti 10 x 10 km a následně rozdělených na kvadráty menší (1 x 1 km). Vzhledem k tomu, že není obvykle známo, ve kterém místě zkoumaného kvadrátu byl druh přesně spatřen, nelze v podrobnějším rozlišení tento kvadrát označit. Jinými slovy kterýkoli z menších kvadrátů pak může být označen za ten, ve kterém byl druh přítomen, ačkoli to nemusí odpovídat skutečnosti. Stejně úskalí má ovšem i převod opačným směrem, tedy z menších kvadrátů na větší, neboť i zde dochází ke ztrátě informace o přesnější lokalizaci druhu.

Řada autorů studuje rozdíly v SDM vytvořených na jednom území, na podkladě stejných proměnných, ale s využitím několika kvadrátů různých velikostí. Moudrý et al. (2012) vytvořili rešerši takovýchto studií, z níž vyplývá, že nejlepších výsledků je dosahováno, pokud je území rozděleno na kvadráty o velikosti 1 – 100 km<sup>2</sup>.

Naopak příliš velká velikost kvadrátů (v řádech tisíců km<sup>2</sup>) je značně nepraktická pro podobné studie, protože dochází k významnému zobecnění podmínek v oblasti. Pokud by bylo například území České republiky rozděleno na čtverce 50 x 50 km, vycházelo by takovýchto čtverců na celé území pouze 32 a environmentální proměnné v nich vypočtených by velmi těžko mohly vysvětlovat příčiny konkrétních jevů.

### **2.2.2 Rozlišení použitých environmentálních proměnných**

Ruku v ruce s využitou velikostí kvadrátů by mělo jít také rozlišení vrstev environmentálních proměnných, na kterých bude model vystavěn. Guisan et Zimmermann (2000) zmiňují, že některé proměnné (například klimatické či o krajinném pokryvu) jsou dnes již bez problému dostupné, ale v méně podrobném rozlišení. Tyto a další proměnné by však pro lokální studie bylo vhodné mít k dispozici v detailnějším rozlišení, což autoři uvádějí na příkladu výzkumů prováděných v horských oblastech, kde se podmínky často rapidně mění ve vertikálním směru. Zpracování dat v podrobnějším rozlišení je limitováno především kvůli vysokým nákladům na jejich získání, úpravu i uložení (Goodchild, 2011). Dle Moudrého et Šímové (2012) vede použití environmentálních proměnných s detailnějším rozlišením ke zpřesnění modelů, protože podrobnější data mohou zobrazovat takové jevy, které sice nejsou plošně rozsáhlé, ale pro výskyt některých druhů mohou být přesto klíčové. Mohou to být například malé či dočasné plochy vodní plochy, které mohou být vhodným habitatem pro některé druhy ptáků a obojživelníků.

Bedia et al. (2013), kteří studovali rozšíření buku lesního na severu Pyrenejského poloostrova, ve své práci došli k závěru, že pro tvorbu SDM v lokálním měřítku je dobré použít data z lokálního zdroje (v jeho případě klimatická data z University of Barcelona Atlas s rozlišením 200 m) než dat globálních databází (zde WorldClim s rozlišením 1 km). Sklenářová (2011) tuto myšlenku podporuje, neboť v její studii prováděné na území ČR s pomocí tří různě podrobných databází vodních ploch (ArcČR 500, CORINE Land Cover a DIBAVOD) byla za nejlepší vyhodnocena data databáze DIBAVOD, která má ze všech tří možností nejpodrobnější rozlišení.

### **2.3 Využití virtuálních druhů při modelování druhové distribuce**

Virtuální druhy jsou v posledních letech hojně využívány pro SDM. Tyto druhy představují zajímavou variantu pro ty studie, v nichž není cílem vysvětlit nebo předpovídat výskyt konkrétního druhu na určitém území, ale spíše zkoumat další možnosti, které SDM skýtají (Moudrý, 2015). Virtuální druh je vymodelován pouze na základě environmentálních proměnných, které jsou autorem zvoleny, a tak výsledek není ovlivněn žádnými dalšími faktory (například výskytem jiných druhů na stanovišti či stochastickými jevy) (Duan et al., 2015). Přesný postup vytvoření virtuálního druhu je zmíněn v Metodice, proto je zde jen obecně uvedeno, že dle definovaného vztahu modelovaného virtuálního druhu s podmínkami prostředí je nejprve vytvořena mapa



pravděpodobnosti výskytu druhu, která je v dalším kroku převedena na presenčně-absenční data o výskytu virtuálního druhu (Leroy, 2015). Tyto presence a absence mohou následně v SDM figurovat jako vysvětlovaná proměnná.

S ohledem na to, jak důležitým nástrojem různých disciplín se modelování druhové distribuce stalo, je vyvíjen neustálý tlak na jejich zlepšování a zpřesňování. (Leroy, 2016) Virtuální druhy, které jsou v podstatě obrazem skutečného rozšíření druhu, umožňují testování toho, jaké aspekty SDM by měly být zváženy a dále rozvedeny. Pokud by byly všechny sporné faktory přezkoumány a podrobeny dalším revizím, mohlo by být dosahováno lepších výsledků u SDM prováděných na reálných druzích (Miller, 2014).

Hirzel et al. (2001) porovnávali na virtuálním druhu dvě metody přístupu k modelování vhodnosti habitatu – Ecological Niche Factor Analysis (ENFA) a GLM. Zatímco ENFA byla provedena pouze na datech presenčních, GLM testoval data presenčně-absenční. Výsledkem jejich studie bylo zjištění, že obě metody nabízejí v podstatě srovnatelné výsledky, čímž autoři potvrzují užitečnost virtuálních druhů při srovnávacích studiích. Huang et al. (2016) tuto myšlenku podporují, ale upozorňují na to, že klíčovou roli hraje vytvoření takového virtuálního druhu, který je adekvátní a realistický.

Existuje řada různých softwarů, které lze pro modelování virtuálních druhů využít. Pro tuto práci byl využit balíček virtualspecies (Leroy et al., 2016). Další možností je například SDMvspecies (Duan et al, 2015), NicheLim (Huang et al., 2016) nebo NicheA (Huijie et al. 2016). Jako nejkompaktnější je momentálně vnímán právě software virtualspecies, který umožňuje modelování na základě vztahu druhu s prostředím. Důležitým aspektem je, že si lze vybrat z velkého množství možností, jak přesně druh bude na podmínky prostředí odpovídat. K tomu lze využít jak jakoukoli funkci, která je dostupná v R, tak jakoukoli funkci definovanou uživatelem. Další výhodou je také funkce, která dovoluje tzv. sampling výskytu druhu (Leroy et al., 2016).

### 3. METODIKA

#### 3.1 Charakteristika území

Pro práci byla využita plocha celé Evropské unie a také některých nečlenských států, pro které byly k dispozici údaje databáze CLC, která je limitujícím prvkem (ostatní zvolené databáze jsou dostupné pro celý svět). Seznam všech dotčených států je uveden v Příloze 1. Celková plocha řešeného území je cca 5 128 650 km<sup>2</sup> (výstup z mapového dokumentu v ArcMapu).

Vzhledem k velké rozloze nelze jednotně charakterizovat podmínky prostředí. Řešená část Evropy se rozkládá na severní polokouli a spadá do dvou podnebných pásů – mírného a subtropického. Evropa je protkána několika miliony kilometrů tekoucích vod, přičemž nejdelší řekou v řešeném území je Dunaj (2860 km), který protéká hned deseti státy. Dále se zde nachází více než 500 000 přírodních jezer (většinou ledovcového původu) s plochou přesahující 1 ha, přičemž zhruba tři čtvrtiny těchto jezer jsou situovány v Norsku, Švédsku a Finsku. V těch oblastech Evropy, které byly ledovci ovlivněny jen minimálně (například Španělsko, jižní Anglie, Česká republika,...), můžeme najít velmi malý počet přírodních jezer. Větší část vodních ploch na těchto územích představují plochy vytvořené člověkem (WISE, 2017).

Zvláště byl průzkum proveden na území České republiky, pro kterou je typické mírné klima, přechodné mezi typem kontinentálním a typem oceánickým. Rozloha ČR činí přibližně 78 900 km<sup>2</sup>. Nejdelší řekou republiky je Vltava, jejíž délka na území státu dosahuje 377 km. Jak bylo výše zmíněno, počet přírodních jezer v ČR není vysoký, jedná se především o karová jezera na Šumavě, z nichž největším je Černé jezero.

## 3.2 Vstupní data

### 3.2.1 Virtuální druh – data o výskytu

Leroy et al. (2015) vytvořili pro generování virtuálních druhů volně dostupný balíček `virtualspecies`, který byl navržen pro použití ve statistickém programu R. Základním předpokladem pro práci s `virtualspecies` jsou správně zpracovaná vstupní data. Veškerá data musí být v rastrovém formátu, tedy mapový podklad musí být rozdělen na menší pravidelné buňky, tzv. pixely. V rámci každého pixelu je pak možné vyjádřit jednu hodnotu zvolených environmentálních proměnných (např. teploty, nadmořské výšky, obvodu vodních ploch,...). Pro každou proměnnou je třeba vytvořit jeden rastr. Všechny rastry jsou pak vloženy do R ve formátu `RasterStack`.

Samotné generování virtuálního druhu pak může být provedeno dvěma základními způsoby (Leroy et al., 2015):

- 1) definováním funkce, podle které druh odpovídá na podmínky prostředí (funkce `generateSpFromFun`)

Úskalí tohoto přístupu spočívá v tom, že je třeba přesně definovat odpovědní funkce všech nezávislých proměnných (např. Bombi et D'Amen, 2012). Dílčí odpovědi pak mohou být přepočítány na škálu 0 (nevhodné prostředí) – 1 (vhodné prostředí), následně skombinovány zvoleným způsobem (např. pomocí váženého průměru). Výsledkem je hodnota vhodnosti prostředí pro daný druh, která může být opět přepočtena na hodnoty 0 – 1.

- 2) prostřednictvím Principal Components Analysis (PCA) (funkce `generateSpFromPCA`)

PCA je neparametrická metoda, která se využívá v případě, kdy je třeba zpracovat větší množství proměnných, u nichž je nemožné předvídat, jaká bude odpovědní funkce druhu na každou z nich. Analýza umožňuje stanovit, které proměnné nejlépe vysvětlují variabilitu v datech. Těmto proměnným lze pak při dalším postupu přidělit vyšší váhu. Výsledkem pak bude opět hodnota vhodnosti prostředí, kterou lze (stejně jako v případě 1)) přepočíst na škálu 0 – 1.

Ať už byla vhodnost prostředí stanovena kterýmkoliv z uvedených způsobů, musí být dále převedena na presenčně-absenční data. Výsledkem převodu je binární mapa presencí a absencí. Do roku 2014 byl běžný postup takový, že byla stanovena přesná hranice pro reklasifikaci (tzv. threshold). Hodnoty, které byly pod touto hranicí, byly vyhodnoceny jako absence, hodnoty nad hranicí jako presence (Bombi et D'Amen, 2012). Tento přístup ale není zcela správný, neboť nereflktuje některé důležité ekologické principy, jež jsou platné na lokální úrovni (stochastické jevy, kompetice, predace,...), a vede k příliš optimistickým výsledkům (Meynard et Kaplan, 2013). Ve skutečnosti se velmi často stává, že druhy chybí na místech, která by byla vyhodnocena jako naprosto ideální, a naopak se vyskytují na místech, kde by teoreticky neměly být přítomny vůbec. Alternativu představuje přístup pravděpodobnostní, u nějž

pravděpodobnost výskytu druhu závisí na hodnotě vhodnosti prostředí. Závislost je popsána např. logistickou křivkou, kde na ose x je vhodnost prostředí a na ose y pravděpodobnost výskytu. Pro každý pixel je stanovena pravděpodobnost výskytu. Je-li její hodnota např. 0,4, znamená to, že ve 4 z 10 případů bude pixel vyhodnocen jako vhodný pro výskyt druhu (tedy bude mít hodnotu 1). Takže pokud budeme výpočet opakovat, bude se výsledná mapa lišit, ale zároveň bude vždy správná. (Leroy et al., 2015).

Výstupem výše zmíněných postupů je mapa, která zachycuje všechny oblasti zájmového území, které jsou vhodné pro výskyt druhu dle definovaných podmínek. Druhy se ale obvykle vyskytují jen na části potenciálně vhodného území. Tento předpoklad lze také naprogramovat pomocí virtualspecies například tak, že výskyt omezíme pouze na určitou plochu (například kontinent, stát,...). Posledním krokem je sámplování výskytu (funkce sampleOccurrences), tedy výběr určitého počtu pixelů z celého rastru, v němž budou stanoveny presence-absence nebo pouze presence (Leroy et al., 2015).

### 3.2.2 Environmentální proměnné

#### DIBAVOD

Digitální BÁze VOdohospodářských Dat (DIBAVOD), která obsahuje nejrůznější vodohospodářská data (například vodní toky, povodí I. – IV. řádu, chráněná území, záplavová území, objekty na toku,...), vznikla jako tematická nadstavba databáze ZABAGED®. Její správou a vývojem se zabývá Oddělení geografických informačních systémů a kartografie VÚV T.G.M., v.v.i., který jí rovněž ve spolupráci s ČHMÚ aktualizuje. Vybrané vrstvy jsou volně přístupné ke stažení ve formátu shapefile a jsou k dispozici v měřítku 1:10 000, resp. 1:50 000. DIBAVOD je rozdělen do 10 skupin dle účelu, každá skupina je dále dělena na objekty, kterých je v databázi zahrnuto celkem 75.

Použité vrstvy:

skupina A – základní jevy povrchových a podzemních vod - objekt A05 – vodní nádrže (v měřítku 1:10 000). Vodní nádrž je vodní útvar vzniklý akumulací vody v uměle vytvořeném prostoru nebo přírodní prohlubni na zemském povrchu, ve kterém se zdržuje nebo zpomaluje odtok vody z povodí. Náleží sem jak nádrže umělé (přehradní nádrže, rybníky) tak přirozené (jezera) reference. (DIBAVOD, 2017).

## **OpenStreetMap**

Projekt OpenStreetMap běží od roku 2004 a jeho hlavním cílem je vytvoření přesné topografické mapy pro celý svět, která bude k dispozici online zdarma a pro každého. Tato databáze je vytvářena tzv. komunitou OSM, jejímž členem může být prakticky kdokoli. Příspěvatelé sbírají geografická data buď přímo v terénu (např. s pomocí GPS) nebo vycházejí z leteckých snímků či z volně dostupných klasických map (popřípadě z licenčních map, jejichž autor však odsouhlasil použití pro OSM) (OPENSTREETMAP, 2016).

Přesnost výsledných map je většinou vyšší, než u klasických map, protože příspěvatelé využívají znalosti svého okolí a mohou také uvádět na pravou míru neaktuální či zkreslené podklady poskytnuté jinými členy komunity. Data jsou velmi přesná nejen fakticky, ale vyznačují se také vysokou prostorovou přesností a vysokým rozlišením. Další výhodou je fakt, že aktualizace mapy probíhá prakticky neustále. Data lze stáhnout například přes webové stránky [download.geofabrik.de](http://download.geofabrik.de) ve formátu shapefile a lze z nich vybrat pouze potřebné vrstvy.

Z portálu geofabrik byly staženy vrstvy jednotlivých evropských států a z nich pak byly využity pouze vrstvy water (obsahující plošné prvky jako nádrže, jezera, rybníky, ledovce,...).

## **Global Inland Water**

Feng et al. (2016) ve spolupráci s University of Maryland vytvořili databázi Global Inland Water (GIW) v roce 2015 na základě satelitních snímků programu Landsat zachycených v 30m rozlišení. Hlavním motivem pro tvorbu GIW bylo poskytnutí dat s vysokým rozlišením, která by byla využitelná pro vědecké účely.

Satelitní snímky byly počítačově zpracovány a výsledky byly poté porovnány s údaji jiných databází mapujících vnitrozemské vodní plochy ve stejném období (tj. v roce 2000). Mapa, která je volně dostupná na stránkách Global Land Cover Facility, zachycuje cca 3 650 000 km<sup>2</sup> vnitrozemských vodních ploch, z čehož zhruba 10 % se nachází v Evropě.

Nepřesnosti databáze jsou způsobeny tím, že vychází ze satelitních snímků, tedy zobrazuje situaci, která může být zkreslena sezónností vodního pokryvu, sněhu a ledu a zbytkovými mraky. Tyto chyby by bylo možné minimalizovat v případě, že by bylo zpracováno velké množství snímků Landsat pro každou oblast.

Výsledná mapa v rastrovém formátu pak obsahuje data s kódovanými hodnotami (0 – žádná data, 1 – pevnina, 2 – voda, 4 – sníh/led, 200 – oblačný stín, 201 – oblak).

## CORINE Land Cover

Program CORINE byl přijat v roce 1985 Evropskou komisí za účelem zajištění informací o stavu životního prostředí a přírodních zdrojů v rámci EU. Znalost současného stavu, minulého stavu a sledování změn je významnou podmínkou pro to, aby mohly orgány na všech možných úrovních zodpovědně rozhodovat o budoucím vývoji prostředí. Program sleduje především krajinný pokryv (CORINE Land Cover), biotopy (Biotopes) a ovzduší (Air).

Mapy pro databázi CORINE Land Cover (CLC) jsou vytvářeny jednotlivými členskými státy. Pověřené agentury členských států připravují data pro svoje území a následně je předávají Evropské agentuře pro životní prostředí (EEA), která je koordinátorem celé databáze. Za ČR zpracovává data CENIA (Česká informační agentura životního prostředí). Klíčovou podmínkou při zpracování dat je použití stejné metodiky, čímž je umožněno vytvoření jednotné databáze a následné sledování změn v jednotlivých oblastech (CENIA, 2016).

Data jsou odvozena ze satelitních snímků programů Landsat nebo SPOT. Ačkoliv snímky získané z různých programů mají různá měřítka, na výslednou mapu nemá výběr zásadní vliv, protože výstupy jsou vždy v měřítku 1:100 000. Při výběru zdroje snímků tak hrají roli především dostupné finance a dostupnost kvalitních snímků pro danou oblast. Snímky jsou pak počítačově zpracovány a upravovány, až ve výsledku vzniknou mapy s výše uvedeným měřítkem. Mapy lze stáhnout zdarma z webových stránek [www.eea.europa.eu](http://www.eea.europa.eu) například ve formátu shapefile, jenž umožňuje okamžité zpracování nástroji GIS. Nejmenší jednotka, která může být na mapě zaznamenána, musí mít plochu alespoň 25 ha a minimální šířku 100 m. Za změnovou plochu pak může být označena souvislá plocha o velikosti alespoň 5 ha.

Krajinný pokryv je v CLC rozdělen do 5 základních kategorií, které se dále dělí na 15 subkategorií a 44 tříd. V současné době jsou k dispozici vrstvy z let 1990, 2000, 2006 a 2012 a příslušné změnové databáze. V diplomové práci jsou využity vrstvy z roku 2006, konkrétně kategorie 5 Vodní plochy, subkategorie 5.1 Pevninské vody, třída 5.1.1. Vodní toky (Přírodní i uměle vytvořené vodní toky sloužící jako odvodňovací kanály. Minimální šířka nutná pro zahrnutí do této kategorie je 100 m.) a třída 5.1.2. Vodní plochy (Přírodní i uměle vytvořená zavodněná území. Tato třída zahrnuje i vodní plochy nádrží, které byly na originálních analyzovaných podkladech výjimečně vypuštěné. Ostrovy by měly být z vrstvy vypuštěny pouze v případě, že je jejich rozloha alespoň 25 ha.) (EEA, 2016).

## WorldClim

Globální databáze pevninského klimatu WorldClim (WC) vznikla pro potřeby studií různých oborů (především v oblasti ochrany přírody, klimatologie,...), v nichž je obvykle nutné pracovat s daty v co nejvyšším rozlišení (De Clercq et al., 2015; Toranza et al., 2016,...). Jedná se o jednu z mála databází, která je zdarma dostupná pro celý svět (s výjimkou Antarktidy) v rozlišení 30 arc s (přibližně 1 km).

Hijmans et al. (2005) vytvořili WC na základě skutečných dat (z let 1960 - 1990) získaných z pozemských klimatických stanic rozmístěných po celém světě, která shromáždili z globálních, regionálních, národních i místních zdrojů. Data byla podrobena kvalitativní kontrole, aby následná interpolace nebyla provedena na základě nepravdivých či zkreslených podkladů. Další kvalitativní kontrola pak proběhla i po samotné interpolaci. Výsledkem interpolace jsou mapy v rastrovém formátu (tj. složené ze stejně velkých čtverců), které mohou být jednoduše využity k práci v nástrojích GIS. Vrstvy jsou koordinovány v referenčním systému WGS84 a je třeba si pamatovat, že teploty v nich obsažené jsou uváděny jako desetinasobek skutečné hodnoty. K dispozici jsou mapy minimálních, průměrných a maximálních teplot a srážek a dále mapy devatenácti odvozených bioklimatických proměnných (např. průměrná roční teplota, maximální teplota v nejteplejším měsíci, srážky v nejlhčím měsíci,...).

Sami autoři zmiňují jisté nepřesnosti, které databáze má a které vycházejí především z neúplných nebo nedostatečných podkladů nebo mohou být následkem nesprávně použité interpolační metody. Tyto nepřesnosti se mohou odrážet různou měrou na jednotlivých vrstvách, tedy v jedné oblasti lze pracovat s poměrně přesnými daty o průměrné teplotě, ale průměrné srážky už mohou být značně odlišné. U bioklimatických proměnných, které jsou odvozeny z primárních dat o teplotách a srážkách, pak mohou být tyto chyby ještě umocněny. Zajímavé jsou pak studie, které pracují s údaji databáze WC, ale využívají i data dostupná z lokálních či místních databází, které obvykle při výzkumech na menších územích přinášejí přesnější výsledky (BEDIA et al., 2013).

V tuto chvíli probíhají dokončovací práce na 2. verzi databáze, která bude zahrnovat roky 1970-2000. Verze je zatím k dispozici pouze pro účely testování, proto tato diplomová práce obsahuje rastry předchozí verze, tedy verze 1.4.

Pro naše účely byly z webových stránek [www.worldclim.org](http://www.worldclim.org) staženy rastrové vrstvy průměrných měsíčních teplot v jarních měsících (březen, duben, květen) v nejvyšším možném rozlišení (30 arc s).

### 3.3 Analýza dat

#### 3.3.1 Příprava dat

Všechny vstupní vrstvy byly během úprav transformovány do stejného projekčního a geografického koordinačního systému. Detailnější informace o úskalích převodů mezi jednotlivými systémy přinesl například Budic et al. (2016). U vrstev, které nevyžadovaly speciální úpravy, byl převod proveden ihned. U ostatních (například GIW) až po spojení do jednotné vrstvy. Zvolen byl European Terrestrial Reference System 1989 (ETRS\_89\_LAEA). Tabulka 1 obsahuje přehled provedených transformací.

<b>Vstupní vrstvy</b>	<b>Původní systém</b>	<b>Systém po transformaci</b>
hranice Evropy, ČR	GCS_WGS_1984	ETRS89_LAEA
WorldClim	GCS_WGS_1984	ETRS89_LAEA
CORINE Land Cover	ETRS89_LAEA	ETRS89_LAEA
DIBAVOD	S_JTSK_KROVAK	ETRS89_LAEA
OpenStreetMap	GCS_WGS_1984	ETRS89_LAEA
Global Inland Water	UTM	ETRS89_LAEA

**Tabulka 1** Převody koordinačních systémů

##### 3.3.1.1. Modelové území

Z webových stránek byly staženy mapy evropských států, nacházejících se ve zkoumaném území (seznam států je uveden v Příloze č. 1). Tyto státy byly funkcí Merge sloučeny do vrstvy *europe.shp*. Na vrstvě *europe.shp* byly funkcí Create Fishnet postupně vystavěny kvadráty (gridy) o velikostech stran 50 km, 10 km a 5 km. Funkce vytvoří obdélník se čtvercovou sítí překrývající celé území, takže je nutné sít oříznout, aby zachycovala opět pouze zájmové území (funkce Clip). Takto byly získány vrstvy *euro\_50*, *euro\_10* a *euro\_5*. Stejným způsobem byly kvadráty připraveny také pro Českou republiku. Na území České republiky nebyla využita vrstva s velikostí gridů 50 km x 50 km, ale byla nahrazena gridy menší velikosti – konkrétně 1 km. Vznikly tak vrstvy *cze\_10*, *cze\_5* a *cze\_1*.

##### 3.3.1.2. Vektorové databáze DIBAVOD, OSM, CLC

Vrstvy databáze OSM byly staženy postupně ve formátech shapefile pro jednotlivé státy zájmového území. Ty byly tedy nejprve funkcí Merge sloučeny tak, aby vznikla jedna vrstva pro veškeré vodní nádrže Evropě dané oblasti nazvaná *osm\_w*. Vrstvy zobrazující vodní nádrže a toky dle databáze CLC byly již staženy pro celou Evropu, takže bylo potřeba je jen oříznout funkcí Clip, aby zachycovaly pouze dotčené území. Vznikly tak vrstvy *clc\_w* a *clc\_ww*. Stejným způsobem byly všechny vrstvy upraveny



také pro území ČR, pro nějž byla navíc obdobně zpracována i databáze DIBAVOD (vrstva *dib\_w*).

Veškeré vrstvy vodních ploch a toků CLC, tedy *osm\_w*, *clc\_w*, *clc\_ww* (pro ČR i *dib\_w*) pak byly znovu oříznuty funkcí Intersect. Výsledkem tohoto kroku je rozdělení každé vrstvy vodních ploch pomocí připravené sítě kvadrátů, takže atributové tabulky těchto vrstev pak uvádějí rozlohy vodních ploch v každém kvadrátu. Tyto rozlohy pak byly ještě přepočteny pomocí Calculate Geometry na km<sup>2</sup>. Protože pro modelování bylo třeba získat pro každý kvadrát jen jednu souhrnnou rozlohu vodních ploch, byla nakonec pro každou vrstvu vodních ploch (resp. toků v případě CLC) pomocí Summary statistics vytvořena tabulka, v níž jsou spočteny celkové rozlohy vodních ploch v každém kvadrátu. Vzniklo tak sedm tabulek s názvy *cz\_clc\_w\_sum*, *cz\_clc\_ww\_sum*, *eu\_clc\_w\_sum*, *eu\_clc\_ww\_sum*, *cz\_dib\_sum*, *cz\_osm\_sum* a *eu\_osm\_sum*.

### 3.3.1.3. Rastrové databáze GIW, WC

Data GIW i WC byla stažena v podobě menších rastrových čtverců a musela být nejprve sloučena do jedné vrstvy. Rastry GIW byly pospojovány podle UTM pásem. Jednotlivá pásma byla reklasifikována na hodnoty 0 a 1 (přičemž hodnota 1 znamená přítomnost vody), a pak spojena pomocí Mosaic To New Raster. V případě překryvu se jako klíčová hodnota brala hodnota maximální. Pokud tedy byla v jedné vrstvě voda přítomna a v překryvné vrstvě ne, je jako výsledek brána přítomnost vody. Vše pak bylo sjednoceno do finální vrstvy *giw\_w*, z níž byly pomocí Zonal Statistics as Table spočítány sumy rastrových buněk v jednotlivých čtvercích kvadrátových vrstev. Každá z buněk má rozlohu 0,009 km<sup>2</sup>, takže touto hodnotou byly sumy vynásobeny. U WC proběhlo sloučení stažených čtverců pomocí funkce Mosaic To New Raster, čímž vznikly tři nové vrstvy (*wc\_3*, *wc\_4* a *wc\_5*). Funkcí Zonal Statistics as Table byla v těchto vrstvách spočtena průměrná teplota v každém čtverci.

Veškeré hodnoty spočtené výše popsány postupy byly nakonec připojeny pomocí Join k příslušným vrstvám zobrazujícím kvadráty (tj. vrstvám *euro\_50*, *euro\_10*, *euro\_5*, *cz\_10*, *cz\_5* a *cz\_1*). V atributové tabulce těchto vrstev pak byly dopočítány další potřebné údaje, jako průměrná teplota v jarních měsících v každém čtverci a pro databázi CLC také celková rozloha vodních objektů, tedy součet rozlohy vodních ploch a rozlohy vodních toků. Dále byly spočteny rozdíly mezi jednotlivými databázemi, které pak byly vizualizovány do map, z nichž je jasně patrné, jak se vodní objekty zachycené různými databázemi vzájemně liší (obrázky 1 – 5). Tmavší barva znamená větší rozlohu vodních ploch v první jmenované databázi, světlejší barva pak větší rozlohu vodních ploch v druhé jmenované databázi. Pro ilustraci byly rozdíly vizualizovány ještě jednou v Přílohách č. 2 a 3, kde jsou znázorněny vodní plochy jednotlivých databází v jednom kvadrátu.

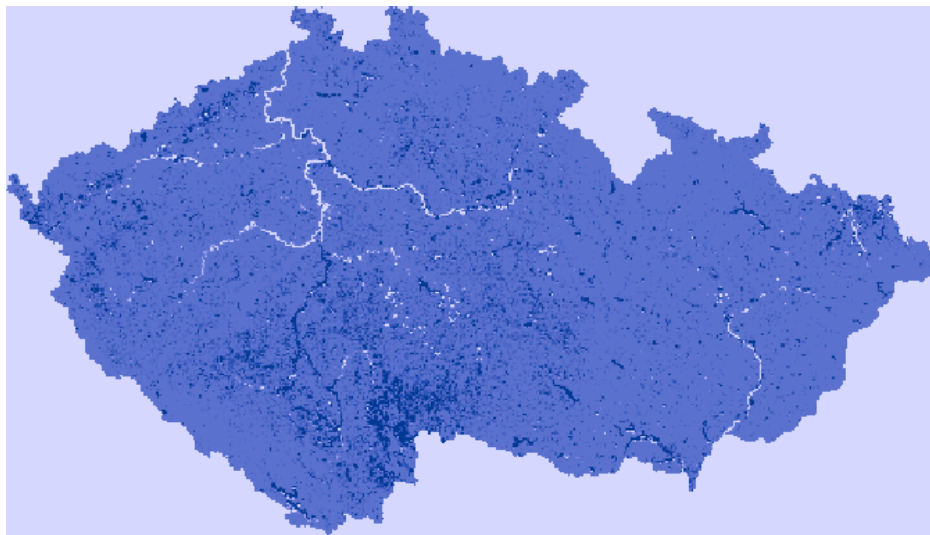
Souhrnné informace o vodních plochách dle jednotlivých databází, které zachycují základní rozdíly v podrobnosti databází, jsou uvedeny v tabulkách č. 2 a 3.

<b>EVROPA</b>	<b>CLC</b>	<b>OSM</b>	<b>GIW</b>
Celková rozloha vodních ploch (km <sup>2</sup> )	124 649	157 362	139 482

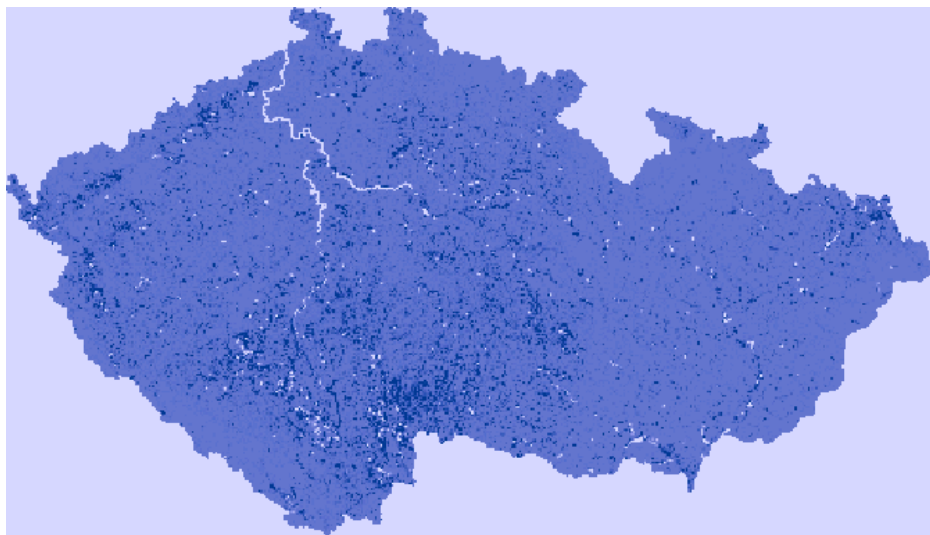
**Tabulka 2 Rozlohy vodních ploch jednotlivých databází (Evropa)**

<b>ČESKÁ REPUBLIKA</b>	<b>CLC</b>	<b>OSM</b>	<b>GIW</b>	<b>DIBAVOD</b>
Celková rozloha vodních ploch (km <sup>2</sup> )	565	1 202	670	835

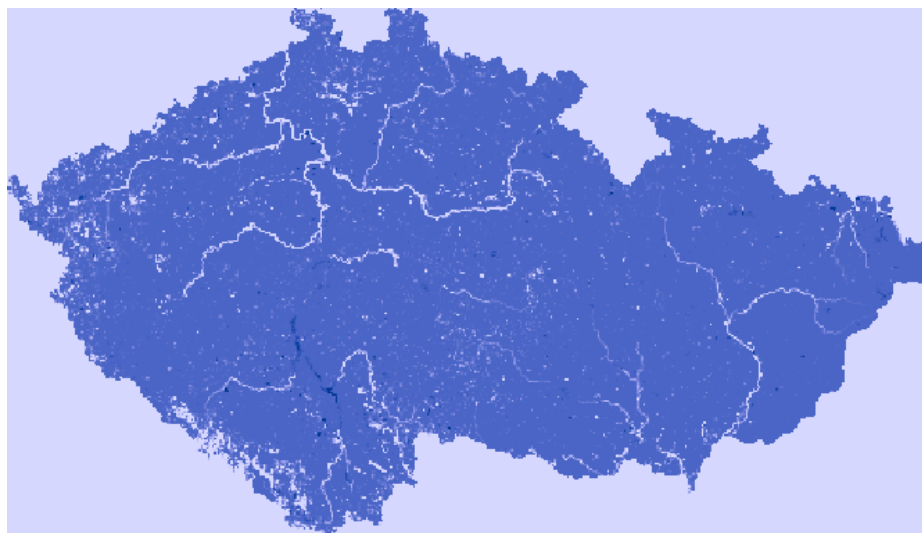
**Tabulka 3 Rozlohy vodních ploch jednotlivých databází (ČR)**



**Obrázek 1 Rozdílová mapa vodních ploch databází DIBAVOD a GIW**



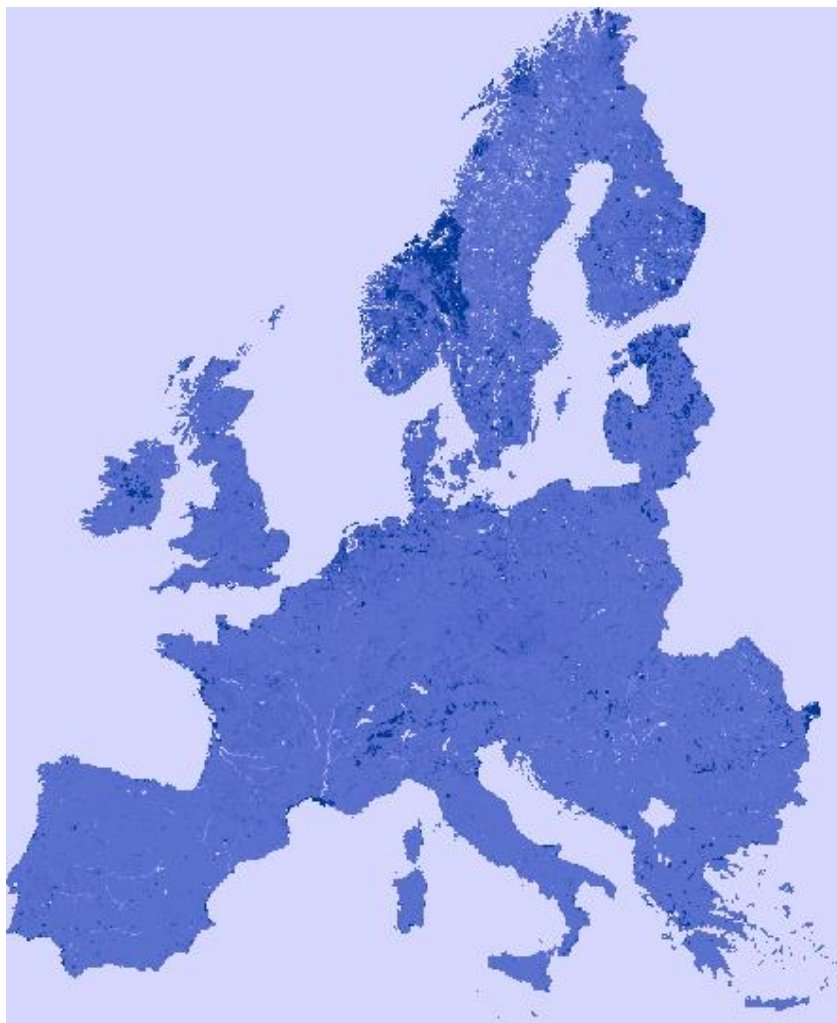
**Obrázek 2 Rozdílová mapa vodních ploch databází DIBAVOD a CLC**



**Obrázek 3 Rozdílová mapa vodních ploch databází DIBAVOD a OSM**



**Obrázek 4 Rozdílová mapa databází OSM a GIW**



Obrázek 5 Rozdílová mapa databází OSM a CLC

### 3.3.2 Výběr vhodné velikosti kvadrátů

Před samotnou tvorbou virtuálního druhu bylo třeba zvolit velikosti kvadrátů, s nimiž bude dále počítáno. Tento krok byl proveden na základě výpočtů korelací mezi jednotlivými databázemi vodních ploch.

S pomocí Spearmanova koeficientu pořadové korelace ( $r_s$ ) byly prozkoumány lineární závislosti mezi jednotlivými dvojicemi databází, jejich síla a směr. Hodnota koeficientu  $r_s$  se pohybuje mezi -1 a +1. V případě, že mezi prediktory není žádný vztah, dosahuje koeficient hodnoty 0. Znaménko před číslem pak podává informaci o tom, zda se jedná o pozitivní či negativní korelaci (Prion et Haerling, 2014). Čím vyšší je absolutní hodnota výsledného čísla, tím dokazuje silnější vztah (Tabulka 4) (Prion et Haerling, 2014):

hodnota $r_s$	hodnocení
0,81 – 1,00	velmi silná
0,61 – 0,80	silná
0,41 – 0,60	mírná
0,21 – 0,40	slabá
0,00 – 0,20	bezvýznamná

**Tabulka 4** Klasifikační tabulka Spearmanova korelačního koeficientu

### 3.3.3 Virtuální druh

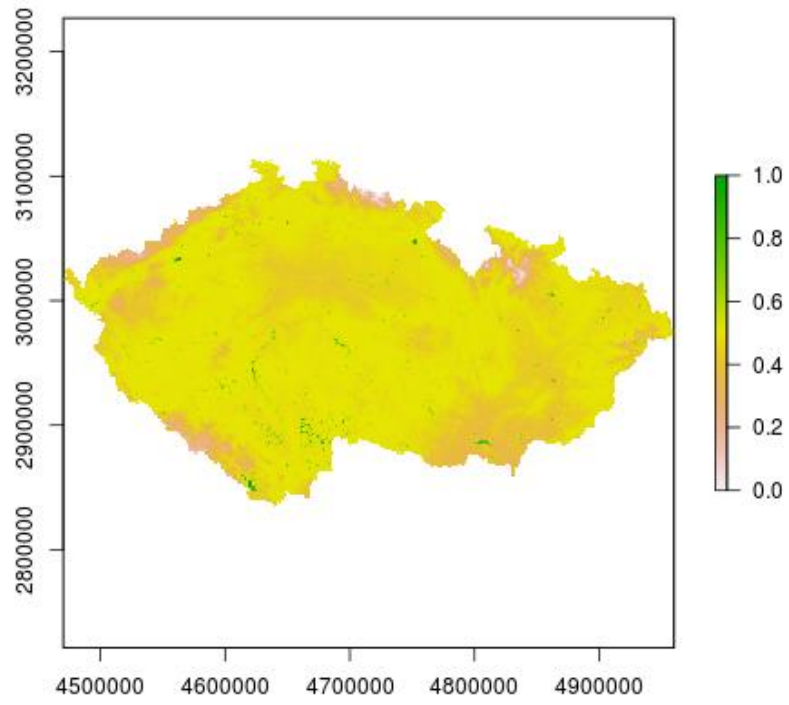
Aby mohla být environmentální data zpracována ve statistickém programu R za pomoci balíku *virtualspecies*, bylo třeba nejprve vyexportovat vrstvy do rastrového formátu, konkrétně do formátu *.tif*. To bylo provedeno v ArcMapu funkcí *Polygon to Raster*. Bylo nutné převést jen ty vrstvy, které přímo vstupují do samotné přípravy virtuálního druhu, tedy pouze hodnoty databází *WorldClim* a *CORINE Land Cover*. Vznikly tak rastrové objekty *cz\_wc.tif*, *cz\_clc.tif*, *eu\_wc.tif* a *eu\_clc.tif*. Ostatní databáze byly použity až pro modelování druhové distribuce, kam vstupují jako sloupce v tabulce v *.csv* formátu.

Po nainstalování všech potřebných balíčků (*virtualspecies*, *raster*, *rgdal*) a nahrání příslušných knihoven (*virtualspecies*, *raster*, *sp*, *rgdal*) byly do R vloženy připravené rastry *cz\_wc.tif* a *cz\_clc1.tif*. Pak byly definovány funkce, podle nichž druh odpovídá na podmínky prostředí (funkce *formatFunctions*). U průměrné teploty bylo počítáno s tím, že na ni druh bude odpovídat normálně (tedy dle normálního rozdělení), a bylo proto nutné stanovit její střední hodnotu ( $\mu$ ) a směrodatnou odchylku ( $\sigma$ ). Byly prozkoumány průměrné teploty jak v rámci Evropy, tak v rámci České republiky. Obě se pohybují kolem 7,2 °C, takže tato hodnota byla zvolena jako  $\mu$ . Parametr  $\sigma$  byl stanoven na hodnotu 3. Dle těchto dat by se měl druh vyskytovat s největší pravděpodobností na místech s teplotou 7,2 °C, naopak na místech příliš chladných či příliš teplých by se vyskytovat neměl. U rozlohy vodních ploch bylo předpokládáno, že čím větší rozlohu vodní plochy mají, tím větší je pravděpodobnost výskytu druhu. Byla zvolena logistická funkce a bylo nutné opět stanovit určité parametry. Pro logistickou funkci jsou to parametry alfa, který udává sklon křivky, a beta, který udává hodnotu inflexního bodu.

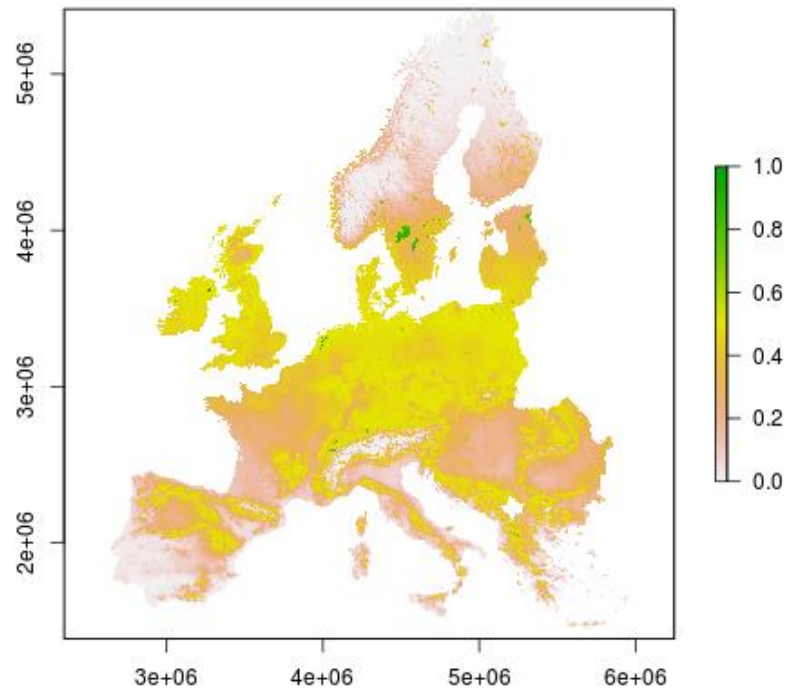
```
my.par.cz<-formatFunctions("cz_wc" = c(fun = 'dnorm', mean = 7.2,
sd = 3), "cz_clc1" = c(fun = 'logisticFun', beta = 15, alpha=-4))
```

V dalším kroku byl na základě definovaných funkcí vygenerován virtuální druh (funkce *generateSpFromFun*). Výsledkem je rastr zobrazující vhodnost prostředí pro virtuální druh. Tento rastr byl pomocí funkce *convertToPA* převeden na rastr presencí/absencí v zájmovém území.

```
my.sp.cz<-generateSpFromFun(raster.stack = cz.stack[[c("cz_wc",
"cz_clc1")]],parameters = my.par.cz, formula="cz_wc +
cz_clc1",plot = TRUE)
```



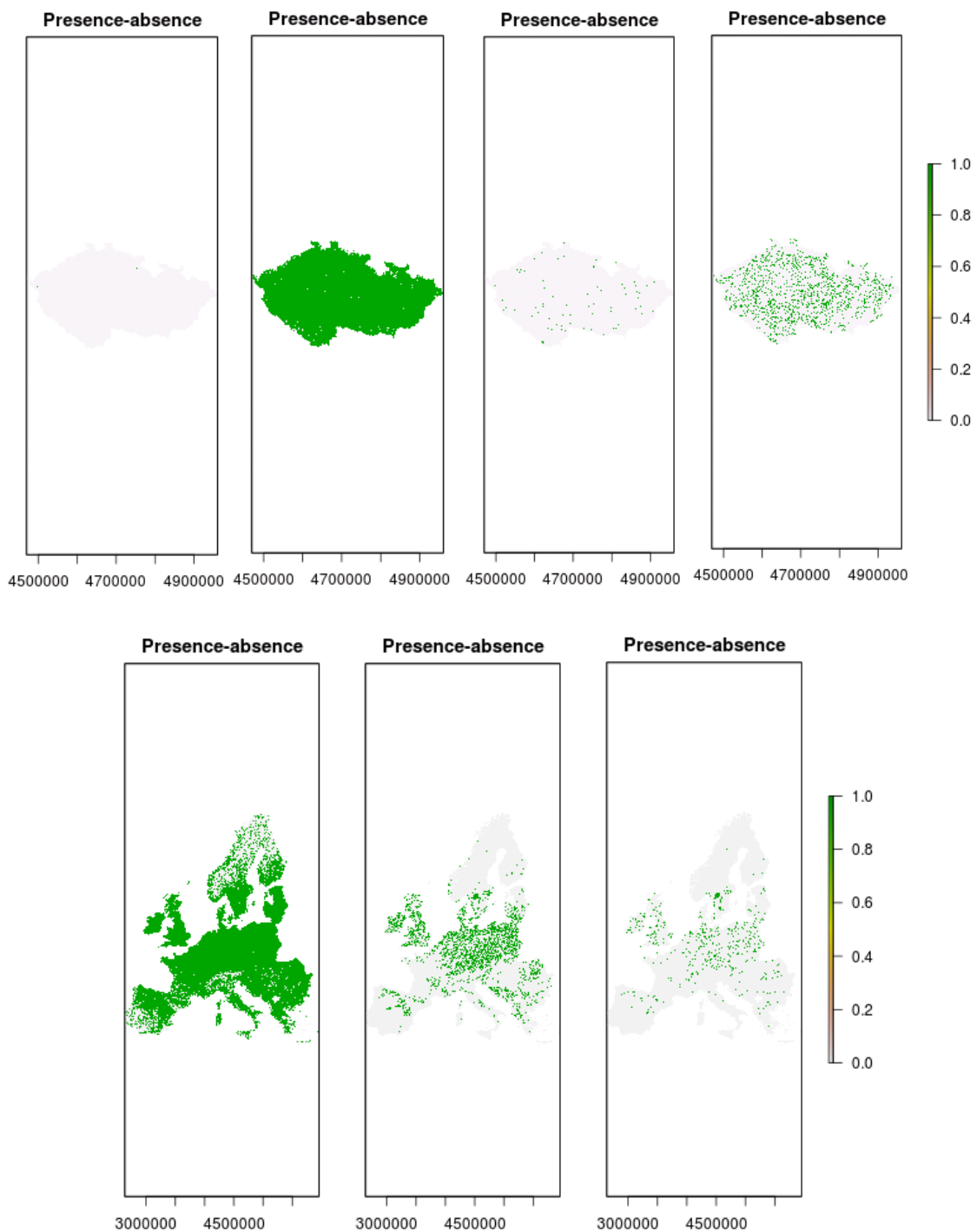
**Obrázek 6** Vhodnost území pro výskyt druhu *my.sp.cz*



**Obrázek 7** Vhodnost území pro výskyt druhu *my.sp.eu*

```
pa_cz<-convertToPA(my.sp.cz, plot=TRUE)
```

Jak již bylo zmíněno, presence a absence jsou z vhodnosti prostředí odvozeny na základě pravděpodobností. Bude-li funkce `convertToPA` spuštěna opakovaně, bude jejím výsledkem vždy odlišná mapa, ačkoliv bude sestavena podle stejných hodnot. Pro ilustraci jsou níže zachyceny výsledky opakovaně generovaných výskytů.



**Obrázek 8** Různé mapy výskytu virtuálních druhů při opakování funkce `convertToPA`

Vzhledem k tomu, že ve skutečnosti by těžko mohl nastat případ, kdy bude perfektně prozkoumáno opravdu celé území, byl ještě proveden sampling presencí a absencí (funkce `sampleOccurrences`). V rámci samplingu je možno stanovit, kolik vzorků má být vytvořeno. Tento parametr byl nastaven na 4000 bodů pro Evropu a 400 bodů pro ČR.

```
PA.pts_cz<- sampleOccurrences(pa_cz, n=400, type="presence-
absence")
```

Aby mohla být takto vzniklá data využita pro modelování druhové distribuce, bylo třeba nasamplované presence/absence vyexportovat a vložit do tabulky s hodnotami jednotlivých prediktorů. Data byla tedy uložena ve formátu `.txt` a pak přes Add XY Data nahrána do ArcMapu. Dále byla převedena na rastr (Points to Raster) a pak byl pomocí Extract Values to Points vytvořen shapefile zachycující výskyt druhu. Tato vrstva pak byla pomocí Intersect sloučena s původní vrstvou kvadrátů, takže vznikla nová vrstva `cz_occ` (resp. `eu_occ`). Vrstvy `cz_occ` (`eu_occ`) obsahují pouze ty kvadráty, v nichž byl sampling proveden. V atributových tabulkách pak přibyl další sloupec s názvem occurrence, který obsahuje hodnoty 0 (absence) a 1 (presence). Atributové tabulky byly nakonec opět vyexportovány do formátu `.csv`. Příklad výsledné tabulky je uveden níže.

square	WC	SUM_CLC_A	SUM_DIB_A	SUM_OSM_A	SUM_GIW_A	occurrence
60	6,885095	0	0	0,103369	0	0
634	6,531147	0	0,000159	0,00354	0	0
818	6,139344	0	0,000989	0,000962	0	1
820	6,362903	0	0	0,021057	0	1

**Tabulka 5** Atributová tabulka vrstvy `cz_occ`

Pro srovnání byl ještě vygenerován další virtuální druh, jehož výskyt závisí pouze na vodních plochách. Postup v R byl totožný jako v původním případě. Jediným rozdílem bylo vynechání teploty v celém procesu. Celkem tedy byly vytvořeny čtyři různé virtuální druhy, jejichž přehled je uveden v tabulce č. 6. Příklad celého skriptu použitého pro vytvoření virtuálního druhu je uveden v Příloze č. 4.

virtuální druh	parametry
cz_1	CLC, WC
cz_2	CLC
eu_1	CLC, WC
eu_2	CLC

**Tabulka 6** Souhrn vytvořených virtuálních druhů a parametrů



### 3.3.4 Logistická regrese

Logistická regrese je metoda matematické statistiky, která odhaduje pravděpodobnost výskytu diskrétní závisle proměnné ( $y$ ), kterou v našem případě představuje výskyt virtuálního druhu, na spojitých nezávisle proměnných (prediktorech)  $X_1, \dots, X_n$ . Jedná se o vyjádření střední hodnoty vysvětlované proměnné jako funkce nezávisle proměnných. Střední hodnota je v tomto případě rovna pravděpodobnosti výskytu sledovaného jevu.

Vzhledem k tomu, že modelovaná veličina  $Y$  zde může nabývat pouze dvou hodnot (0 a 1), jedná se o speciální případ binomického rozdělení, kterým je alternativní rozdělení s parametrem  $\mu$ , tedy  $Y \sim A(\mu)$ . Pravděpodobnost výskytu obou možných hodnot lze zapsat jako

$$P(Y=k) = \mu^k (1 - \mu)^{1-k} \quad \begin{array}{l} k = 0; 1 \\ 0 < \mu < 1 \end{array} \quad (\text{Zvára, 2008}).$$

Přímo pak můžeme dopočítat střední hodnotu a rozptyl veličiny  $Y$ .

$$EY = \mu \quad DY = \mu(1-\mu)$$

Pokud by byl vytvořen model  $P(Y = 1) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$ , mohlo by se stát, že odhady koeficientů  $\beta_0, \beta_1, \dots, \beta_n$  pro některé realizace  $x_1, \dots, x_n$  veličin  $X_1, \dots, X_n$  vyjdou mimo interval  $(0;1)$ , protože prediktor může nabývat všech hodnot z  $\mathbb{R}$ , tedy z množiny reálných čísel (Zvára, 2008).

Tento problém řeší výpočet tzv. šance (odds), která vyjadřuje, kolikrát vyšší je pravděpodobnost, že  $Y$  má hodnotu 1, než pravděpodobnost, že má hodnotu 0. Výsledná šance leží v intervalu  $(0; \infty)$ .

$$\text{odds}(P(Y = 1)) = \frac{P(Y = 1)}{P(Y = 0)} = \frac{P(Y = 1)}{1 - P(Y = 1)}$$

Interval  $(0; \infty)$  je následně potřeba transformovat logitovou funkcí tak, aby byl rozšířen na celou množinu reálných čísel, tedy na interval  $(-\infty; \infty)$

$$\text{logit}(P(Y = 1)) = \ln(\text{odds}(P(Y = 1))) = \ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right)$$

Za předpokladu, že  $\beta = (\beta_0, \beta_1, \dots, \beta_n)'$  a  $X = (1, X_1, \dots, X_n)'$ , lze vzorec pro pravděpodobnost vyjádřit jako

$$P(Y = 1) = \frac{1}{1 + e^{-X'\beta}}$$

Tato pravděpodobnost je podmíněná, jelikož pro různé realizace  $x$  náhodného vektoru  $X$  nabývá různých hodnot

$$P(Y = 1|X = x) = \frac{1}{1 + e^{-x'\beta}}$$

Výše uvedeným modelem predikujeme střední hodnotu náhodné veličiny  $Y$  v závislosti na realizacích  $x$ . Hodnoty koeficientů  $\beta = (\beta_0, \beta_1, \dots, \beta_n)'$ , kde  $0 < \mu_i < 1$ , se odhadují metodou maximální věrohodnosti (maximum likelihood estimation) (Zvára, 2008).

Pokud je předmětem logistické regrese diskrétní (binární) závisle proměnná a spojitá nezávisle proměnná, je šance při zvolené hodnotě  $x$  rovna

$$odds = \frac{\frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}}{\frac{1}{1 + \exp(\beta_0 + \beta_1 x)}} = e^{\beta_0 + \beta_1 x}$$

Porovnáme-li šance výsledku pokusu pro dvě hodnoty nezávisle proměnné, lišící se o jednotku, dostaneme

$$\frac{odds(x=1)}{odds(x=0)} = \frac{e^{(\beta_0 + \beta_1)}}{e^{\beta_0}} = e^{\beta_1}$$

Parametr  $\beta_1$  pak informuje o změně logaritmu poměru šancí k jednotkovému přírůstku nezávisle proměnné  $x$  (Zvára, 2008).

### 3.3.5 Modelování

Pro účely modelování byl využit statistický software R 3.1.2., do jehož prostředí byly postupně nahrány připravené tabulky eu.csv a cz.csv. Na základě dat ze vstupních tabulek byly vytvořeny modely, které vyjadřují vztah mezi výskytem virtuálního druhu rozloze rozlohou vodních ploch podle databází OSM a GIW (resp. OSM, DIBAVOD a GIW pro ČR). Pro srovnání byl model vytvořen i pro databázi CLC. Byla tedy opakovaně spuštěna funkce `glm`, do níž byly vkládány uvažované prediktory. Pro ilustraci je níže uveden konkrétní příkaz pro výskyt druhu v evropském měřítku, v němž jako prediktor figuruje rozloha vodních ploch dle OSM.

```
eu_osm<glm(occurrence~SUM_OSM_A, family=binomial, data=eu)

summary(eu_osm)
```

Výsledky analýzy byly vyvolány funkcí `summary` a její výsledky byly zaznamenány do tabulek v Přílohách č. 3 a 4. Funkce `summary` zahrnuje tyto údaje (Crawley, 2007):

- (Intercept)                      absolutní člen ( $\beta_0$ ), který představuje hodnotu lineárního prediktoru pro výskyt, pokud se rozloha změní o jednotku

- SUM\_OSM\_A odhadovaná hodnota ( $\beta_1$ ), o kterou se změní šance výskytu, pokud rozloha bude o jednotku vyšší; znaménko před číslem pak udává, zda odhadovaná hodnota ovlivňuje výskyt pozitivně či negativně
- Std. Error střední chyba odhadu
- z value hodnota Waldova testu
- Pr (>|z|) interval spolehlivost Waldova testu
- Signif. codes významnost
- AIC Akaikeho informační kritérium
- Null deviance vypovídá o tom, jak dobře model předpovídá výskyt druhu v případě, že by nebyly zahrnuty vysvětlující proměnné
- Residual deviance ukazuje, jak dobře model předpovídá výskyt druhu, pokud zahrneme prediktory

Na základě vypočtených hodnot  $\beta_0$  a  $\beta_1$  lze dopočítat pravděpodobnost výskytu druhu v jednotlivých kvadrátech dle vzorce

$$p_i = \frac{e^{(\beta_0 + \beta_1 x_i + \beta_2 y_i)}}{1 + e^{(\beta_0 + \beta_1 x_i + \beta_2 y_i)}}$$

### 3.4 Zhodnocení modelů

Jako nejjednodušší ukazatel vypovídající o rozdílech jednotlivých modelů může sloužit AIC (Akaikeho informační kritérium), jehož hodnota je přímým výstupem funkce glm v R. Výsledná hodnota sama o sobě žádnou vypovídací hodnotu o kvalitě modelu nemá. Může být ovšem využita pro výběr nejlepšího modelu, pokud je porovnáváno více modelů. AIC přiřazuje vyšší hodnotu těm modelům, které jsou komplikovanější, tedy pracují s více prediktory. Při srovnávání by měl být za nejlepší označen ten model, který má nejnižší hodnotu AIC a je tím pádem nejjednodušší.

Hlavním kritériem při srovnání výsledných modelů byla hodnota AUC (Area Under Curve). Jedná se o plochu pod ROC křivkou (receiver operating characteristic curve), která bývá využívána při zařazování objektů do dvou možných tříd v případě, že je známo, že objekt náleží právě do jedné z těchto tříd (presence/absence) (Kutálek, 2010). AUC podává informaci o kvalitě rozhodovacího kritéria a bývá velmi často používáno jako evaluační metoda.

Aby mohla být ROC křivka vytvořena, je třeba rozdělit objekty binární klasifikací do čtyř skupin (Pearce et Ferrier, 2000):

- true positive – objekty jsou pozitivní a klasifikačním modelem byly vyhodnoceny také jako pozitivní,
- false positive – objekty jsou negativní, ale klasifikačním modelem byly nesprávně vyhodnoceny jako pozitivní,

- C. false negative – objekty jsou pozitivní, ale klasifikačním modelem byly nesprávně vyhodnoceny jako negativní,
- D. true negative – objekty jsou negativní a klasifikačním modelem byly vyhodnoceny také jako negativní.

	zaznamenaná presence	zaznamenaná absence	
vyhodnocená presence	A (true positive)	B (false positive)	A + B
vyhodnocená absence	C (false negative)	D (true negative)	C + D
	A + C	B + D	A + B + C + D

**Tabulka 7 Binární klasifikační tabulka pro ROC**

Na základě výše uvedených skupin bývají ještě určovány hodnoty tzv. senzitivity a specificity. Senzitivita je podmíněná pravděpodobnost, s jakou výsledek testu nabývá hodnoty 1. Je to tedy relativní četnost správně klasifikovaných pozitivních případů vztážená k veškerým zaznamenaným presencím daná poměrem  $A/(A+C)$ . Specificita je podmíněná pravděpodobnost, s jakou výsledek testu nabývá hodnoty 0. Je to tedy relativní četnost správně klasifikovaných absencí vztážená k veškerým absencím daná poměrem  $D/(B+D)$  (Fawcett, 2006).

ROC křivka je dvojrozměrné vyjádření tzv. false positive rate (poměr  $B/(B+D)$ ) na ose x a tzv. true positive rate (poměr  $A/(A+C)$ ) na ose y. Hodnoty na obou osách nabývají pouze hodnot 0 a 1, takže plocha celého čtverce, v němž je ROC křivka vykreslena je roven 1. Hodnota 1 je tedy maximum, kterého může ukazatel AUC nabývat. Teoreticky by se žádný realistický model neměl dostat na AUC nižší než 0,5.

Výsledné hodnoty AUC modelů druhové distribuce lze klasifikovat podle tabulky č. 8 (Bortlíček, 2008):

hodnota AUC	hodnocení
0,9 – 1,0	výborně
0,8 – 0,9	velmi dobře
0,7 – 0,8	dobře
0,6 – 0,7	dostatečně
0,5 – 0,6	nedostatečně

**Tabulka 8 Klasifikace AUC**

Poslední evaluace byla provedena pomocí kalibračních grafů, které umožňují porovnání kvality jednotlivých modelů z hlediska predikovaných pravděpodobností. Pro tvorbu takového modelu je třeba nejprve vypočítat pravděpodobnost výskytu druhu dle výsledků glm. Hodnoty se pak rozdělí do několika skupin (například po 0,1) a pro každou skupinu se spočítá poměr pozorovaných presencí k celkovému počtu kvadrátů ve skupině. Na ose x figurují středy každé skupiny, vypočítaný poměr se pak vynese na osu y. V rámci modelu se také vykresluje přímkou se sklonem 45°. Křivka modelu, který je dobře kalibrovaný, by se pak měla nacházet v těsné blízkosti této přímky a měla by ji ideálně svým průběhem kopírovat (Pearce et Ferrier, 2000).

## 4. VÝSLEDKY

Hodnoty Spearmanova korelačního koeficientu pro všechny uvažované velikosti kvadrátů na území Evropy vycházely vyšší než 0,6. Dle klasifikační tabulky se jednalo v případě kvadrátů velikosti 50 km i 10 km o velmi silnou korelaci. Jako nejvhodnější pro použití pro tvorbu virtuálního druhu a následné modelování druhové distribuce byly tedy zvoleny kvadráty velikosti 5 km (plocha 25 km<sup>2</sup>), jejichž korelační koeficienty byly vyhodnoceny jako silně korelované a jsou uvedeny v tabulce č. 9

	SUM_CLC_A	SUM_OSM_A	SUM_GIW_A
SUM_CLC_A	1	0,6541186	0,719066
SUM_OSM_A		1	0,6850134
SUM_GIW_A			1

Tabulka 9 Spearmanův korelační koeficient pro Evropu

V rámci České republiky byla situace obdobná, tedy hodnoty korelačních koeficientů byly vysoké u větších kvadrátů – v případě kvadrátů velikosti 10 km se jednalo o velmi silnou korelaci, u 5km kvadrátů o korelaci silnou. Pro další práci byl tedy opět zvolen nejmenší z připravených rozměrů a sice 1 km (1 km<sup>2</sup>). Korelační koeficienty v tomto případě dosahovaly slabých či mírných interakcí, jak dokazuje tabulka č. 10. Výjimku tvořila silná korelace mezi databázemi DIBAVOD a OSM.

	SUM_CLC_A	SUM_DIB_A	SUM_OSM_A	SUM_GIW_A
SUM_CLC_A	1	0,3145645	0,3221737	0,5076195
SUM_DIB_A		1	0,7914303	0,5809673
SUM_OSM_A			1	0,5655975
SUM_GIW_A				1

Tabulka 10 Spearmanův korelační koeficient pro ČR

### 4.1 Srovnání modelů dle AIC

Jak již bylo uvedeno, AIC penalizuje ty modely, v nichž figuruje více proměnných. Lepších hodnot tak bylo dosaženo v případech, kde jako prediktory vystupují pouze rozlohy vodních ploch. Vzhledem k tomu, že už samotný výskyt druhu byl modelován v jiných podmínkách, nelze tohoto kritéria využít pro srovnání modelů distribuce druhů vytvořených bez teploty s modely druhů vytvořenými s teplotou. Hodnoty AIC byly tedy využity jen ke kontrole modelů v rámci jednotlivých srovnatelných kategorií.

<b>Model a prediktory</b>	<b>AIC</b>
EU ~ CLC, WC	2146,1
EU ~ GIW, WC	2161,3
EU ~ OSM, WC	2217,9
EU ~ CLC	23,444
EU ~ GIW	48,821
EU ~ OSM	144,01

**Tabulka 11 Hodnoty AIC pro Evropu**

<b>Model a prediktory</b>	<b>AIC</b>
CZ ~ CLC, WC	491,98
CZ ~ GIW, WC	494,31
CZ ~ DIB, WC	494,93
CZ ~ OSM, WC	495,15
CZ ~ CLC	174,3
CZ ~ DIB	183,11
CZ ~ GIW	186,38
CZ ~ OSM	187,46

**Tabulka 12 Hodnoty AIC pro ČR**

## 4.2 Srovnání modelů dle AUC

Vypočtené hodnoty AUC modelů na území vybraných evropských států vykazují velmi vysoké hodnoty pro všechny modely zahrnující jak rozlohu vodních ploch tak teplotu. Ještě o něco málo vyšší jsou tyto hodnoty u modelů databází CLC a GIW v případě, že teplota zahrnuta nebyla. Bez teploty ovšem značně pokleslo AUC modelu vycházejícího z databáze OSM.

<b>Model a prediktory</b>	<b>AUC</b>
EU ~ CLC, WC	0,9590941
EU ~ GIW, WC	0,958479
EU ~ OSM, WC	0,9560508
EU ~ CLC	0,9997887
EU ~ GIW	0,9673761
EU ~ OSM	0,8339522

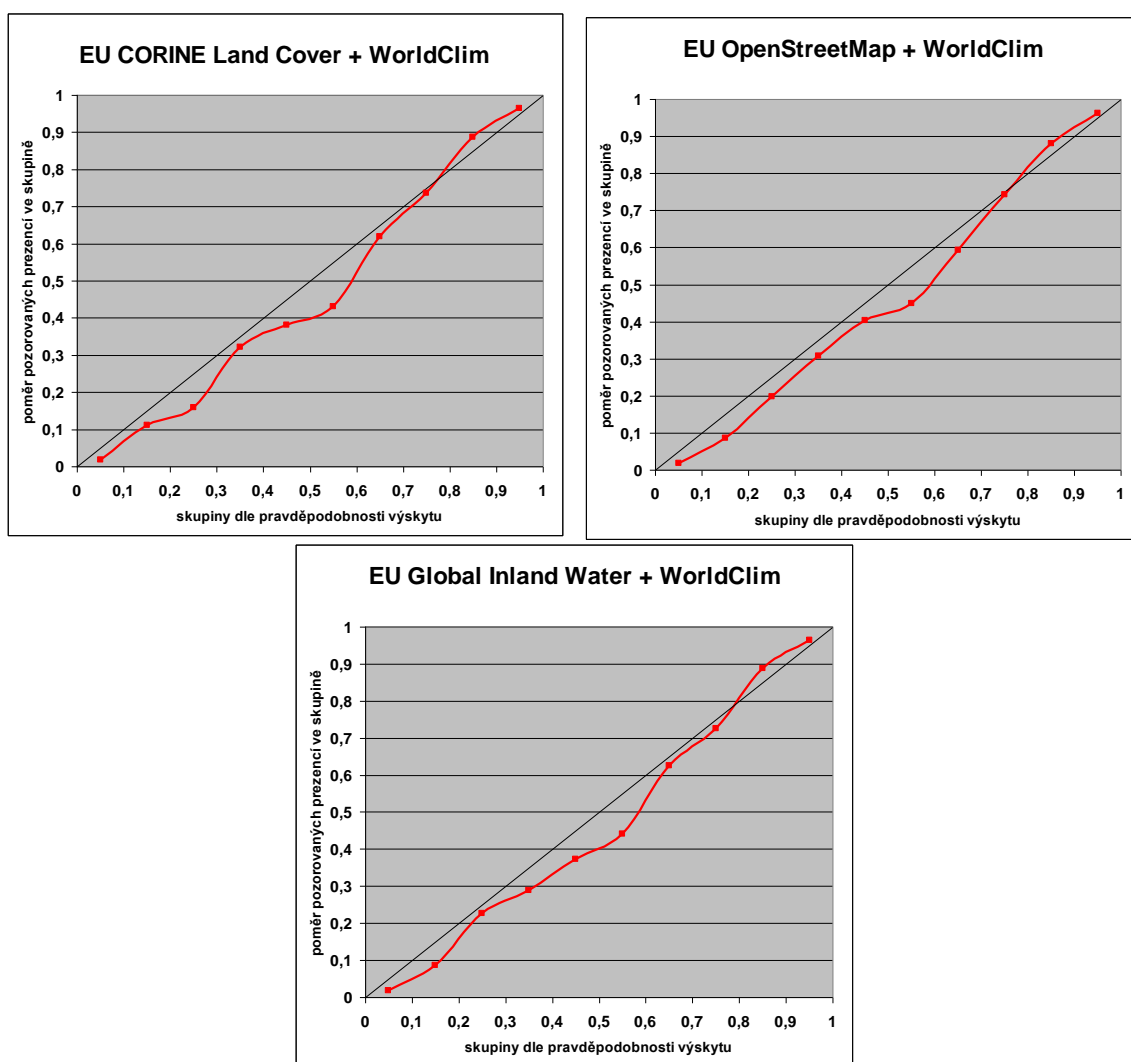
**Tabulka 13 Hodnoty AUC pro Evropu**

U modelů vytvořených na území České republiky na základě rozlohy vodních plocha a teploty byly hodnoty AUC dostatečné. Po vynechání teploty při tvorbě virtuálního druhu i v GLM byly hodnoty nižší a tudíž klasifikované jako nedostatečné.

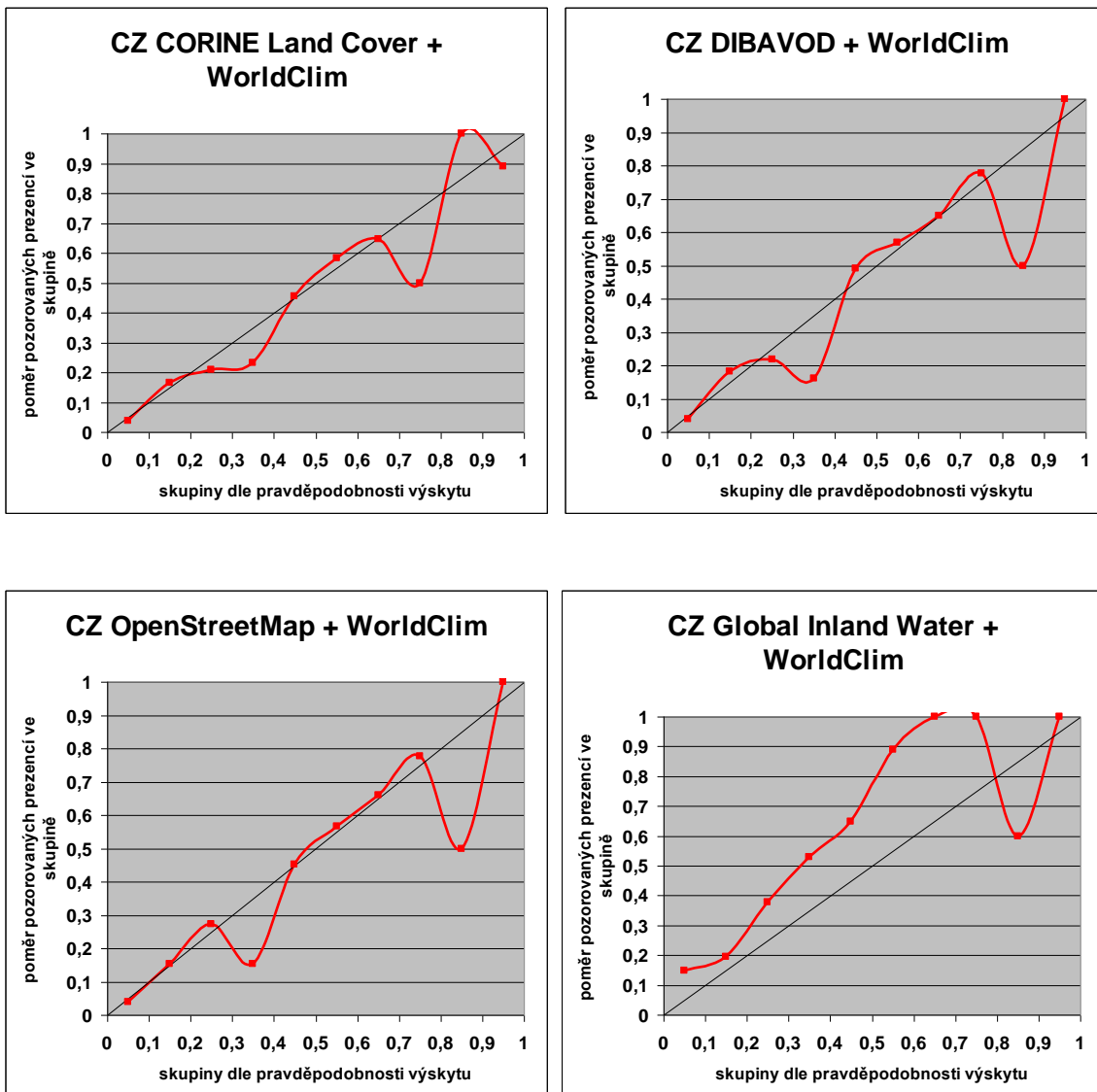
Model a prediktory	AUC
CZ ~ GIW, WC	0,6873547
CZ ~ OSM, WC	0,6826286
CZ ~ DIB, WC	0,6825036
CZ ~ CLC, WC	0,6820035
CZ ~ DIB	0,5528075
CZ ~ OSM	0,5544015
CZ ~ GIW	0,5364562
CZ ~ CLC	0,5291547

Tabulka 14 Hodnoty AUC pro ČR

### 4.3 Srovnání modelů dle kalibračních grafů



Obrázek 9 Kalibrační grafy modelů počítajících s průměrnou teplotou (Evropa)



Obrázek 10 Kalibrační grafy modelů počítajících s průměrnou teplotou (ČR)



Nakonec byly ještě sesumírovány vodní plochy dle jednotlivých databází pro vybrané velikosti kvadrátů v každém z uvažovaných území. Tyto údaje byly spočteny, aby mohly být použity při hledání vysvětlení dosažených výsledků.

<b>EVROPA (5 x 5 km<sup>2</sup>)</b>	<b>CLC</b>	<b>OSM</b>	<b>GIW</b>
Celková rozloha vodních ploch (km <sup>2</sup> )	124 649	157 362	139 482
Průměrná rozloha vodních ploch v kvadrátu ( km <sup>2</sup> )	0,60	0,76	0,67
Počet kvadrátů obsahujících vodní plochy	60 233	148 112	125 420
Počet kvadrátů neobsahujících vodní plochy	148 009	60 130	82 822

**Tabulka 15** Souhrnné hodnoty vodních ploch jednotlivých databází (Evropa)

<b>ČESKÁ REPUBLIKA (1 x 1 km<sup>2</sup>)</b>	<b>CLC</b>	<b>OSM</b>	<b>GIW</b>	<b>DIB</b>
Celková rozloha vodních ploch (km <sup>2</sup> )	565	1 202	670	835
Průměrná rozloha v. p. v kvadrátu ( km <sup>2</sup> )	0,007	0,015	0,008	0,01
Počet kvadrátů obsahujících vodní plochy	3 176	44 274	11 421	36 781
Počet kvadrátů neobsahujících vodní plochy	76 830	35 732	68 585	43 225

**Tabulka 16** Souhrnné hodnoty vodních ploch jednotlivých databází (ČR)

## 5. DISKUSE

Vypočítané hodnoty Spearmanova korelačního koeficientu naznačují, že v evropském měřítku by při studiích využívajících velkých kvadrátů neměl být mezi jednotlivými databázemi významný rozdíl. Mělo by tedy být relativně jedno, zda bude v hrubých měřítcích pro modelování použita databáze CORINE Land Cover, OpenStreetMap či Global Inland Water. Obdobný závěr bychom mohli učinit u dvou větších kvadrátů modelovaných v České republice. Naopak výsledky dosažené u kvadrátů velikosti 1 km<sup>2</sup> v ČR napovídají, že mezi databázemi by měly být patrné rozdíly. Výjimku tvoří silná korelace mezi databázemi DIBAVOD a OpenStreetMap. Je tedy možno usuzovat, že při modelování druhové distribuce by měly tyto dvě databáze dosahovat srovnatelných výsledků.

K modelování virtuálních druhů byly nejprve využity rozlohy vodních ploch databáze CORINE Land Cover a průměrná teplota v jarních měsících dle databáze WorldClim. Virtuálnímu druhu byly přisouzeny určité odpovědní funkce, které byly zvoleny autorkou, a jedná se tedy o volbu čistě subjektivní. Zrovna tak byly experimentálně zvoleny i hodnoty parametrů obou funkcí. Při výběru těchto hodnot je třeba dbát na to, co přesně je cílem studie. Dle Guisan et Zimmermann (2000) by měl být SDM vždy realistický, takže i do něj vstupující proměnné by měly být v jistém souladu s tím, co v přírodě může existovat. Pro posouzení, zda dosažené hodnoty prvního typu modelů nejsou příliš zkresleny teplotou, byly dále ještě vymodelovány další virtuální druhy, jejichž výskyt závisí čistě na rozloze vodních ploch. Tyto druhy však podmínku o realističnosti téměř jistě nesplňují, neboť druh, který by se vykytoval na všech místech s dostatečně velkou vodní plochou na světě těžko najdeme.

Při hledání možných příčin dosažených výsledků byly prozkoumány tabulky obsahující souhrnné informace o rozlohách vodních ploch v jednotlivých kvadrátech. V Evropě bylo zaznamenáno 60 233 kvadrátů obsahujících vodní plochu dle databáze CLC, 148 112 kvadrátů s vodní plochou dle OSM a 125 420 s vodní plochou dle GIW. Dáme-li tyto hodnoty do souvislosti s hodnotami AUC, lze říci, že čím více kvadrátů je vodními plochami obsazeno, tím horších výsledků model dosahuje. Na území České republiky se z uvedených statistik žádný podobný trend vyzorovat nepodařilo. Ani kalibrační grafy nenaznačují jiné výsledky. V rámci Evropy nejsou v grafech modelů využívajících rozlohy vodních ploch i průměrnou teplotu žádné viditelné rozdíly. U modelů vytvořených pro Českou republiku bylo horších výsledků dosaženo pouze u databáze Global Inland Water.

Ve výsledcích by bylo možné zohlednit také další charakteristiky vodních ploch jako je například jejich obvod či počet vodních ploch v každém kvadrátu. Tyto aspekty však byly vynechány zejména kvůli použití databáze Global Inland Water, která je jako jediná z datových sad dostupná pouze v rastrovém formátu. Zjištění výše uvedených detailních údajů je tak téměř nemožné.

Důležitou roli by bezpochyby mělo hrát rozlišení a metoda tvorby jednotlivých databází. Obecně lze předpokládat, že databáze, které jsou odvozeny pouze ze satelitních snímků by měly být zatíženy větší nepřesností. Ta by v případě databáze Global Inland Water mohla být navíc umocněna tím, že pro diplomovou práci musela být převedena z rastrového formátu tak, aby bylo možno spočítat rozlohu vodních ploch. Při převezech mezi formáty a souřadnicovými systémy mohlo dojít k částečnému zkreslení některých dat. Jistá odchylka může být také způsobena upřednostněním výskytu vody na místech, kde se překrývalo více stažených rastrových čtverců.

Problémem databáze OpenStreetMap naopak může být nedostatečný průzkum některých lokalit a z něj vycházející absence dat pro tyto oblasti. Tento fakt demonstroval například Bittner (2017), který se zabýval dostupností mapových podkladů OSM na území Izraele a Palestiny. V jeho práci je viditelné, že v Pásmu Gazy nebyly ještě v roce 2008 v rámci OSM vyznačeny žádné cestní sítě, zatímco v roce 2014 již byla oblast poměrně dobře zmapována. Tyto chybějící informace mohou být případem zemí, v nichž probíhají válečné konflikty, nebo se může jednat o nátlak určitých skupin, které nemají zájem o mapování některých území (Bittner, 2017). Dalším problémem pak zůstává jistá prostorová nepřesnost, kterou zkoumali například Estima et Painho (2013), kteří prokázali 76% shodu map OSM s oficiálními mapami Portugalska.

Nejpřesnější data pro Českou republiku by teoreticky měla nabídnout databáze DIBAVOD, která je tvořena odborníky z Výzkumného ústavu vodohospodářského T. G. Masaryka a je průběžně doplňována. Podle statistik jednotlivých databází a rozdílových map uvedených v Metodice se však zdá, že databáze OpenStreetMap je ještě podrobnější než DIBAVOD.

Hlavní překážkou při srovnávání dat z různých databází vodních ploch se zdá být rozdílnost zobrazení a klasifikace jednotlivých prvků. Zatímco CORINE Land Cover zobrazují i vodní toky jako polygony, což znamená, že i jejich plochy lze spočítat a při práci použít, datové sady OpenStreetMap a DIBAVOD tuto možnost nenabízí. Obě zachycují vodní toky jako liniové prvky a tudíž by bylo možno počítat maximálně s jejich délkou. V tom případě by opět nebylo možno porovnávat s Global Inland Water, kde je délka stejně těžko určitelná, jako obvod vodních ploch. Databáze GIW však přesto obsahuje i vodní toky, které tedy do výpočtů vstoupily jako součást celkové rozlohy vodních ploch.

Rozdílná klasifikace je pak komplikací u CORINE Land Cover a DIBAVOD v tom smyslu, že kromě samostatných vrstev vodních toků a vodních nádrží jsou v nich vyčleněny ještě další vrstvy jako například močály, bažiny či pobřežní laguny. U OpenStreetMap může být jiná klasifikace určitých prvků dána opět tím, že databázi tvoří dobrovolníci. Při vizualizaci jednotlivých databází je také patrné, že některé z nich zachycují i velké vodní toky, ačkoliv byly staženy pouze jako vrstvy vodních nádrží.

U OpenStreetMap bylo navíc zjištěno, že vrstva vodních ploch obsahuje některé plochy dvakrát, přičemž tyto plochy pokaždé mají tytéž parametry (ID, název a rozlohu), ale jsou klasifikovány jako jiný druh vodní plochy (např. jednou jako water, podruhé jako reservoir). Pokud by tedy nebyly duplicitní záznamy předem vymazány, byly by plochy započteny dvakrát a opět by došlo ke zkreslení výsledků. Je tedy zřejmé, že před samotným modelováním je nutné provést důkladnou revizi vstupních dat, což může být problematické především u studií prováděných na velkých plochách ale v podrobném rozlišení. Limitujícím faktorem může být také dostupná výpočetní technika.

Studii by bylo vhodné dále rozvést a podrobit dalšímu výzkumu. Například by stálo za úvahu rozdělení území na menší kvadráty, než jaké byly využity v této diplomové práci. Prostor pro další výzkum nabízí také správné nastavení parametrů funkcí, kterými bude druh na zvolené environmentální prediktory odpovídat, a zvážení využití dalších vrstev databází CLC a DIBAVOD. Bylo by možné využít také dalších databází zachycujících vodní plochy (v České republice například národní databázi ArcČR 500 či DMÚ 25).

## 6. ZÁVĚR

V diplomové práci byla vytvořena rešerše shrnující současné poznatky ze studované problematiky. Rešerše posloužila při volbě některých důležitých cest vedoucích k dosažení dílčích cílů práce. Na základě známých studií modelování druhové distribuce byly vybrány možné velikosti kvadrátů, které by mohly být pro zájmové území vhodné a opodstatněné. Podle Spearmanova korelačního koeficientu byly za nejlepší vyhodnoceny kvadráty velikosti 5 km (25 km<sup>2</sup>) pro území Evropy a 1 km (1 km<sup>2</sup>) pro území České republiky.

Virtuální druh byl vytvořen pro obě území v uvedených velikostech kvadrátů v závislosti na rozloze vodních ploch databáze CORINE Land Cover a průměrné teplotě v jarních měsících dle databáze WorldClim. Pro srovnání byl ještě vytvořen další druh závislý pouze na rozloze vodních ploch. Pro vygenerování druhů byl použit softwarový balíček virtualspecies. Z dvou možných přístupů k převedení vhodnosti prostředí na presence a absence byl zvolen přístup pravděpodobnostní. Následně byl ještě proveden sampling presencí a absencí. Takto získaná binární data byla následně využita k modelování druhové distribuce.

Druhová distribuce byla postupně modelována za použití průměrné jarní teploty databáze WorldClim a rozloh vodních ploch všech databází. Model s CORINE Land Cover byl vytvořen jako referenční. SDM byly rovněž vytvořeny i pro virtuální druhy (jeden evropský, jeden český), pro jejichž tvorbu nebyla využita teplota. Modely jejich distribuce tedy opět zohledňovaly pouze rozlohy vodních ploch jednotlivých databází.

Všechny cíle, které byly autorkou stanoveny, byly splněny. Výsledky ovšem naznačují, že nelze jednoznačně stanovit, při jaké velikosti kvadrátů přestává mít přesnost použitých datových sad vliv na modely druhové distribuce. Hodnoty korelačních koeficientů by napovídaly, že v Evropě bude tato velikost kvadrátu menší než 5 km. V České republice se zdálo, že při velikosti kvadrátu 1 km už by měly být rozdíly mezi jednotlivými databázemi patrné, avšak podle ukazatele AUC ani podle vykreslených kalibračních grafů nelze tuto teorii spolehlivě potvrdit.

Problematika by zasloužila další výzkum s použitím menších kvadrátů a s rozsáhlejším výzkumem parametrů, které by bylo možno zohlednit při tvorbě virtuálního druhu.

## 7. SEZNAM POUŽITÉ LITERATURY

ARAÚJO M.B., THUILLER W., WILLIAMS P.H., REGINSTER I., 2005: Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography* 14: 17 – 30.

BEDIA J., HERRERA S., GUTIÉRREZ J.M., 2013: Dangers of using global bioclimatic datasets for ecological niche modeling. Limitations for future climate projections. *Global and Planetary Change* 107: 1 - 12.

BITTNER C., 2017: *Political Geography* 57: 34 – 48.

BOMBI P., D'AMEN M., 2012: Scaling down distribution maps from atlas data: a test of different approaches with virtual species. *Journal of Biogeography* 39: 640 – 651.

BORTLÍČEK Z., 2008: ROC křivky. [Diplomová práce], Masarykova univerzita v Brně, Přírodovědecká fakulta.

BOSCUCCI F., MARCORIN I., SIGURA M., BRESSAN E., TAMBERLICH F., VIANELLO A., CASOLO V., 2015: Distribution modeling of seagrasses in brackish waters of Grado-Marano lagoon (Norther Adriatic Sea). *Estuarine, Coastal and Shelf Science* 164: 183 – 193.

BUDIC L., DIDENKO G., DORMANN C.F., 2016: Squares of different sizes: effect of geographical projection on model parameter estimates in species distribution modeling. *Ecology and Evolution* 2016: 202 – 211.

CENIA, 2016: CORINE 2012. Online: <http://www1.cenia.cz/www/projekty-cenia/CORINE>, cit. 18.11.2016.

DE CLERCQ E.M., LETA S., ESTRADA-PEÑA A., MADDER M., ADEHAN S., VANWAMBEKE S.O., 2015: Species distribution modelling for *Rhipicephalus microplus* (Acari: Ixodidae) in Benin, West Africa: Comparing datasets and modelling algorithms. *Preventive Veterinary Medicine* 118: 8 – 21.

DIBAVOD, 2017. Online: <http://www.dibavod.cz/17/geodatabaze-dibavod.html>, cit. 28.02.2017.

DUAN R.-Y., KONG X.-Q., HUANG M.-Y., WU G.-L., WANG Z.-G., 2015: SDMvspecies: a software for creating virtual species for species distribution modelling. *Ecography* 38: 108-110.

EEA, 2016: CORINE Land Cover 2012 seamless vector data. CORINE Metodology. European Environment Agency. Online: <http://www.eea.europa.eu/>, cit. 20.12.2016.

ELITH J., LEATHWICK J.R., 2009: Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *The Annual Review of Ecology, Evolution and Systematics* 40: 677 – 697.

- ESTIMA J., PAINHO M., 2015: Investigating the Potential of OpenStreetMap for Land Use/Land Cover Production: A Case Study for Continental Portugal. *OpenStreetMap in GIScience*: 273 – 293.
- FAWCETT T., 2006: An introduction to ROC analysis. *Pattern Recognition Letters* 27: 861 – 874.
- FENG M., SEXTON J.O., CHANNAN S., TOWNSHEND J.R., 2016: A global, high-resolution (30-m) inland water body dataset for 2000: first results of a topographic-spectral classification algorithm. *International Journal of Digital Earth* 9:2: 113 – 133.
- FERNÁNDEZ M., HAMILTON H., ALVAREZ O., GUO Q., 2012: Does adding multi-scale climatic variability improve our capacity to explain niche transferability in invasive species? *Ecological Modelling* 246: 60 – 67.
- FLOUSEK J., GRAMSZ B., TELENSKÝ T., 2015: Ptáci Krkonoš – atlas hnízdního rozšíření 2012 – 2014. Správa Krkonošského národního parku, Vrchlabí: 471 s.
- GLOBAL LAND COVER FACILITY, 2016: Global Inland Water. Online: <http://glcf.umd.edu/data/watercover/>, cit. 28.11.2016.
- GOODCHILD M.F., 2011: Scale in GIS: An overview. *Geomorphology* 130: 5 – 9.
- GUISAN A., ZIMMERMANN N.E., 2000: Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147 – 186.
- GUISAN A., THUILLER W., 2005: Predicting species distribution: offering more than simple habitat models. *Ecology letters* 8: 993 – 1009.
- HIJMANS R.J., CAMERON S.E., PARRA J.L., JONES P.G., JARVIS A., 2005: Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965 – 1978.
- HIJMANS R.J., ELITH J., 2017: Species distribution modeling with R. Online: <https://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf>, cit. 10.01.2017.
- HIRZEL A.H., HELFER V., METRAL F., 2001: Assessing habitat-suitability models with a virtual species. *Ecological Modelling* 145: 111 – 121.
- HUANG M., KONG X., VARELA S., DUAN R., 2016: The Niche Limitation Method (NicheLim), a new algorithm for generating virtual species to study biogeography. *Ecological Modelling* 320: 197 – 202.
- HUIJIE Q., TOWNSEND PETERSON A., CAMPBELL L.P., SOBERÓN J., JI L., ESCOBAR L.E., 2016: NicheA: creating virtual species and ecological niches in multivariate environmental scenarios. *Ecography* 39: 1 – 9.
- KUTÁLEK D., 2010: Statistická analýza ROC křivek. [Diplomová práce], Vysoké učení technické v Brně, Fakulta strojního inženýrství.
- LEROY B., 2015: The virtualspecies R package: a complete tutorial. Online: <http://borisleroy.com/files/virtualspecies-tutorial.html>, cit. 03.01.2017.

- LEROY B., MEYNARD C.N., BELLARD C., COURCHAMP F, 2015: virtualspecies, an R package to generate virtual species distributions. *Ecography* 39: 599 – 607.
- MEYNARD C.N., KAPLAN D.M., 2013: Using virtual species to study species distributions and model performance. *Journal of Biogeography* 40: 1 – 8.
- MILLER J.A., 2014: Virtual species distribution models: Using simulated data to evaluate aspects of model performance. *Progress in Physical Geography* 38: 117 – 128.
- MOUDRÝ V., 2015: Modelling species distributions with simulated virtual species. *Journal of Biogeography* 2015: 1 – 2.
- MOUDRÝ V., ŠÍMOVÁ P., 2012: Influence of positional accuracy, sample size and scale on modelling species distributions: a review. *International Journal of Geographical Information Science*, 26:11: 2083 – 2095.
- MOUDRÝ V., KOMÁREK J., ŠÍMOVÁ P., 2017: Which breeding bird categories should we use in models of species distribution? *Ecological Indicators* 74: 526 – 529.
- OPENSTREETMAP, 2016: O projektu. Online: <http://www.openstreetmap.org/about>, cit. 22.12.2016.
- PEARCE J., FERRIER S., 2000: Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133: 225 – 245.
- PRION S., HAERLING K.A., 2014: Making Sense of Methods and Measurement: Spearman-Rho Ranked-Order Correlation Coefficient. *Clinical Simulation in Nursing* 10: 535 – 536.
- REIF J., VOŘÍŠEK P., ŠŤASTNÝ K., BEJČEK V., PETR J., 2008: Agricultural intensification and farmland birds: new insights from a central European country. *Ibis* 150: 596 – 605.
- ROBERTSON M.P., CUMMING G.S., ERASMUS B.F.N., 2010: Getting the most out of atlas data. *Diversity and Distributions* 16: 363 – 375.
- SKLENÁŘOVÁ L., 2011: Spolehlivost použitých geodat pro fitování biotopových efektů na prostorovou distribuci vybraných ptačích druhů. Diplomová práce.
- SZYMKOWIAK J., SKIERCZYNSKI M., KUCZYNSKI L., 2014: Are buntings good indicators of agricultural intensity? *Agriculture, Ecosystems and Environment* 188: 192 – 197.
- ŠMILAUER P., 1998 – 2007: Moderní regresní metody. Biologická fakulta JU. Online: <http://regent.jcu.cz/MRM.pdf>, cit. 5.4.2017.
- ŠŤASTNÝ K., BEJČEK V., HUDEC K., 2006: Atlas hnízdního rozšíření ptáků v České republice 2001 – 2003. Aventinum, Praha: 463 s. + 4 mapové přílohy.
- TORANZA C., BRAZEIRO A., MANEYRO R., 2016: Anfíbios amenazados de Uruguay: efectividad de las áreas protegidas ante el cambio climático. *Ecología Austral* 26: 138 - 149.



WHITTAKER R.J., ARAÚJO M.B., JEPSON P., LADLE R.J., WATSON J.E.M., WILLIS K.J., 2005: Conservation Biogeography: assessment and prospect. Diversity and Distributions 11: 3 –23.

WISE, 2008: European waters – overview. Online: <http://www.eea.europa.eu/themes/water/european-waters>, citováno 4.1.2017.

WORLDCLIM, 2016: Global Climate Data. Online: <http://www.worldclim.org/>, cit. 20.11.2016.

ZVÁRA K., 2008: Regrese. Matfyzpress, Praha: 253 s.

## **8. SEZNAM PŘÍLOH**

Příloha 1 Evropské státy v zájmovém území

Příloha 2 Rozdíly jednotlivých databází v jednom modelovém kvadrátu v Evropě

Příloha 3 Rozdíly jednotlivých databází v jednom modelovém kvadrátu v ČR

Příloha 4 Skript pro tvorbu virtuálního druhu, GLM a výpočet AUC pro Evropu

Příloha 5 Tabulka výsledků logistické regrese a AUC pro evropské státy

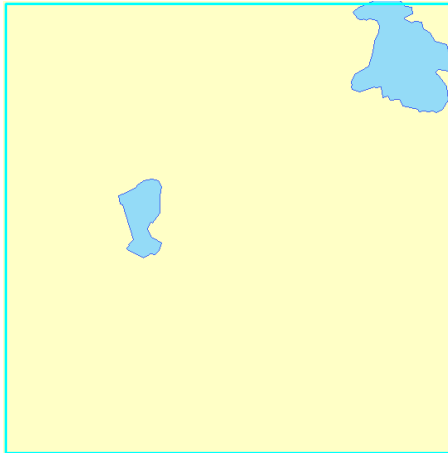
Příloha 6 Tabulka výsledků logistické regrese a AUC pro Českou republiku

## 9. PŘÍLOHY

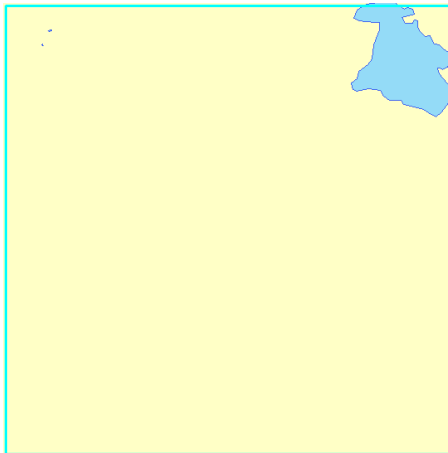
### Příloha 1 Evropské státy v zájmovém území

Albánie	Itálie	Portugalsko
Andorra	Lichtenštejnsko	Rakousko
Belgie	Litva	Rumunsko
Bosna a Hercegovina	Lotyšsko	Řecko
Bulharsko	Lucembursko	San Marino
Černá Hora	Maďarsko	Slovensko
Česká republika	Makedonie	Slovinsko
Dánsko	Malta	Srbsko
Estonsko	Monako	Španělsko
Finsko	Německo	Švédsko
Francie	Nizozemsko	Švýcarsko
Chorvatsko	Norsko	Vatikán
Irsko	Polsko	Velká Británie

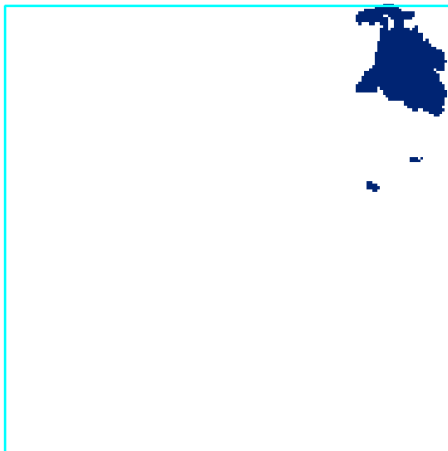
**Příloha 2 Rozdíly jednotlivých databází v jednom modelovém kvadrátu v Evropě**



CORINE Land Cover – 1,12 km<sup>2</sup>

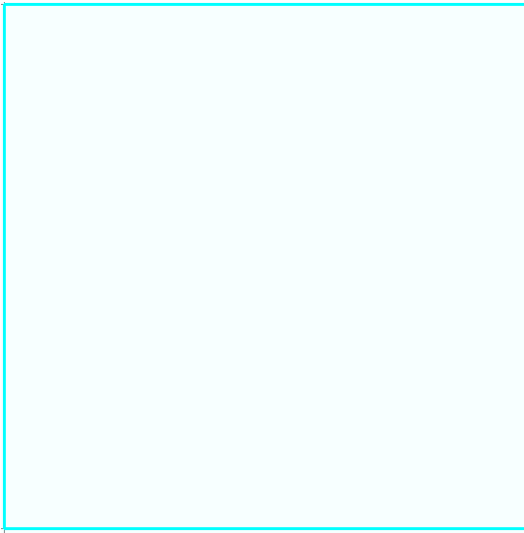


OpenStreetMap – 0,83 km<sup>2</sup>

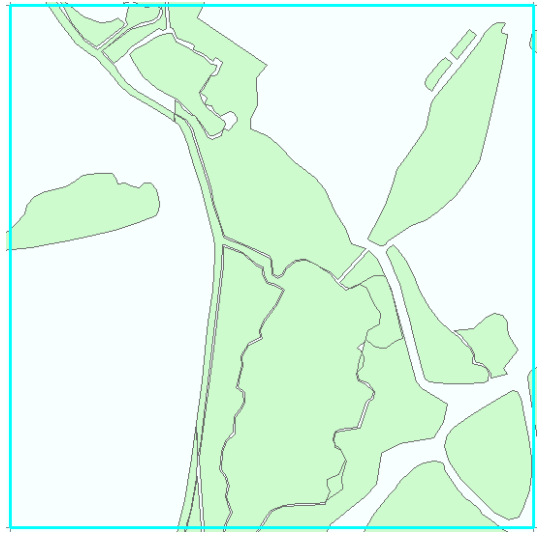


Global Inland Water – 0,74 km<sup>2</sup>

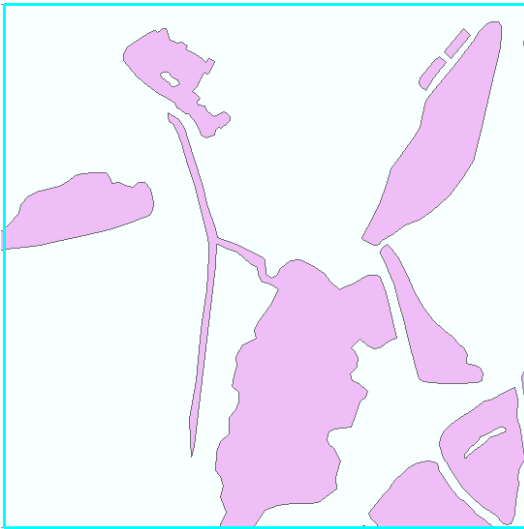
### Příloha 3 Rozdíly jednotlivých databází v jednom modelovém kvadrátu v ČR



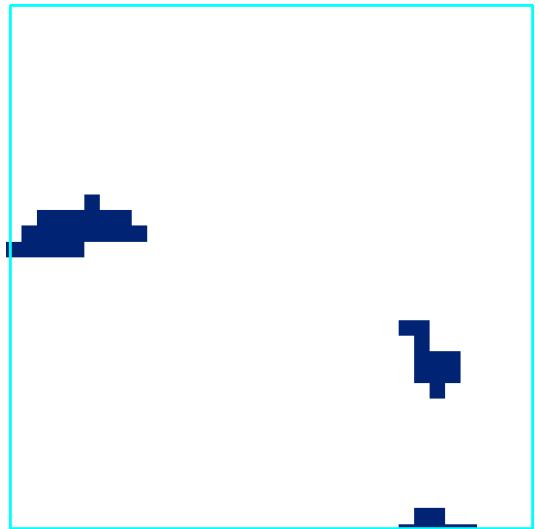
CORINE Land Cover – 0 km<sup>2</sup>



OpenStreetMap – 0,43 km<sup>2</sup>



DIBAVOD – 0,26 km<sup>2</sup>



Global Inland Water – 0,03 km<sup>2</sup>

## Příloha 4 Skript pro tvorbu virtuálního druhu, GLM a výpočet AUC pro Evropu

```
(setwd("~/Dokumenty/R/"))
install.packages("virtualspecies")
install.packages("maps")
install.packages("mapdata")
install.packages("raster")
install.packages("rgdal")
install.packages("PresenceAbsence")

library(PresenceAbsence)
library(virtualspecies)
library(raster)
library(sp)
library(rgdal)
library(maps)
library(mapdata)

eu_Wc<-raster("eu_wc.tif")
png('EU_WorldClim.png')
plot(eu_Wc, main="WorldClim")
dev.off()
eu_Clc<-raster("eu_clc11.tif")
png('EU_Corine.png')
plot(eu_Clc, main="CORINE Land Cover")
dev.off()

eu.stack<-stack(eu_Wc, eu_Clc)

my.par.eu<-formatFunctions("eu_wc" = c(fun = 'dnorm', mean = 7.2, sd =
3), "eu_clc11" = c(fun = 'logisticFun', beta = 15, alpha=-4))

my.sp.eu<-generateSpFromFun(raster.stack = eu.stack[[c("eu_wc",
"eu_clc11")]],parameters = my.par.eu, formula="eu_wc + eu_clc11",plot =
TRUE)
plotResponse(my.sp.eu)

pa_eu<-convertToPA(my.sp.eu, plot=TRUE)
PA.pts_eu<- sampleOccurrences(pa_eu, n=4000, type="presence-absence")
PA.pts_eu
eu_sampled<-PA.pts_eu$sample.points

write.table(eu_sampled, "~/Dokumenty/R/eu_sampled.txt", sep="\t")

eu<-read.csv2("eu.csv",header=TRUE)

eu_giw<-glm(occurrence~SUM_GIW_A+WC_T+I(WC_T^2),family=binomial,data=eu)

nobs<-nrow(eu);nobs
group<-rep(1:5,length=nobs)
group<-sample(group)
D_1<-numeric(0)
D_1_all<-numeric(0)
PA_1_all<-numeric(0)
```

```

for(i in 1:5) {group_1.glm<-
glm(occurrence~SUM_GIW_A+WC_T+I(WC_T^2),family=binomial,data=eu[group!=1
,])
predict_S1<-predict(group_1.glm,newdata=eu[group==i,],type="response")
D_1<-as.vector(predict_S1)
D_1_all<-c(D_1_all,D_1)
PA_1<-eu[group==i,c("square","occurrence")]
PA_1_all<-rbind(PA_1_all,PA_1)
}

PA_1_all<-PA_1_all[,c("square","occurrence")]
eu_1<-data.frame(PA_1_all,D_1_all)
auc(eu_1,st.dev=TRUE,which.model=1,na.rm=FALSE)
optresh<-optimal.thresholds(eu_1,opt.methods="ObsPrev",which.model=1)
t<-optresh[,2];t
cm<-cmx(eu_1,threshold=t,which.model=1,na.rm=TRUE);cmx
pcc(cm,st.dev=TRUE)
presence.absence.accuracy(eu_1,threshold=t,find.auc=TRUE,st.dev=TRUE)

```

**Příloha 5 Tabulka výsledků logistické regrese a AUC pro evropské státy**

	prediktor	parametr	střední chyba odhadu	z value	Pr(> z )	hladina významnosti	AIC	AUC				
<b>eu_clc</b>	(Intercept)	β0	-11,93851	0,53849	-22,171	<2e-16	2146,1	0,9590941				
	SUM_CLC_A	β1	0,43248	0,04504	9,601	<2e-16					sensitivity	0,9279012
	WC_T	β2	4,18408	0,15810	26,465	<2e-16					specificity	0,8698734
	I(WC_T^2)	β3	-0,29376	0,01060	-27,722	<2e-16						
<b>eu_osm</b>	(Intercept)	β0	-10,900279	0,500467	-21,780	<2e-16	2217,9	0,9560508				
	SUM_OSM_A	β1	0,267628	0,032606	8,208	2,25e-16					sensitivity	0,9274074
	WC_T	β2	3,898838	0,146819	26,555	<2e-16					specificity	0,8632911
	I(WC_T^2)	β3	-0,275784	0,009872	-27,936	<2e-16						
<b>eu_giw</b>	(Intercept)	β0	-11,77603	0,53361	-22,07	<2e-16	2161,3	0,958479				
	SUM_GIW_A	β1	0,43756	0,04736	9,24	<2e-16					sensitivity	0,9269136
	WC_T	β2	4,12937	0,15598	26,47	<2e-16					specificity	0,8648101
	I(WC_T^2)	β3	-0,28988	0,01044	-27,77	<2e-16						
<b>eu_clc</b>	(Intercept)	β0	-18,776	6,165	-3,046	0,00232	23,444	0,9997887	sensitivity	1		
	SUM_CLC_A	β1	1,326	0,450	2,946	0,00322					specificity	0,9949686
<b>eu_osm</b>	(Intercept)	β0	-6,75340	0,40896	-16,51	<2e-16	144,01	0,8339522	sensitivity	0,72		
	SUM_OSM_A	β1	0,37155	0,03545	10,48	<2e-16					specificity	0,9605031
<b>eu_giw</b>	(Intercept)	β0	-9,41895	1,20126	-7,841	4,47e-15	49,821	0,9673761	sensitivity	0,96		
	SUM_GIW_A	β1	0,63234	0,09296	6,802	1,03e-11					specificity	0,981635

Vysvětlivky – hladina významnosti: (\*\*\*) 0,001, (\*\*) 0,01, (\*) 0,05, (.) 0,1, ( ) 1



**Příloha 6 Tabulka výsledků logistické regrese a AUC pro Českou republiku**

	prediktor	parametr	střední chyba odhadu	z value	Pr(> z )	hladina významnosti	AIC	AUC			
<b>cz_clc</b>	(Intercept)	$\beta_0$	-29,5101	5,7206	-5,159	2,49e-07	491,98	0,6820035			
	SUM_CLC_A	$\beta_1$	12,9221	6,4485	2,004	0,0451			*	sensitivity	0,8029557
	WC_T	$\beta_2$	8,5156	1,5966	5,334	9,63e-07			***	specificity	0,5177665
	I(WC_T^2)	$\beta_3$	-0,6020	0,1105	-5,448	5,09e-08			***		
<b>cz_dib</b>	(Intercept)	$\beta_0$	-28,5997	5,6100	-5,098	3,43e-07	494,93	0,6825036			
	SUM_DIB_A	$\beta_1$	9,0676	4,4865	2,021	0,0433			*	sensitivity	0,7881773
	WC	$\beta_2$	8,2371	1,5649	5,264	1,41e-07			***	specificity	0,5177665
	I(WC^2)	$\beta_3$	-0,5815	0,1082	-5,373	7,73e-08			***		
<b>cz_osm</b>	(Intercept)	$\beta_0$	-29,1241	5,6424	-5,162	2,45e-07	495,15	0,6826286			
	SUM_OSM_A	$\beta_1$	8,6580	4,1327	2,095	0,0362			*	sensitivity	0,7980286
	WC	$\beta_2$	8,3766	1,5738	5,323	1,02e-07			***	specificity	0,5177665
	I(WC^2)	$\beta_3$	-0,5910	0,1088	-5,430	5,64e-08			***		
<b>cz_giw</b>	(Intercept)	$\beta_0$	-28,8844	5,6548	-5,108	3,26e-07	494,31	0,6873547			
	SUM_GIW_A	$\beta_1$	11,4666	5,4782	2,093	0,0363			*	sensitivity	0,7931034
	WC	$\beta_2$	8,3233	1,5770	5,278	1,31e-07			***	specificity	0,5177665
	I(WC_T^2)	$\beta_3$	-0,6020	0,1105	-5,448	5,09e-08			***		
<b>cz_clc</b>	(Intercept)	$\beta_0$	-2,8905	0,2257	-12,805	<2e-16	174,3	0,5291547	sensitivity	0,1923077	
	SUM_CLC_A	$\beta_1$	41,0881	12,1174	3,391	0,000697			***	specificity	0,0,986631
<b>cz_dib</b>	(Intercept)	$\beta_0$	-2,902	0,227	-12,783	<2e-16	183,11	0,5528075	sensitivity	0,2692308	
	SUM_DIB_A	$\beta_1$	19,332	5,744	3,366	0,000763			***	specificity	0,8957219
<b>cz_osm</b>	(Intercept)	$\beta_0$	-2,8850	0,2271	-12,705	<2e-16	187,46	0,5544015	sensitivity	0,2692308	
	SUM_OSM_A	$\beta_1$	11,5070	3,6176	3,181	0,00147			**	specificity	0,842246
<b>cz_giw</b>	(Intercept)	$\beta_0$	-2,8312	0,2191	-12,920	<2e-16	186,38	0,5364562	sensitivity	0,2307692	
	SUM_GIW_A	$\beta_1$	16,0883	5,1493	3,124	0,00178			**	specificity	0,9171123

Vysvětlivky – hladina významnosti: (\*\*\*) 0,001, (\*\*) 0,01, (\*) 0,05, (.) 0,1, ( ) 1