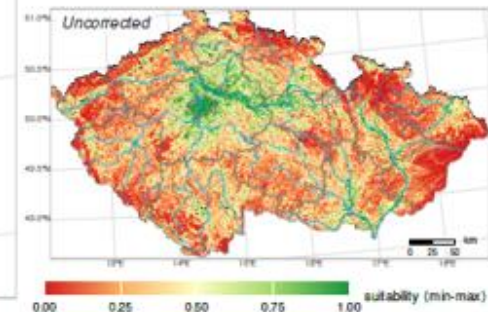
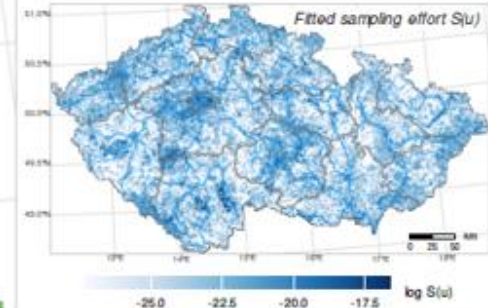
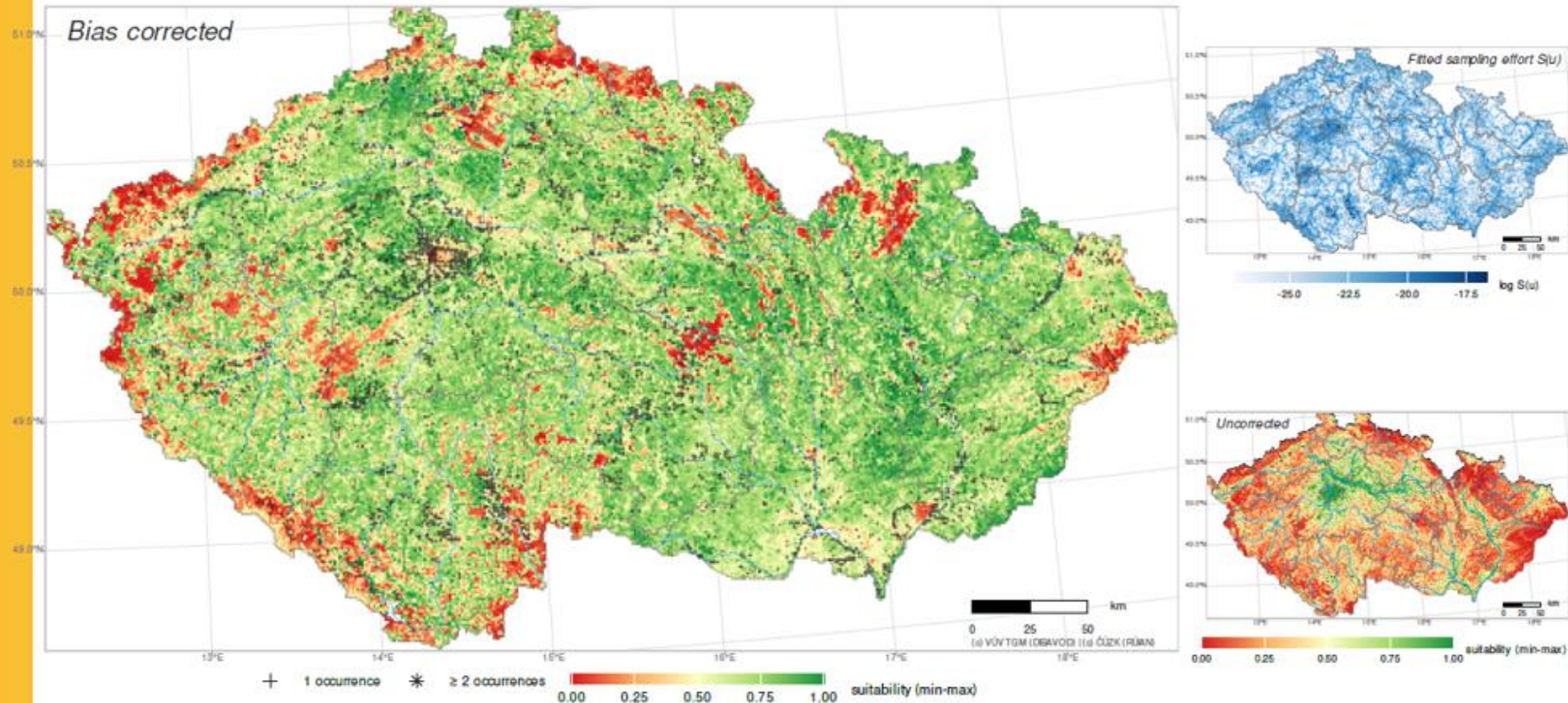


Komplexní modelování biodiverzity na úrovni ČR

Petra Šímová, Petr Balej, Dominika Prajzlerová, Jiří Prošek, Richard Bittman,... Vítězslav Moudrý

Ardea cinerea (volavka popelavá) | Ptáci



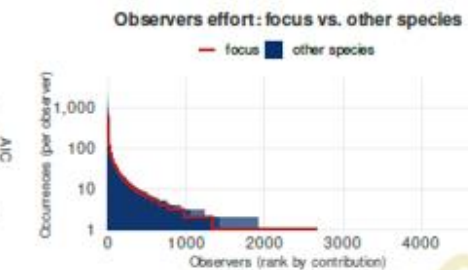
Period: 2021-2025 (months 3-8)

Occurrence data source: NDOP

Occurrence filtering:
84 129 raw -> 21 489 dedup -> 5 833 per pixel

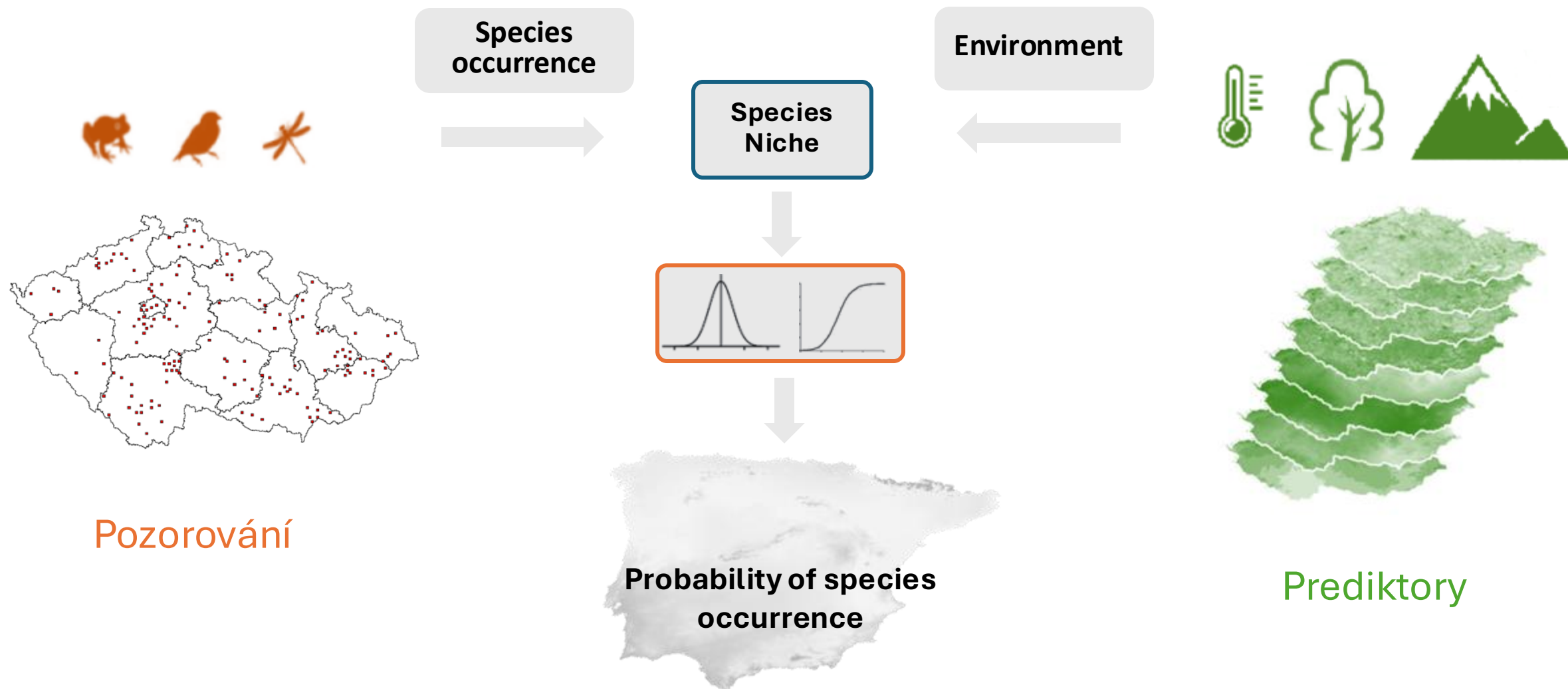
Unique observers: 2663

Prediction similarity (corrected vs. uncorr.):
Pearson's $r = 0.071$, Schoener's $D = 0.684$

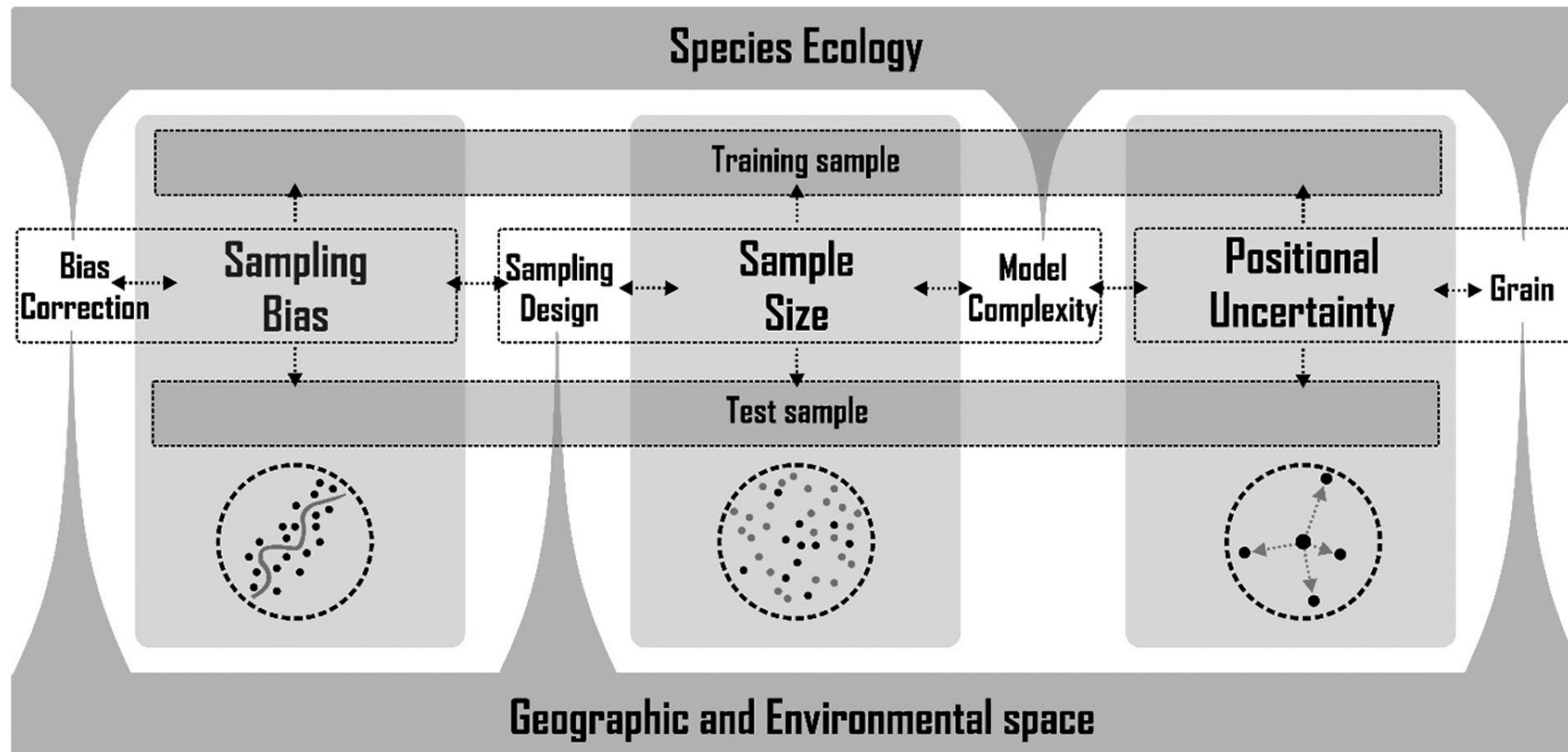


Species occurrence data source: Název a databáze ochrany přírody (NDOP), Agentura ochrany přírody a krajiny ČR [S602030018 Centrum pro krajiny a biodiverzitu | Observer-Oriented Multi-Species Point-Process Model (spatialstat-omppm)]
Prediction transformation: cloglog suitability [line is rescaled for display] Continuous Boyce Index (CBI: spatial-block [n=3, size: 50 km] cross-validation (CV), Thin dist=10 km, 387 presences left, Akaike information criterion (AIC).

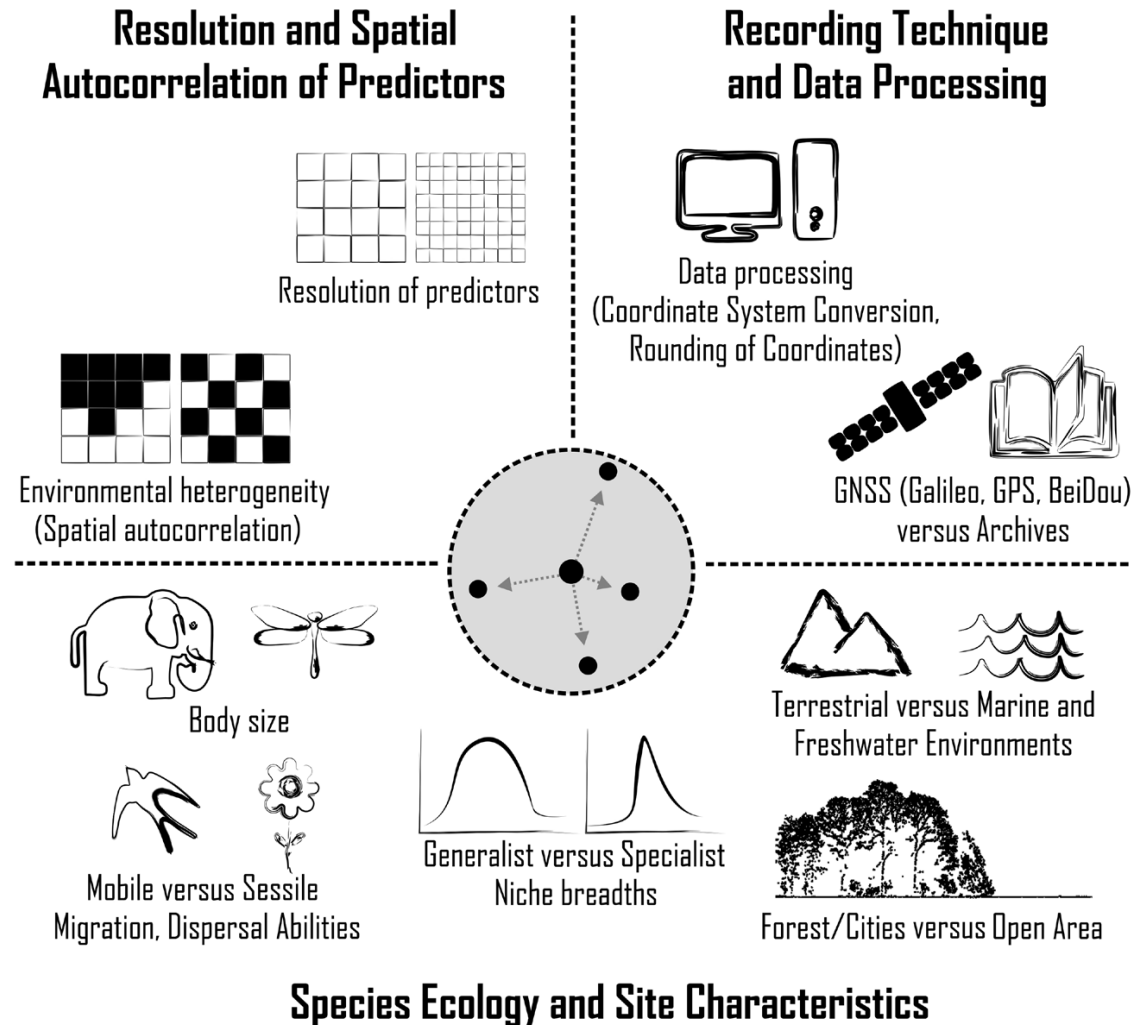
Modelování druhové distribuce



Vstupní data (**Pozorování**)

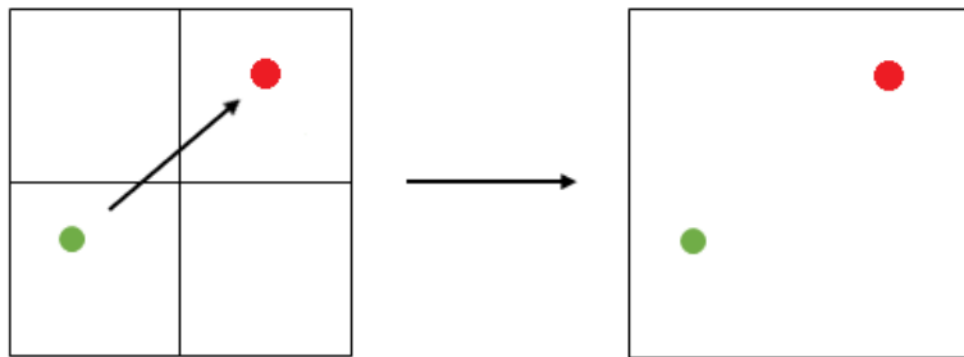


Vstupní data (Polohová chyba)



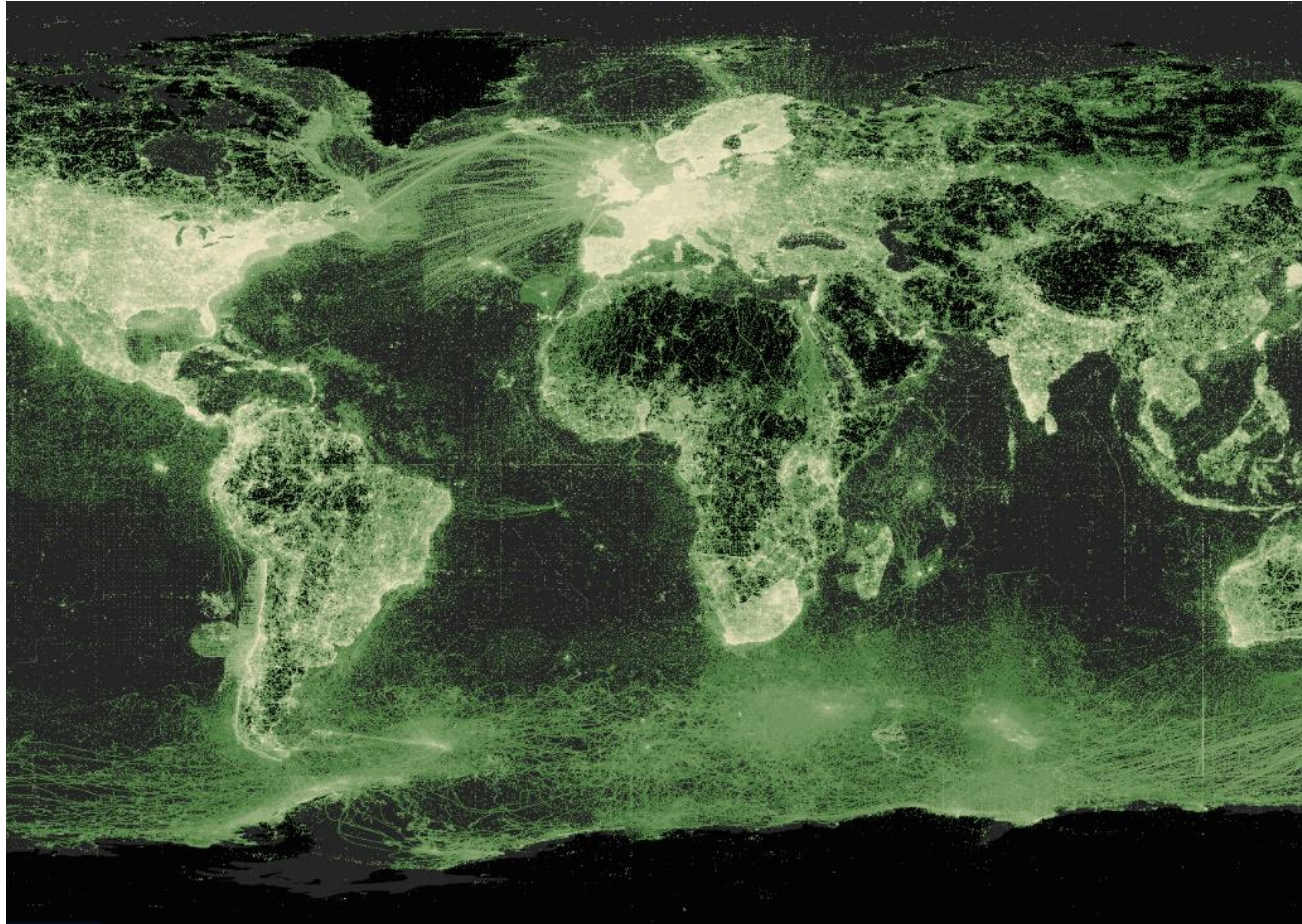
Vstupní data (Polohová neurčitost vs měřítko)

Zhoršení rozlišení vstupních dat (prediktorů) účelem kompenzace poziční chyby nevede ke zlepšení výkonnosti modelu.



Gábor, L., Jetz, W., Lu, M., Rocchini, D., Cord, A., Malavasi, M., ... & Moudrý, V. (2022). Positional errors in species distribution modelling are not overcome by the coarser grains of analysis. *Methods in Ecology and Evolution*, 13(10), 2289-2302.

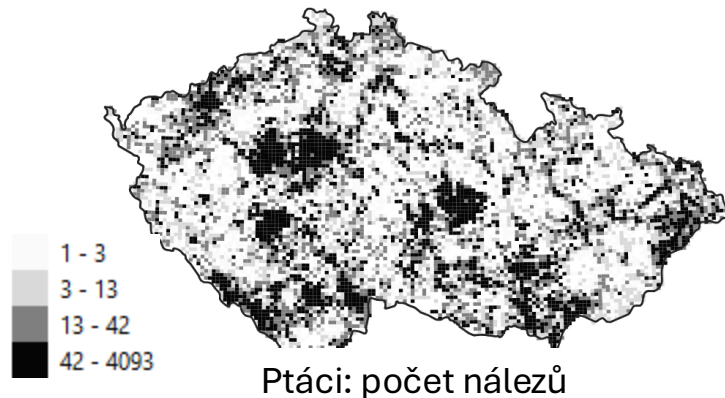
Vstupní data (**Sampling bias**)



Paradox “záplavy” dat

- Odborníci i občanská věda nasbírali **miliony** (NDOP) až **miliardy** záznamů druhů (GBIF, iNaturalist, eBird).
- **Více dat ≠ lepší modely.** Záznamy odrážejí *lidské chování* stejně jako biologii.
- Pozorovatelé se koncentrují u silnic, měst a atraktivních lokalit.
- Když to ignorujeme, model se učí **geografii, ne biologii.**
- Problémem není nedostatek dat — ale jak nedokonalá data spolehlivě využít.

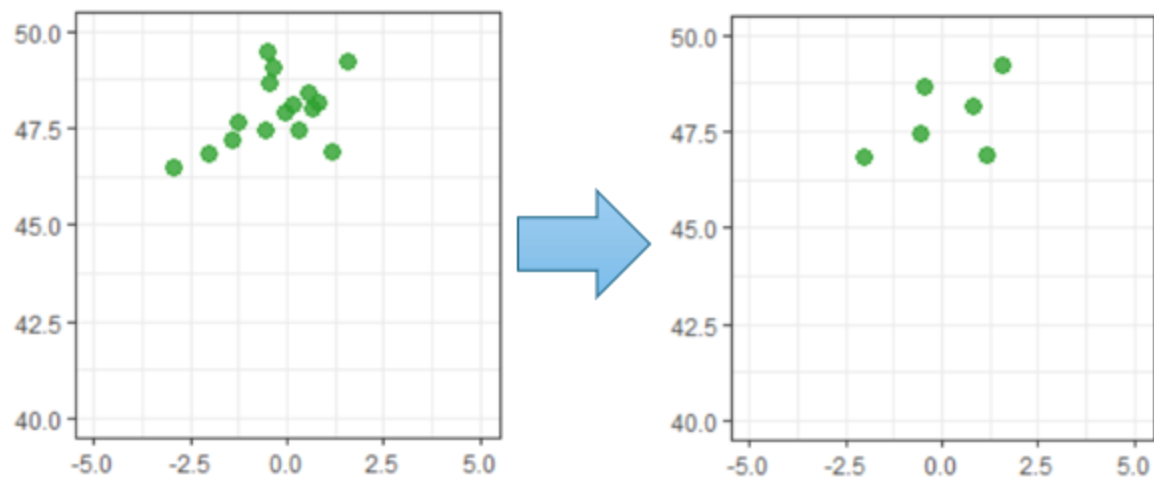
Vstupní data (**Sampling bias**)



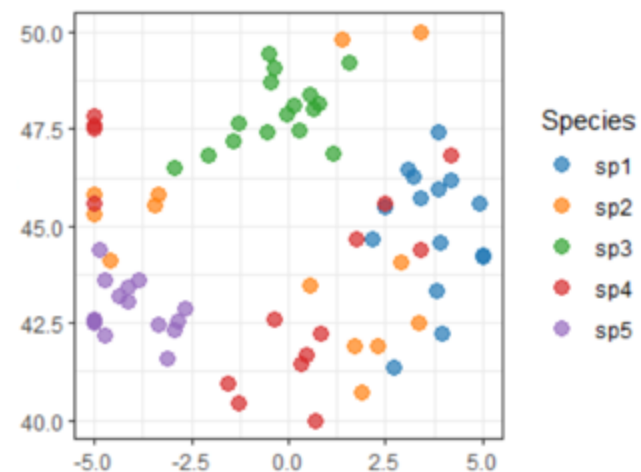
- Data typu **přítomnost (presence)**: víme, kde byl druh *spatřen*, ne kde *chybí*.
- Nerovnoměrné **vzorkovací úsilí** činí některé oblasti “lepšími” jen proto, že tam hledalo více lidí.
- Nejsilnějším a nejpřehlíženějším faktorem je **heterogenita pozorovatelů**: různí lidé chodí na různá místa a zaznamenávají různé taxony.
- *NDOP je nejobsáhlejší česká národní databáze — ale jako všechna oportunistická data je “(ne)přirozeně” zkreslená nerovnoměrnou intenzitou sběru druhových nálezů v různých oblastech a prostředích.*

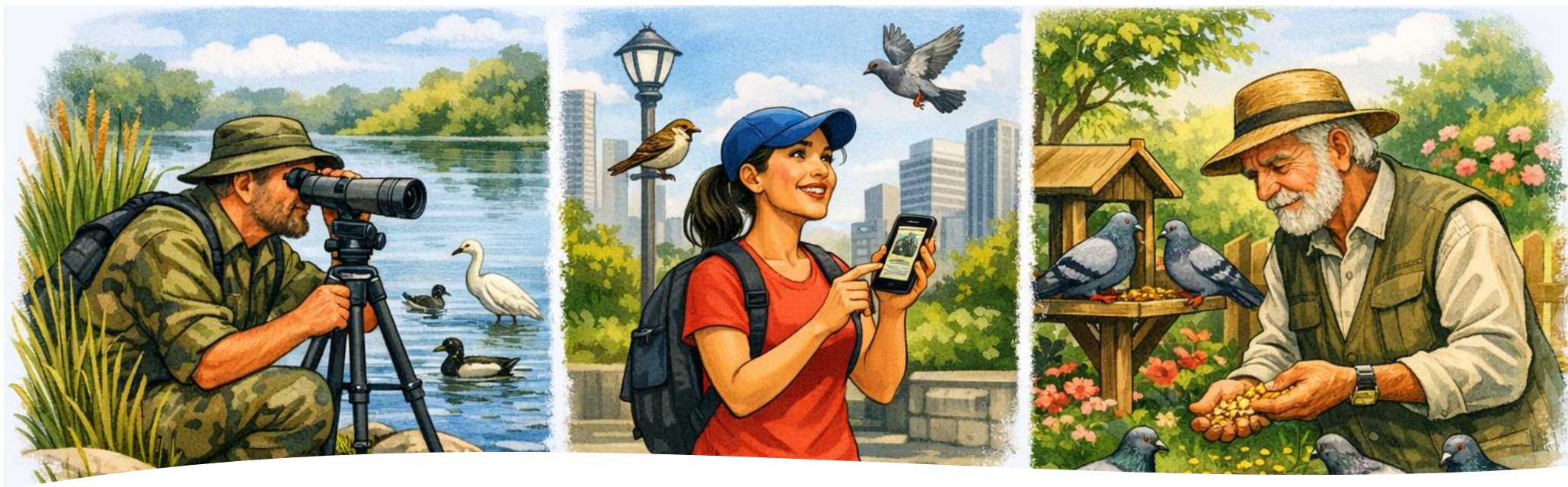
Metoda	Myšlenka	Omezení
Prostorové proředění (např. spThin)	Smazat shluky bodů	Nelze rozlišit co je důvodem shluku
Manipulace prostředí cílové skupiny (TGOB)	Použít všechny záznamy skupiny jako “úsilí”	Zachází se všemi pozorovateli stejně

Spatial Thinning of Species Presences (STSP)



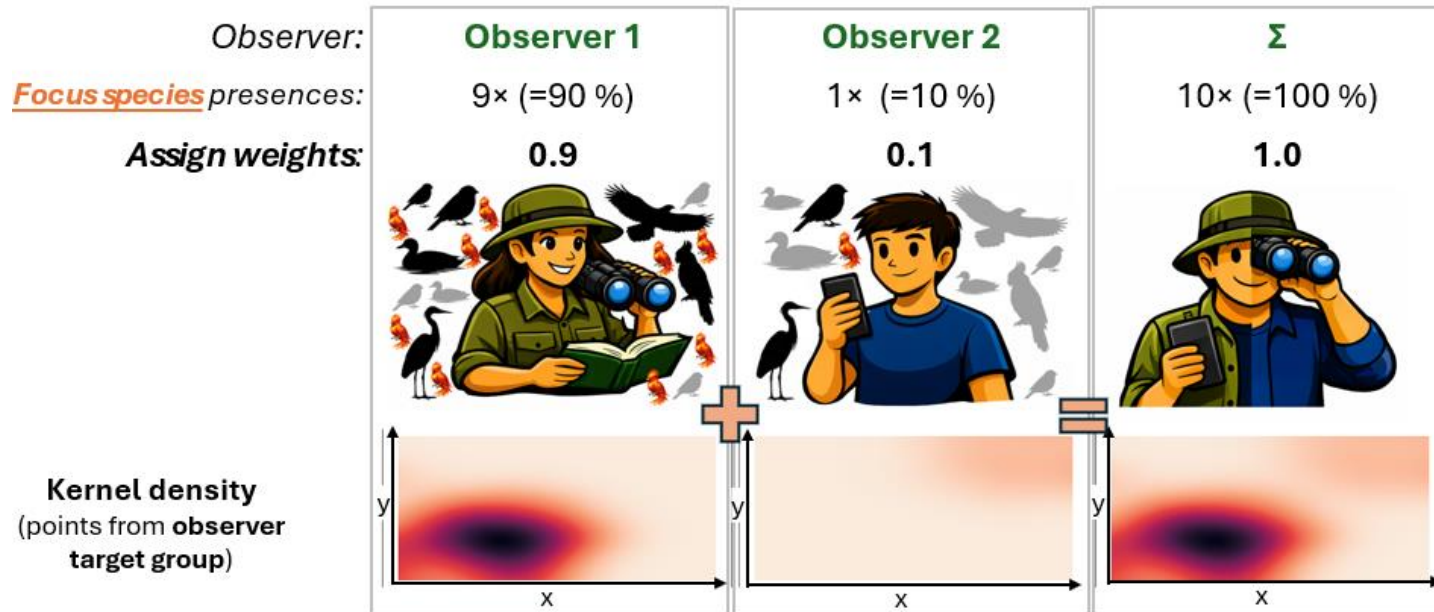
Target Group Occurrences Background (TGOB)





Pozorovatelé nejsou stejní

- Expert systematicky navštěvující mokřady $\neq$ příležitostný pozorovatel na zahradě.
- Každý pozorovatel má specifický **prostorový otisk** (kde hledá) a **příspěvek** (kolik přidává k danému druhu).
- Modelovat *kdo, kde a co hledal* — ne jen odkud máme k dispozici nálezy druhu.



- **Prostorový otisk pozorovatele** — hladký povrch jádrové hustoty (*Kernel density*) toho, kde pozorovatel zaznamenává.
- **Váha pozorovatele (OW)** — podíl pozorovatele na obsazených pixelech cílového druhu.
- **Povrch zkreslení = Σ (otisk $\times$ váha)** přes všechny pozorovatele.

- Metoda: **PW-OOA** — presence-weighted observer-oriented approach
- 108 druhů českých ptáků, testováno proti **nezávislému** terénnímu průzkumu (LSD).
- PW-OOA překonala většinou ostatní metody korekce.

Vstupní data (**Prediktory**)

Zdroj:

- Data z pozemního měření
- (Satelitní) dálkový průzkum Země
- Modely (predikce)
- ...

Typ prediktorů:

- Klima
- Terén
- Krajinový pokryv
- Povrchová odrazivost
- Spektrální indexy
- Struktura vegetace
- ...

Vstupní data (Prediktory)

(BIO)KLIMA

- BIO02 – Mean Diurnal Range = průměrný denní teplotní rozsah, tedy průměrný rozdíl mezi denní maximální a minimální teplotou.

- BIO13 – Precipitation of Wettest Month = úhrn srážek v nejvlhčím měsíci.

- BIO14 – Precipitation of Driest Month = úhrn srážek v nejsušším měsíci.

NADMOŘSKÁ VÝŠKA

SVAŽITOST

POVRCHOVA ODRAZIVOST (SENTINEL-2)

Období	Pásmo	Název	Přibližná centrální vlnová délka	Co zachycuje
3–4	B3	Green	560 nm	zelená vegetace, pigmentace
	B6	Red Edge 2	740 nm	vegetační stav, chlorofyl, struktura
	B12	SWIR 2	2190 nm	obsah vody, struktura vegetace,
5–6	B4	Red	665 nm	chlorofyl, absorpce červeného záření
	B8	NIR	842 nm	biomasa, struktura a vitalita vegetace
	B11	SWIR 1	1610 nm	obsah vody ve vegetaci/půdě
7–8	B7	Red Edge 3	783 nm	chlorofyl, struktura a fyziologický stav vegetace
9–10	B2	Blue	490 nm	pigmentace, vegetace, půdní/atmosférické charakteristiky
	B8	NIR	842 nm	biomasa, struktura a vitalita vegetace



Classified data:

-  time-consuming
-  accuracy only 65 – 80%
-  limited resolution

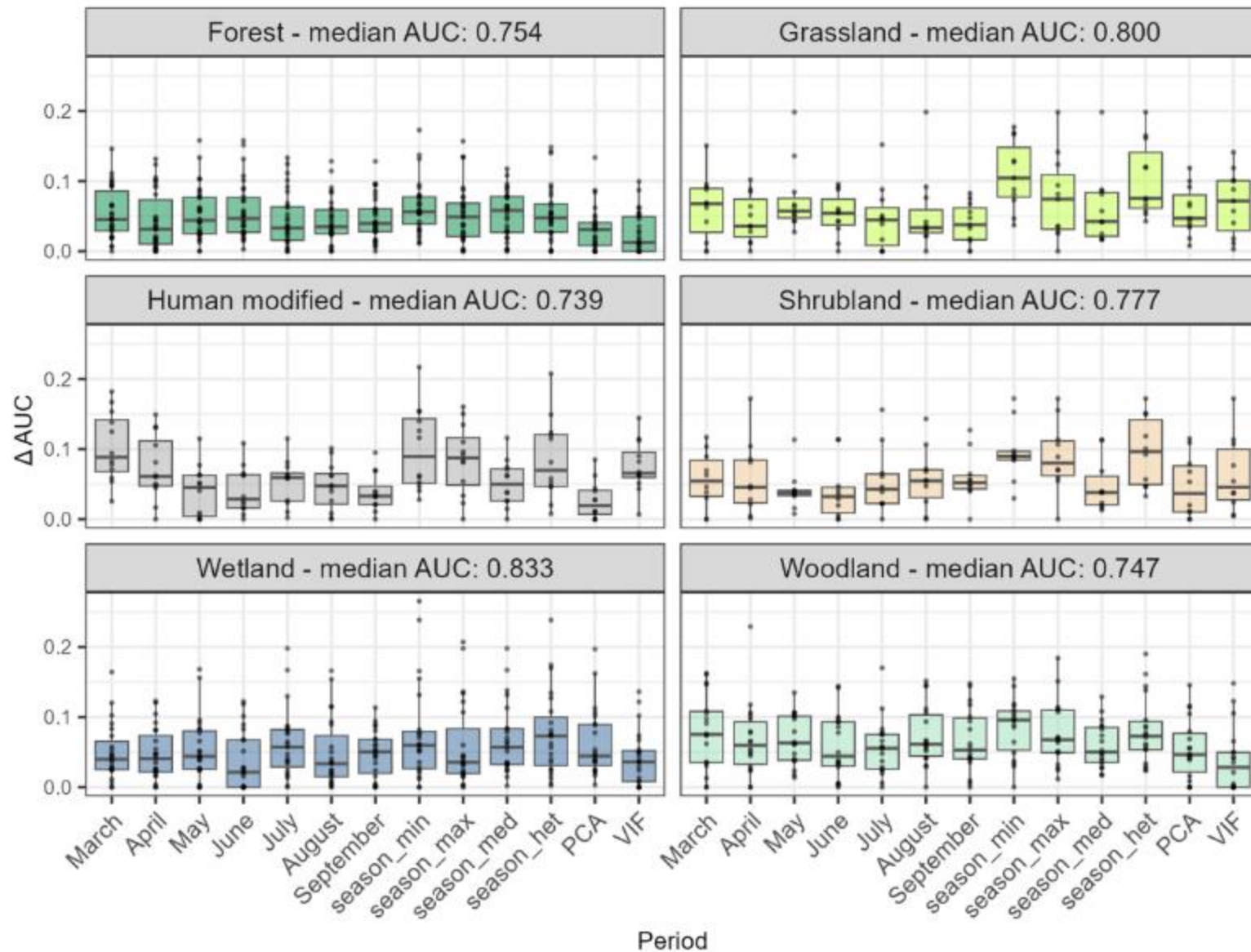


Unclassified data:

-  available
-  regular, large areas
-  original information for each pixel

Výsledky ukazují, že spektrální heterogenita odvozená z neklasifikovaných multispektrálních dat je účinným nástrojem pro hodnocení diverzity.

Prajzlerová, Dominika, et al. "The relationship between remotely-sensed spectral heterogeneity and bird diversity is modulated by landscape type." *International Journal of Applied Earth Observation and Geoinformation* 128 (2024): 103763.

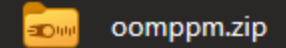


Boxplots illustrating the distribution of ΔAUC values for different models across various habitat types. The x-axis represents different periods, while the y-axis shows ΔAUC , representing the difference between the AUC for a given period and the best-performing period for each species (birds).

Výsledek

- *R* balíček **oomppm** :

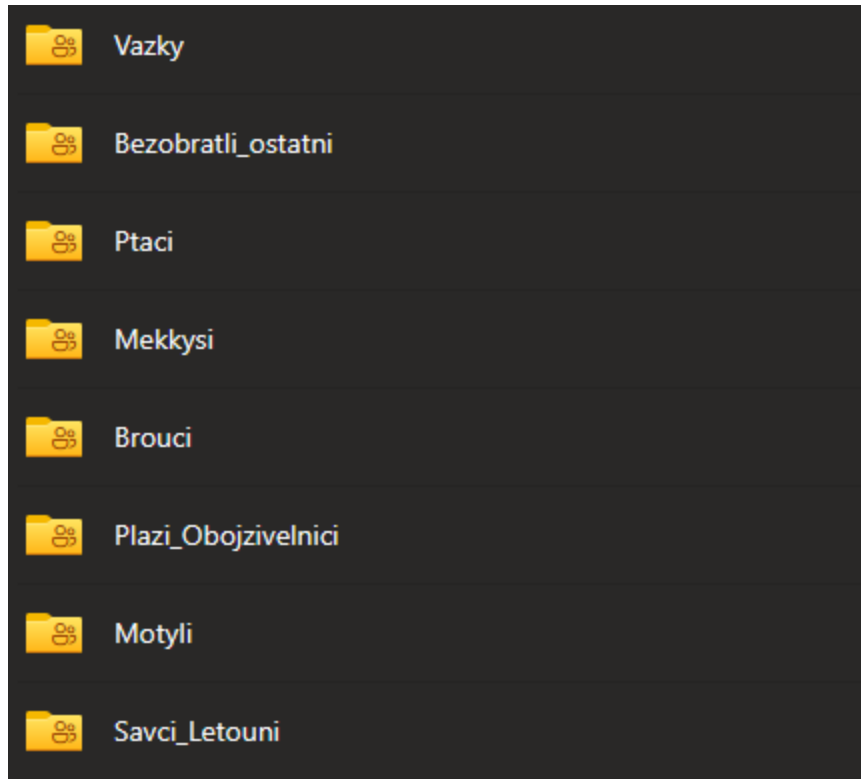
- prediktory → nálezy → **úsilí pozorovatelů** → model → validace → mapy.



- Model současně odhaduje **stanoviště** (klima, terén, vegetace Sentinel-2) a **vzorkovací úsilí** (otisky jednotlivých pozorovatelů).
- Při predikci se **úsilí vypne** → čistá, nezkreslená mapa vhodnosti stanoviště (= *habitat suitability*).
- Bonus: vydá i odhadnutý **povrch vzorkovacího úsilí**.
- Testováno na běžných PC (Windows 11) a finální modely na HPC Metacentrum (Linux).
- Nativně víceprocesově (bez potřeby dalších pomocných knihoven).

Výsledek

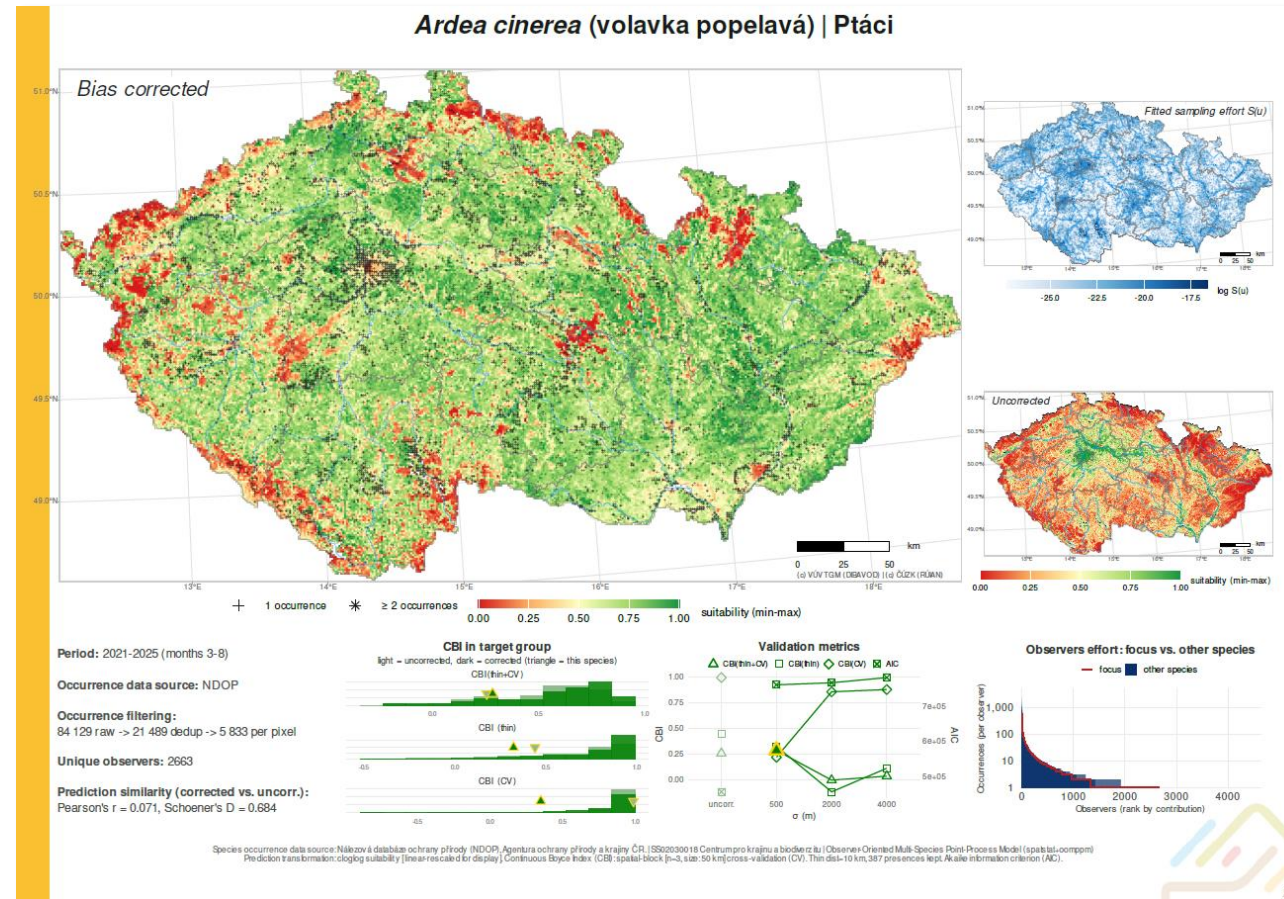
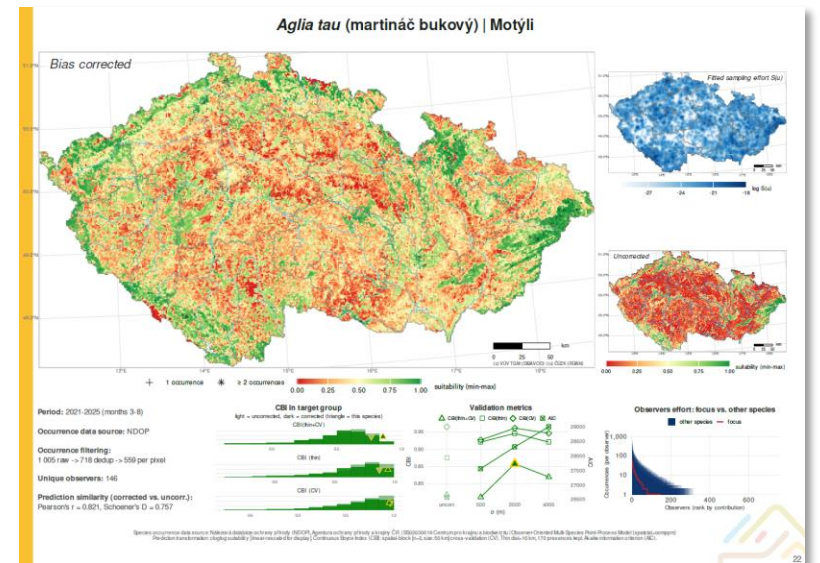
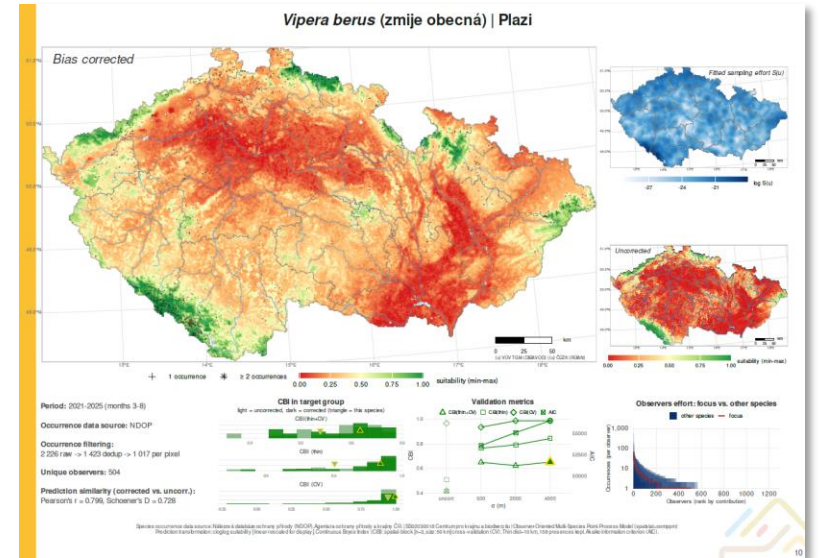
- Aplikováno na celý NDOP



- **Mapu vhodnosti korigovanou o vzorkovací zkreslení** s překrytím existujících nálezů.
- **Odhadnutý povrch vzorkovacího úsilí** (kde lidé skutečně hledali).
- **Nekorigovanou mapu** vedle — přesně vidíte, co korekce změnila.
- **Metriky kvality** (křížově validovaný Boyceho index) a statistiky pozorovatelů.
- *Vše v automaticky generovaném PDF reportu.*

Co získáte pro každý druh

- PDF reporty, ukázky:

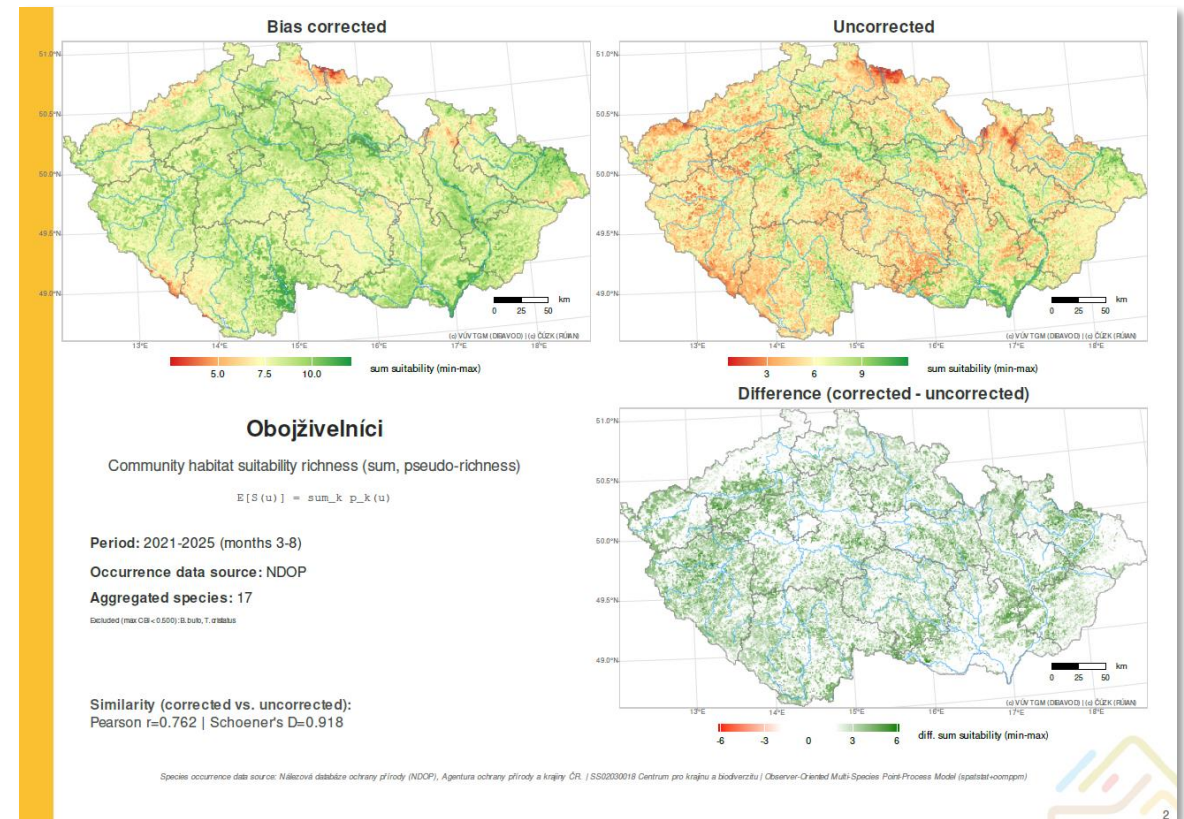
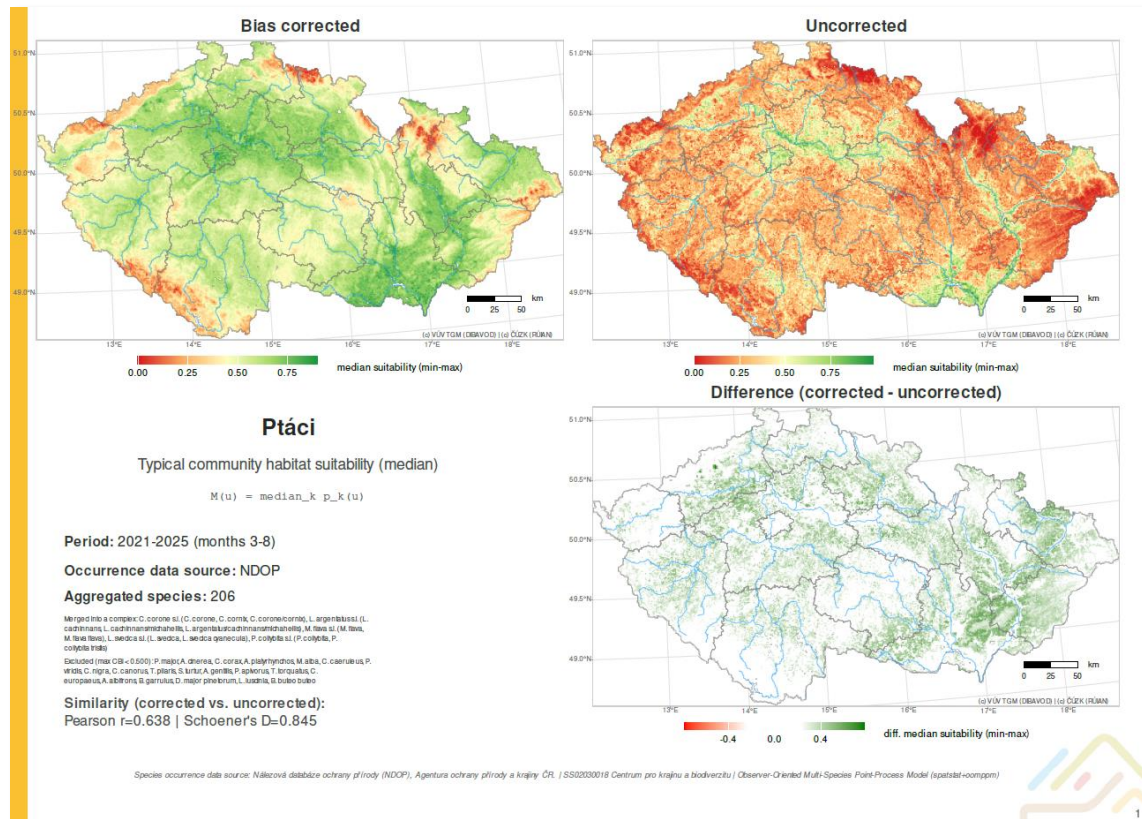


Co získáte pro celé společenstvo (skupinu)

- Mapa **komunitní bohatosti**: sečtené korigované vhodnosti (pseudo-bohatost).
- Mapa **typické vhodnosti**: medián přes druhy.
- Korigované vs nekorigované, plus jejich **rozdíl** — zviditelněný otisk zkreslení.

Co získáte pro celé společenstvo (skupinu)

- PDF reporty, ukázky:



Proč to používat

- **Mění zkreslené záznamy NDOP v lépe obhajitelné mapy** výskytu pro ochranu a reporting.
 - Modelujme, kde druhy žijí — ne kde jsme “náhodou” našli.
- **Transparentní:** ukazuje korekci a její efekt vůči nekorigovanému modelu.
- **Reprodukovatelné a škálovatelné:** jeden skript, mnoho taxonů, připraveno pro HPC.

Paradox hodnocení (validace)

- Testování na **stejných zkreslených datech** (interní validace) může upřednostnit *nekorigovaný* model — “odměňuje” predikci zkreslení.
- Pravdu odhalí jen **nezávislá** nebo **prostorově oddělená** validace.
- Poučení vestavěné do *oomppm*: validujeme **prostorově-blokovou křížovou validací** a **Continuous Boyce Indexem**, ne naivními interními metrikami.

