

LATVIJAS UNIVERSITĀTE
MEDICĪNAS UN DZĪVĪBAS ZINĀTŅU FAKULTĀTE
EKOLOĢIJAS NODAĻA

SUGU IZPLATĪBAS MODEĻA KVALITĀTES METRIKAS

Vita Šakele

2025

SATURA RĀDĪTĀJS

IEVADS	2
1. METRIKAS	2
1.1. Pārpratumu matrica	4
1.2. AUC	5
1.3. TSS	6
1.4. Koena kappa	6
1.5. Nepārtrauktais Boisa indekss (CBI)	7
2. TIKAI KLĀTBŪTNES DATU IZMANTOŠANAS RADĪTĀS PROBLĒMAS	8
3. METRIKU PIELIETOJUMS DAŽĀDĀM KOPĀM	9
4. IZMANTOTĀ LITERATŪRA	12

IEVADS

Neatkarīgi no modelējamās problēmas un lietotajiem modelēšanas algoritmiem neatņemama modelēšanas sastāvdaļa ir modeļu hiperparametru pielāgošana un izveidoto modeļu kvalitātes novērtēšana un salīdzināšana. Šajā referātā ir apkopots, kādas metrikas tiek izmantotas sugu izplatības modeļu izvērtēšanai, ko tās nozīmē un kādi ir to ierobežojumi tikai klātbūtnes modeļos, kuras metrikas rēķina katrai no modelēšanas kopām – apmācības, validācijas un testa.

1. METRIKAS

Sugu izplatības modelēšanā ir izmantojamas vairākas metrikas. Ir pētījumi (Konowalik & Nosol, 2021), kur vispirms izmantotas praktiski visas attiecināmās metrikas, tad meklēta Spīrmēna rangu korelācija starp patieso vērtību (*ground-truth*) novērtējumiem un metrikām, kā arī beigās izmantojamās metriku kopas noteikšanai pielietota lineārā regresija. Minētajā pētījumā sākotnēji izmantoja šādas metrikas:

- laukums zem uztvērēja darba raksturlīknes (*Area under the ROC curve, AUC*),
- vidējā kvadrātiskā kļūda (*Root Mean Squared Error, RMSE*),
- pareizība (*Accuracy*),
- nobīde (*Bias*),
- vidējā absolūtā kļūda (*Mean Absolute Error, MAE*),
- kvadrātisko kļūdu summa (*Sum of Squared Errors, SSE*),
- sagaidāmā kalibrācijas kļūda (*Expected Calibration Error, ECE*),
- maksimālā kalibrācijas kļūda (*Maximum Calibration Error, MCE*),
- nesvērtā Kappa statistika (*unweighted Kappa statistic*),
- jutība jeb pārklājums (*Sensitivity or Recall*),
- specifiskums (*Specificity*),
- pozitīvā prognozējošā vērtība (*Positive Predictive Value, PPV*),
- negatīvā prognozējošā vērtība (*Negative Predictive Value, NPV*),
- precizitāte (*Precision*),
- F1 rādītājs,
- sastopamība (*Prevalence*),
- noteikšanas biežums (*Detection Rate*),
- balansētās pareizības Boisa indekss (*Balanced Accuracy Boyce Index*),
- patiesās spējas statistika (*True Skill Statistic, TSS*),

- Žakāra līdzības indekss (*Jaccard's Similarity Index*),
- Sērensena līdzības indekss (*Sørensen's Similarity Index*),
- pārprognozēšanas biežums (*Overprediction Rate, OPR*),
- zemprognozēšanas biežums (*Underprediction Rate, UPR*).

Iepriekšminētajā pētījumā (Konowalik & Nosol, 2021) visaugstāko saskaņu ar patieso vērtību uzrādīja AUC, TSS, MAE, Bias un ECE. Citi autori par biežāk izmantotajām uzskata AUC, TSS un Koena kappa (Bracho-Estévez u.c., 2024), (Leroy u.c., 2018), (McPherson u.c., 2004), (Allouche u.c., 2006), (Grimmett u.c., 2020) un sāk parādīties pētījumi, kur izmanto arī Boisa indeksu (Avotiņš, 2025), (Burdon et al., 2025), (Liu u.c., 2025). Tomēr tādas absolūtas vienprātības, kuras metrikas būtu jāizmanto, joprojām nav un tiek izmantotas vairākas metrikas kopā (Avotiņš, 2025). Literatūrā atrodamas diskusijas par metriku trūkumiem. Dažas no tām:

- AUC un TSS rezultāti ir augstāki datu kopās ar vairāk punktiem (Konowalik & Nosol, 2021).
- Sākotnēji uzskatīja, ka TSS nav atkarīgs no sastopamības biežuma (*prevalence*) (Allouche u.c., 2006). Tomēr ir pierādījumi, ka ir TSS neatkarība no sastopamības biežuma ir saistīta ar datu kopas izmēru (Somodi u.c., 2017).
- Līdzīgi bija ar AUC un to arī uzskatīja par neatkarīgu no sastopamības biežuma (Manel u.c., 2001), (McPherson u.c., 2004).
- AUC kļūdas var būt lielas, ja paraugi ir mazi (Valavi u.c., 2022), arī Koena kappa ir atkarīga no sastopamības (Allouche u.c., 2006). AUC korelē ar Koena kappa rādītāju, kurš ir vienkāršāks aprēķinos (Manel et al., 2001).
- AUC vienādi novērtē I tipa kļūdas (viltus pozitīvie jeb liekās klātbūtnes) un II tipa kļūdas (viltus negatīvie jeb trūkstošās klātbūtnes) (Lobo u.c., 2008).
- AUC ir patiesi informatīvs tikai tad, ja ir pieejami patiesie iztrūkuma dati un pētījuma mērķis ir realizētās izplatības novērtēšana. Ja pētījuma mērķis ir potenciālā izplatība, AUC nav atbilstošs veikspējas rādītājs, jo I tipa kļūdas pēc būtības ir ar mazāku svaru nekā II tipa kļūdas (Jiménez-Valverde, 2012).
- AUC nesniedz informāciju par modeļa kļūdu telpisko sadalījumu (Lobo u.c., 2008).
- Jālieto metrikas, kas novērtē saskaņotību starp atkārtotām prognozēm (mēra telpisko stabilitāti), nevis tikai veikspēju, kā to dara AUC (Grimmett u.c., 2020).
- Ir ieteikums TSS aizstāt ar meteoroloģijā izmantoto simetrisko ekstrēmās atkarības indeksu (*Symmetric Extremal Dependence Index, SEDI*), jo, novērtējot retu sugu klātbūtni, TSS ātri konverģē uz jutības metriku (Wunderlich u.c., 2019).

Nākamajās apakšnodaļās ir aprakstīts AUC, TSS, Koena kappa un Boisa indekss, kas, neskatoties uz to trūkumiem, ir būtiskas sugu izplatības modeļu novērtēšanas metrikas. Tā kā metrikas balstās uz pārpratumu matricu, tad vispirms dots tās apraksts un tikai pēc tam pašas metrikas.

1.1. Pārpratumu matrica

Pārpratumu matrica (confusion matrix) (1. attēls) ir 2 x 2 matrica, kas apkopo datus par to, cik bieži modelis pareizi vai kļūdaini prognozē katru klasi, salīdzinot prognozes ar faktiskajiem datiem. Sugu izplatības modelēšanas kontekstā klases ir klātbūtne un iztrūkums. Attiecīgi prognozētās klases atbilst prognozētajai sugas klātbūtnei vai iztrūkumam un aktuālās klases – faktiski novērotajai sugas klātbūtnei vai iztrūkumam. Tikai klātbūtnes modeļos faktiski novērotais iztrūkums ir aizstāts ar pseidoiztrūkuma vai fona datiem (Phillips u.c., 2009), bet no pārpratumu matricas viedokļa, tas tomēr ir pārkāpums, jo tur būtu nepieciešami faktiskā iztrūkuma dati (Li & Guo, 2013). Tradicionāli pārpratumu matricu veido šādi elementi (Li & Guo, 2013):

- patiesi pozitīvais (True Positive, TP) – pareizi prognozēto klātbūtnes gadījumu skaits,
- viltus pozitīvais (False Positive, FP) – iztrūkumu skaits, kas ir prognozēti kā klātbūtnes; šo sauc arī par I tipa kļūdu jeb komisijas kļūdu (*Commission Error*) – kļūdaini prognozētā klātbūtne,
- viltus negatīvais (False Negative, FN) – faktisko klātbūtnes novērojumu skaits, kas ir prognozēti kā iztrūkums; šo sauc arī par II tipa kļūdu jeb (*Omission Error*) – kļūdaini prognozētais iztrūkums,
- patiesi negatīvais (True Negative, TN) – pareizi prognozētais iztrūkumu gadījumu skaits.

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

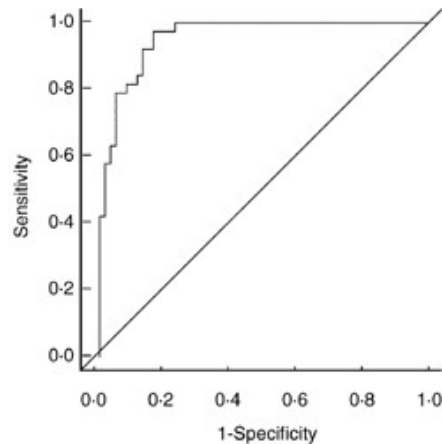
1. attēls. Pārpratumu matrica un metrikas (<https://medium.com/swlh/how-to-remember-all-these-classification-concepts-forever-761c065be33>).

1.2. AUC

Klasiskā variantā AUC skatās ROC līknei un var apzīmēt kā AUC_{ROC} , kuru aprēķina ņemot vērā *1-specifiskums* jeb viltus pozitīvo līmenis attiecībā pret *jūtīgumu* jeb pareizi prognozēto klātbūtni (patiesi pozitīvo) līmeni pie daudziem sliekšņiem (Valavi u.c., 2022):

$$\text{specifiskums} = \frac{TN}{TN+FP} \quad \text{un} \quad \text{jūtīgums} = \frac{TP}{TP+FN}$$

Uz x ass attēlo *1- specifiskums*, bet uz y ass – *jūtīgums* (2. att.).

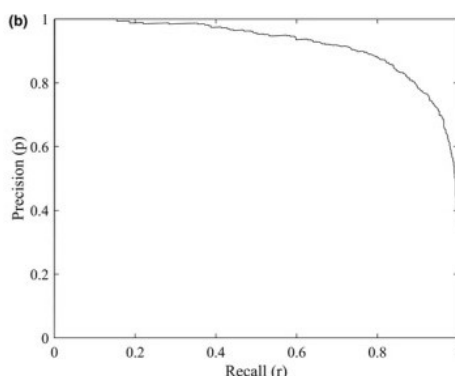


2. attēls. AUC_{ROC} piemērs (Manel u.c., 2001).

AUC_{ROC} vērtība ir diapazonā no 0 līdz 1, kur 1 norāda uz perfektu diskrimināciju, bet 0.5 – uz diskrimināciju, kas nav labāka par nejaušu klasifikāciju (Valavi u.c., 2022). Ar diskrimināciju sugu izplatības modelēšanā saprot modeļa spēju pareizi nošķirt klātbūtni no iztrūkuma (Leroy u.c., 2018).

Tomēr iespējama arī alternatīva AUC, kad skatās nevis ROC līkni, bet gan precizitātes – pārklājuma līkni (*Area under Precision-Recall Curve*, AUC_{PR}), tādā veidā pielāgojot AUC tikai klātbūtnes modeļiem, lai fokusētos tieši uz klātbūtnes prognozēšanu (Valavi u.c., 2022). AUC_{PR} (3. att.) ir ieteikts izmantot, kad negatīvo (iztrūkuma) vērtību ir daudz vairāk nekā pozitīvo (klātbūtnes) jeb datu kopa ir nesabalansēta un šī situācija atbilst tikai klātbūtnes modeļiem un retu sugu modelēšanai sugām, t.i., sugām ar zemu sastopamību. Līdzīgi kā AUC_{ROC} , AUC_{PR} arī aprēķina vienu veikspējas mēru visā iespējamo sliekšņu vērtību diapazonā. AUC_{PR} gadījumā uz x ass ir *jūtīgums* jeb *pārklājums*, kas parāda, cik liels ir to patieso klātbūtnu īpatsvars, kuras modelis ir spējis pareizi atklāt. Uz y ass ir *precizitāte*, kas parāda, cik liels ir patiesās klātbūtnes īpatsvars visās modeļa prognozētajās klātbūtnēs.

$$\text{jūtīgums jeb pārklājums} = \frac{TP}{TP+FN} \quad \text{un} \quad \text{precizitāte} = \frac{TP}{TP+FP}$$



3. attēls. AUC_{PR} piemērs (Li & Guo, 2021).

Atšķirībā no ROC līknes, kurai ir fiksēta bāzes līnija 0,5, precizitātes-pārklājuma līknes bāzes līnija ir atkarīga no sastopamības testa datos (Valavi u.c., 2022).

1.3. TSS

TSS ir modeļu novērtēšanas metrika, kas ņem vērā II tipa, gan I tipa kļūdas un nejaušību. TSS ir apvieno jutīgumu un specifiskumu un no tā, atņemot precizitāti, kas varētu rasties nejaušības dēļ (*random guessing*) (Allouche u.c., 2006).

$$\text{TSS} = \text{jutīgums} + \text{specifiskums} - 1 = \frac{\text{TP}}{(\text{TP} + \text{FN})} + \frac{\text{TN}}{(\text{TN} + \text{FP})} - 1$$

TSS vērtība ir diapazonā [0; 1]. Dažādos pētījumos atšķiras uzskati par to, kāda TSS vērtība norāda, kas ir vājš, vidējs un labs modelis (Allouche u.c., 2006), (Sadeghi u.c., 2024), (Bania u.c., 2025). Bet orientējoši tas ir (Avotiņš, 2025):

- TSS < 0.2 raksturīgs vājiem (*poor*) modeļiem;
- 0.2 ≤ TSS < 0.4 – viduvējiem (*fair*) modeļiem;
- 0.4 ≤ TSS < 0.6 – mēreni labiem (*moderately good*) modeļiem;
- 0.6 ≤ TSS < 0.8 – labiem (*good*) modeļiem;
- TSS ≥ 0.8 – izcili prognozējošiem (*excellent*) modeļiem.

1.4. Koena kappa

Kādreiz tika uzskatīts, ka Koena kappa, koriģē modeļa prognozes precizitātes novērtējumu, ņemot vērā precizitāti, kas varētu rasties nejaušības dēļ (Allouche u.c., 2006). Metrika ņem vērā arī I tipa un II tipa kļūdas un tam ir relatīva tolerance pret nulles vērtībām pārpratumu matricā. Metrikas vērtības ir diapazonā [-1, +1], kur +1 norāda uz izcilu modeli, bet nulle vai zemāka vērtība norāda uz veiktspēju, kas nav labāka par nejaušu.

$$k = \frac{\text{pareizība} - \text{nejaušības korekcija}}{1 - \text{nejaušības korekcija}} =$$

$$= \frac{\frac{TP + TN}{TP + FP + FN + TN} - \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TP + FP + FN + TN)^2}}{1 - \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TP + FP + FN + TN)^2}}$$

1.5. Nepārtrauktais Boisa indekss (CBI)

Originālajā Boisa indeksā biotopa piemērotība ir vērtību diapazonu klasēs (*bins*) (Boyce u.c., 2002). Bet tā vēlākajā modifikācijā - nepārtrauktajā Boisa indeksā (continuous Boyce index, CBI) klašu vietā lieto slīdošo jeb pārklājošos logu (Hirzel u.c., 2006) (4. attēls). CBI balstās uz Spīrmena ranga korelācijas koeficientu (R_s) starp klātbūtnes punktu proporcijas (P) un modelējamajā reģionā pieejamās teritorijas proporcijas (E) attiecību, vērtējot sugas sastopamību no zemas līdz augstai relatīvai biotopa piemērotībai (Glenn u.c., 2016).

Precīzāk, katrai klasei i tiek aprēķinātas divas frekvences: P_i – prognozētā klātbūtnes punktu frekvence un E_i – sagaidāmā klātbūtnes punktu frekvence, t. i., frekvence, ko paredzētu nejaušs punktu izkliedējums visā pētījuma teritorijā, to nosaka katras klases aizņemtās platības relatīvais lielums (Hirzel u.c., 2006).

$$P_i = \frac{p_i}{\sum_{j=1}^b p_j}$$

kur b – klašu skaits,

p_i - to novērtējuma punktu skaits, ko modelis prognozē nokrītam biotopa piemērotības klasē i ,

p_j - kopējais novērtējuma punktu skaits.

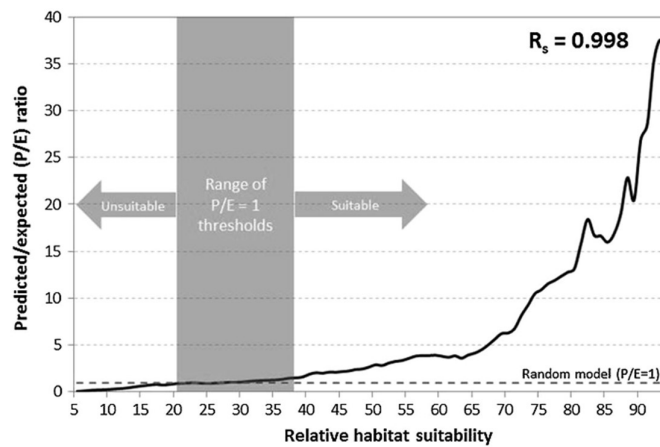
$$E_i = \frac{a_i}{\sum_{j=1}^b a_j}$$

kur a_i ir režģa šūnu skaits, kas pieder biotopa piemērotības klasei i , vai arī klasei i atbilstošās platības lielums,

a_j - kopējais šūnu skaits visā pētījuma teritorijā.

Katrai klasei i aprēķina prognozēts-sagaidāms proporciju F_i :

$$F_i = \frac{P_i}{E_i}$$



4. attēls. Prognozētā un sagaidāmā (P/E) līkne un nepārtrauktā Boisa indeksa (CBI) vērtība, kas aprēķināta, izmantojot Spīrmena rangu (R_s) analīzi. X ass ir saīsināta no 5 līdz 95, jo izmantots 10 intervālu pārvietojošais vidējais (Glenn u.c., 2016).

CBI ir pieskaitāma pie kalibrācijas novērtēšanas metodēm, jo palīdz noteikt, cik labi modelis ir kalibrēts un pārnesams (*transferable*) (Manzoor u.c., 2018). Iepriekš apskatītās diskriminācijas novērtējuma metodes vairāk ir saistītas ar klātbūtnes un iztrūkuma klasificēšanas pareizības novērtēšanu. Kalibrācija ir raksturo, cik labi modeļa prognozētās klātbūtnes varbūtības atbilst faktiskajai sastopamībai. Pat ja modelis sasniedz augstu diskrimināciju apmācības vidē (piemēram, augstu AUC vai TSS), tas var būt slikti kalibrēts un tādēļ nespēj uzticami darboties citā telpā vai vides apstākļos.

2. TIKAI KLĀTBŪTNES DATU IZMANTOŠANAS RADĪTĀS PROBLĒMAS

Tikai klātbūtnes modeļi parasti novērtē relatīvās sastopamības varbūtības vai biotopa piemērotības indeksus, nevis absolūtās klātbūtnes varbūtības. Šī atšķirība rodas tādēļ, ka datus trūkst iztrūkumu, un tāpēc jāpaļaujas uz fona datiem kā aizstājējiem. Tomēr šādā pieejā mēra to, cik labi modelis atšķir klātbūtni no fona, nevis no reāliem iztrūkumiem, kas sarežģī interpretāciju. Prognozētā varbūtība atspoguļo vietas izvēles iespējamību klātbūtnei, nevis sugas faktisko sastopamību, kas neizpētītās teritorijās var novest pie pāraprognozēšanas vai nepietiekamas prognozēšanas.

Modeļu, kas izmanto tikai klātbūtnes datus, novērtēšana ir saistīta ar virkni problēmu:

- **Pseidoiztrūkuma vai fona dati** - lai arī modeļus sauc par tikai klātbūtnes datu modeļiem, tie tāpat izmanto iztrūkuma datus - pseidoiztrūkuma vai fona datus. Savukārt pašas novērtēšanas metodes AUC, Koena kappa un TSS ir problemātiski izmantot šādām modelēšanās metodēm, jo tās pēc definīcijas izmanto iztrūkuma datus, pieņemot, ka tas ir patiess iztrūkums, nevis pseidoiztrūkums vai fons (Leroy u.c., 2018), (Grimmett u.c., 2020).

- **Datu kopas izmērs** – AUC un TSS vērtības ir augstākas datu kopās ar lielāku datu punktu skaitu (Konowalik & Nosol, 2021), kā arī AUC kļūdas ir lielākas, ja paraugi ir maza izmēra (Valavi u.c., 2022).

- **Sastopamības biežuma ietekme** - AUC, Koena kappa un TSS nav sastopamības biežuma neatkarīgas (Leroy u.c., 2018), (Allouche u.c., 2006), (Somodi u.c., 2017), (Manel u.c., 2001), (McPherson u.c., 2004). Šādā gadījumā metrikas vairs nesniedz informāciju par modeļa veikspēju, bet drīzāk raksturo datu dabu. Metrikas, kas izmanto pareizību, retām sugām dod kļūdaini augstus novērtējumus, jo patiesībā novērtē iztrūkumus.

Ne gluži metrikas problēma, bet tomēr saistīta ar tikai klātbūtnes modeļiem ir **paraugošanas novirze (sampling bias)** (Phillips u.c., 2009) – tā gan nav tieši saistīta ar modeļu novērtēšanu, tomēr ņemot vērā, ka tikai klātbūtnes dati aizvien biežāk ir oportūniski novērojumi, attiecīgi dati bieži paši par sevi neatspoguļo pašas piemērotākās sugas dzīvotnes.

3. METRIKU PIELIETOJUMS DAŽĀDĀM KOPĀM

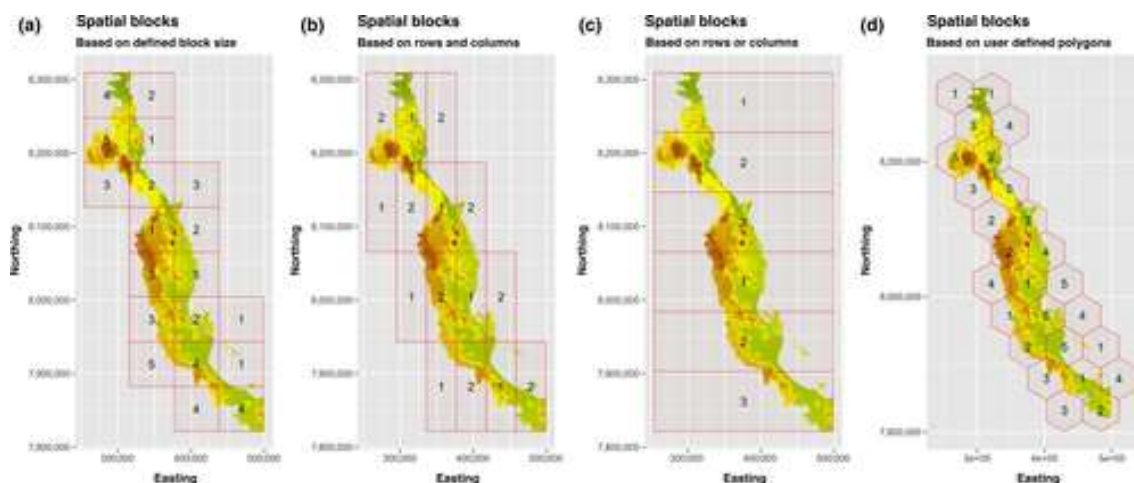
Modelēšanā ir pieņemts visus datus dalīt divās vai trīs daļās:

- apmācības un testa datu kopās gadījumos, ja neveic hiperparametru pielāgošanu (*hyperparameter tuning*);
- apmācības, validācijas un testa kopās, ja veic hiperparametru pielāgošanu.

Tā kā katrai kopai ir savs izmantošanas mērķis, arī modeļu novērtēšanas metrikas tiek pielietotas atšķirīgi (1. tabula). Parasti ar apmācības datiem veic modeļa sarežģītības, nevis veikspējas novērtēšanu. Bieži metrikas izmanto, salīdzinot to vērtējumus apmācības un validācijas kopās (gan kopā, gan atsevišķi), lai modeļiem vērtētu pārpielāgošanos un veiktu sākotnējo modeļu atlasu un hiperparametru pielāgošanu. Savukārt testa datus izmanto – gala modeļa veikspējas novērtēšanai. Šie dati iepriekš nav bijuši iesaistīti modeļa apmācībā un validācijā.

Pirms metožu izmantošanas dažādām datu kopām apkopojuma, ir jāapraksta šķērvalidācijas (*cross-validation*) process, kas īsti nav novērtēšanas metode, jo tur var izmantot dažādas novērtēšanas metodes, bet drīzāk novērtēšanas process (Arlot & Celisse, 2010). Šķērvalidācijā izmanto apvienotu datu kopu, ko veido apmācības un validācijas kopas, bet testa kopa paliek neizmantota. Šķērvalidācijā izmantoti dati tiek sadalīti k apakškopās (*folds*) un modeli apmāca k reizes, katru reizi viena daļa ir apmācības dati un k-1 ir validācijas dati. Katrā apmācības reizē cita apakškopa tiek izmantota validācijai.

Nemot vērā, ka sugu izplatības modelēšanā strādā ar telpiskiem datiem, tad šķēršvalidācijā datus nevar sadalīt gadījuma ceļā, bet ir jāievēro to telpiskais novietojums un dati ir jādala telpiskos blokos (5. attēls) (Valavi u.c., 2019).



5. attēls. Dažādas telpas sadalīšanas stratēģijas (Valavi u.c., 2019).

Zemāk (1. tabulā) ir dots apkopojums par metriku izmantošanu noteiktās datu kopās.

1. tabula.

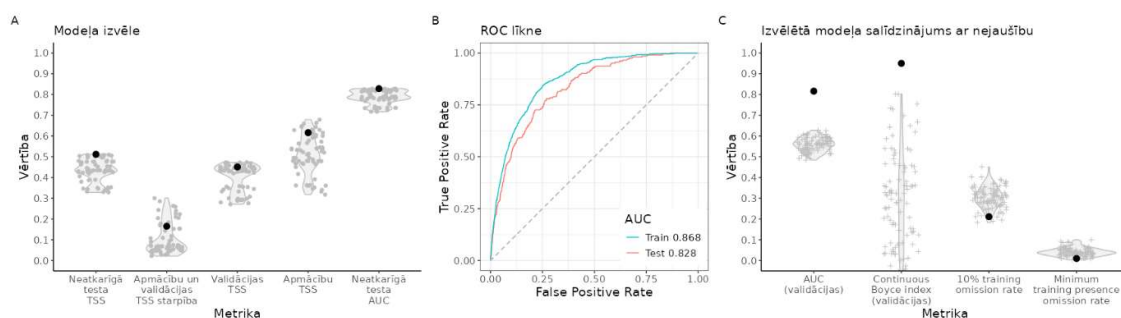
Metriku lietojums putnu sugu izplatības modelēšanā (Avotiņš, 2025).

Kopa	Metrika	Izmantošanas mērķis
Apmācības kopa	TSS*	Informatīvs nolūks.
Apmācības kopa + validācijas kopa	$TSS_{\text{apmācības}} - TSS_{\text{validācijas}}^*$	Izmanto labākā modeļa izvēlei. Zemāka vērtība norāda uz modeļa stabilitāti. Iespējams, ka līdzīgi varētu izmantot arī vidējo TSS indeksu, kas ir vidējais no $TSS_{\text{testa}} / TSS_{\text{apmācības}}$ un apkopo modeļa pārpielāgošanās pakāpi (Parimbelli u.c., 2025).
Validācijas kopa	TSS*	Novērtēt modeļa spēju prognozēt vides piemērotību telpiski nodalītās vietās. Jo augstāka šī vērtība, jo labāka ir modeļa spēja
Validācijas kopa	AUC	Informatīvs nolūks. Tomēr sniedz informāciju par modeļa stabilitāti un izmantojams salīdzināšanai ar nulles modeli.

Kopa	Metrika	Izmantošanas mērķis
Validācijas kopa	CBI	Raksturo modeļa prognozēšanas spēju. Jo lielāks, jo labāk.
Testa kopa	TSS	Novērtē modeļa veiktspēju darbā ar jauniem, iepriekš neredzētiem datiem.
Testa kopa	AUC	Informatīvs nolūks.

* Metrikas nav izmantotas tieši konkrētajai kopai. Bet gan ir veikta šķērsvalidācija, kur ir izmantoti visi apmācības un validācijas kopas dati.

6. attēlā ir dots piemērs, kā izskatās vienas putnu sugas labākā modeļa novērtējumi ar dažādām metriķām, salīdzinot ar alternatīvo modeļu vērtējumiem.



6. attēls. Dažādu metriķu pielietojums modeļu izvēlei un novērtēšanai (Avotiņš, 2025).

Papildus modeļu veiktspējas metriķām, tiek izmantotas arī modeļu sarežģītības novērtēšanas metriķas, piemēram, koriģētais Akaike informācijas kritērijs (*Akaike Information Criterion corrected, AICc*). To izmanto visai novērojumu kopai, nedalot to sīkāk (Vignali u.c., 2020). Šo rādītāju izmanto, lai salīdzinātu modeļu sarežģītību, nevis to prognozēšanas spējas, un to izmanto, lai salīdzinātu konkurējošus modeļus, kas apmācīti, izmantojot tos pašus datus, bet ne galīgai modeļa novērtēšanai.

IZMANTOTĀ LITERATŪRA

- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223–1232. <https://doi.org/10.1111/J.1365-2664.2006.01214.X>;WEBSITE:WEBSITE:BESJOURNALS;ISSUE:ISSUE:DOI
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40–79. <https://doi.org/10.1214/09-SS054>
- Avotiņš, A. (2025, maijā 28). *Putnu sugu izplatības modelēšana biodaudzveidības aizsardzībai: materiāli reproducēšanai*. https://aavotins.github.io/PutnuSDMs_gramata/
- Bania, J. K., Deka, J. R., Paul, A., Nath, A. J., Sileshi, G. W., & Das, A. K. (2025). Highly suitable areas for tea (*Camellia sinensis*) production will decline under future climate change scenarios. *Environmental and Sustainability Indicators*, 26, 100720. <https://doi.org/10.1016/J.INDIC.2025.100720>
- Boyce, M. S., Vernier, P. R., Nielsen, S. E., & Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecological Modelling*, 157(2–3), 281–300. [https://doi.org/10.1016/S0304-3800\(02\)00200-4](https://doi.org/10.1016/S0304-3800(02)00200-4)
- Bracho-Estévez, C. A., Arenas-Castro, S., González-Varo, J. P., & González-Moreno, P. (2024). Spatially explicit metrics improve the evaluation of species distribution models facing sampling biases. *Ecological Informatics*, 84, 102916. <https://doi.org/10.1016/J.ECOINF.2024.102916>
- Glenn, E. M., Lesmeister, D. B., Davis, R. J., Hollen, B., & Poopatanapong, A. (2016). Estimating density of a territorial species in a dynamic landscape. *Landscape Ecology* 2016 32:3, 32(3), 563–579. <https://doi.org/10.1007/S10980-016-0467-6>
- Grimmett, L., Whitsed, R., & Horta, A. (2020). Presence-only species distribution models are sensitive to sample prevalence: Evaluating models using spatial prediction stability and accuracy metrics. *Ecological Modelling*, 431, 109194. <https://doi.org/10.1016/J.ECOLMODEL.2020.109194>
- Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199(2), 142–152. <https://doi.org/10.1016/J.ECOLMODEL.2006.05.017>
- Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, 21(4), 498–507. <https://doi.org/10.1111/J.1466-8238.2011.00683.X>
- Konowalik, K., & Nosol, A. (2021). Evaluation metrics and validation of presence-only species distribution models based on distributional maps with varying coverage. *Scientific Reports*, 11(1). <https://doi.org/10.1038/S41598-020-80062-1>
- Leroy, B., Delsol, R., Hugueny, B., Meynard, C. N., Barhoumi, C., Barbet-Massin, M., & Bellard, C. (2018). Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography*, 45(9), 1994–2002. <https://doi.org/10.1111/JBI.13402>
- Li, W., & Guo, Q. (2013). How to assess the prediction accuracy of species presence-absence models without absence data? *Ecography*, 36(7), 788–799. <https://doi.org/10.1111/J.1600-0587.2013.07585.X>

- Li, W., & Guo, Q. (2021). Plotting receiver operating characteristic and precision–recall curves from presence and background data. *Ecology and Evolution*, 11(15), 10192. <https://doi.org/10.1002/ECE3.7826>
- Liu, C., Newell, G., White, M., & Machunter, J. (2025). Improving the estimation of the Boyce index using statistical smoothing methods for evaluating species distribution models with presence-only data. *Ecography*, 2025(1). <https://doi.org/10.1111/ecog.07218>
- Lobo, J. M., Jiménez-valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17(2), 145–151. <https://doi.org/10.1111/J.1466-8238.2007.00358.X>
- Manel, S., Ceri Williams, H., & Ormerod, S. J. (2001). Evaluating presence-absence models in ecology: The need to account for prevalence. *Journal of Applied Ecology*, 38(5), 921–931. <https://doi.org/10.1046/J.1365-2664.2001.00647.X>;ISSUE:ISSUE:DOI
- Manzoor, S. A., Griffiths, G., & Lukac, M. (2018). Species distribution model transferability and model grain size – finer may not always be better. *Scientific Reports 2018 8:1*, 8(1), 7168-. <https://doi.org/10.1038/s41598-018-25437-1>
- McPherson, J. M., Jetz, W., & Rogers, D. J. (2004). The effects of species' range sizes on the accuracy of distribution models: Ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, 41(5), 811–823. <https://doi.org/10.1111/J.0021-8901.2004.00943.X>;ISSUE:ISSUE:DOI
- Parimbelli, A., Johnson, M. P., Howell, K., Laguionie-Marchais, C., & Allock, A. L. (2025). Licence to predict – Investigating approaches to modelling low-occurrence deep-sea Irish Antipatharia with a new evaluation metric. *Mediterranean Marine Science*, 26(2), 400–417. <https://doi.org/10.12681/mms.39523>
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19(1), 181–197. <https://doi.org/10.1890/07-2153.1>
- Sadeghi, M., Malekian, M., & Esfahani, M. T. (2024). Interspecific niche overlap and climatic associations of native Quercus species in the Zagros forests of Iran. *Global Ecology and Conservation*, 51, e02878. <https://doi.org/10.1016/J.GECCO.2024.E02878>
- Somodi, I., Lepesi, N., & Botta-Dukát, Z. (2017). Prevalence dependence in model goodness measures with special emphasis on true skill statistics. *Ecology and Evolution*, 7(3), 863. <https://doi.org/10.1002/ECE3.2654>
- Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10(2), 225–232. <https://doi.org/10.1111/2041-210X.13107>;REQUESTEDJOURNAL:JOURNAL:2041210X;WEBSITE:WEBSITE:BE SJOURNALS;WGROU:STRING:PUBLICATION
- Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2022). Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs*, 92(1), e01486. <https://doi.org/10.1002/ECM.1486>;WEBSITE:WEBSITE:ESAJOURNALS;JOURNAL: JOURNAL:15577015;WGROU:STRING:PUBLICATION

- Vignali, S., Barras, A. G., Arlettaz, R., & Braunisch, V. (2020). SDMtune: An R package to tune and evaluate species distribution models. *Ecology and Evolution*, 10(20), 11488–11506. <https://doi.org/10.1002/ECE3.6786>;WGROU:STRING:PUBLICATION
- Wunderlich, R. F., Lin, Y.-P., Anthony, J., & Petway, J. R. (2019). *Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models Launched to accelerate biodiversity conservation*. <https://doi.org/10.3897/natureconservation.35.33918>