

EKOĢEOGRĀFISKO MAINĪGO IZVĒLE

Tulkojumi un saīsinājumi

- *Species distribution modelling/models* Sugu izplatības modelēšana/modeli SDM (lietots, kad nav svarīgi nošķirt, vai tiek runāts par modelēšanu vai modeļiem)
- *Ecogeographical/Environmental variables/predictors* Ekoģeogrāfiskie mainīgie EGV
- *Maximum entropy* Maksimālā entropija MaxEnt
- *Variance inflation factor* Dispersijas inflācijas koeficients VIF
- *Primary component analysis* Galveno komponentu analīze PCA
- *Pairwise correlation coefficient* Pārveida korelācijas koeficients

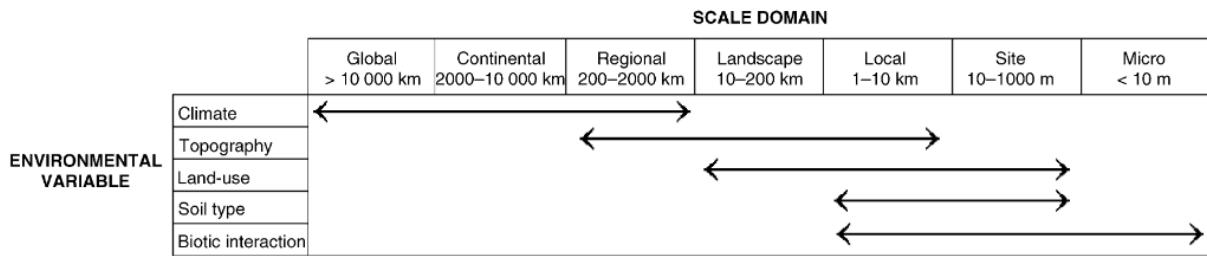
Vispārīgi apsvērumi EGV izvēlē

Iepazīstoties ar literatūru, izkristalizējas dažas pamatprasības un apsvērumi, kad tiek izvēlēti EGV. Lai arī liela daļa no turpmāk minētā var likties pašsaprotama, tomēr nekaitē par to atgādināt, tādēļ nodaļā vēlos iepazīstināt ar apsvērumiem, kas vienmēr ir jāņem vērā pirms citiem EGV atlases veidiem.

Viens no svarīgākajiem faktoriem, kas jāapsver EGV izvēlē, ir modelēšanai izvēlēties tādus mainīgos, kas patiešām ir nozīmīgi sugas dzīves vides aptveršanai (Araújo & Guisan, 2006; Austin, 2002). Jāatcerās, ka SD modelēšana nav “tīra” matemātika, bet gan darbs ar īstām sugām, tādēļ ekoloģiskiem apsvērumiem vajadzētu dot priekšroku (Dormann *et al.*, 2013).

Ir jādomā par EGV mērogu salīdzinājumā ar modelējamo teritoriju (Dormann, 2007). Ir EGV, piemēram, klimata, kuriem bieži nav pietiekami liela variabilitāte mazos mērogos, lai tie būtu īpaši noderīgi SDM pētījumos mazās teritoijās (Dormann, 2007; Mackey & Lindenmayer, 2001). Šādos pētījumos visdrīzāk noderīgāki būs tādi EGV, kuriem ir iespējama liela variabilitāte attiecīgajā mērogā, piemēram, zemes lietojuma, veģetācijas indeksa EGV (Phillips *et al.*, 2006). Kopumā, manuprāt, šis liek noprast, ka projekta vajadzībām daudz lielāka nozīme būs EGV ar augstu izšķirtspēju, kā zemes lietojuma EGV, nevis klimata

(1. attēls). Tomēr Latvija nav arī tik maza, lai apgalvotu, ka klimats visur būtu vienāds, tādēļ tos tāpat visdrīzāk ir jāņem vērā (1. attēls).



1. attēls. Shēma, kas parāda, kuras EGV grupas būtu piemērotas izmantot attiecīgajos mērogos (Pearson & Dawson, 2003).

Projektā saskaramies ar EGV, kas ir par vienu un to pašu “tēmu”, bet dažādos mērogos, piemēram, skujkoku mežu platības īpatsvars 1 ha, 500 m, 1250 m, 3000 m, 10 000 m. Ja pastāv šāda EGV struktūra, tad ir svarīgi domāt par to, kas ietekmē dažādus modelējamās sugas ekoloģijas aspektus kāda ir dzīves telpa, kāda ir raksturīgā dispersija? Parasti EGV, kas vēsta par kādu dzīvībai nepieciešamo, piemēram, barības avotu, ūdens vai arī specifisku dzīivotņu, resursu pieejamību vajadzētu modelēt sugas dzīves telpas mērogā (Fournier *et al.*, 2017). Savukārt par klimata un citus dispersijas nozīmīgos EGV vajadzētu izmantot attiecīgi dispersijas mērogā (Fournier *et al.*, 2017).

Svarīgi ir ņemt vērā temporālo variabilitāti. EGV, kas radīti no pēdējo, teiksim, 50 gadu datiem, visdrīzāk nebūs noderīgi, ja pētījumā tiek lietoti pēdējo piecu gadu novērojumi, jo 50 gadu dati visdrīzāk precīzi neapraksta pēdējo gadu variabilitāti (Anderson & Martinez-Meyer, 2004). Par šo projekta ietvaros visdrīzāk nav jāuztraucas, jo gan EGV, gan novērojumi nāk no viena laika perioda.

Izvēloties EGV, jāatcerās, ka to kombinācija ietekmē to, cik plaši izmantojams būs modelis (Phillips *et al.*, 2006). Tas nozīmē to, ka ne vienmēr EGV, kas atzīti par nozīmīgiem vienā pētījumā, būs tik pat noderīgi modelēšanai pētījumam citā teritorijā. EGV, kas atspoguļo resursu, kā ūdens, saules, barības vielu, pieejamību, parasti veiksmīgi pārnesīsies uz citiem modeļiem (Mackey & Lindenmayer, 2001), taču EGV kā augstums virs jūras līmeņa gan ne (Phillips *et al.*, 2006). Spekulēju, ka tas visdrīzāk ir tādēļ, ka ir sugars, kuras citās vietās var pastāvēt tikai noteiktos apstākļos, kas citos reģionos ir pieejami tikai kalnainā vidē, kamēr, piemēram, Latvijā, sugars pamatvajadzības tiek izpildītas arī teritorijās, ar augstumu ~ 0 m v.j.l. Šo svarīgi ir atcerēties, ja tiek atrasts raksts par SDM, kurā modelētas interesējošas sugars citā pasaules reģionā, kas radikāli atšķiras no Latvijas.

Lai varētu veikt turmāku atlasi, ir nepieciešams izveidot kopu ar EGV, kas ļem vērā augstāk minētos apsvērumus.

Esošās pieejas EGV atlasē MaxEnt SDM

Viens no būtiskākajiem iemesliem EGV atlasīšanai ir multikolinaritātes novēršana. Tādēļ liela daļa no EGV atlasīšanas pieejām būtībā ir multikolinearitātes novēršanas metodes. Jāņem vērā, ka multikolinearitāti nav iespējams pilnīgi izskauст, tādēļ pirms, balstoties uz statistiskajiem testiem, atbrīvojas no kāda EGV, jāpadomā, vai mainīgā izķļaušana no analīzes nebūtu sliktāka nekā tā atstāšana (Dormann *et al.*, 2013). Pastāv vairākas pieejas, ar kurām tiek atlasīti EGV MaxEnt SDM, taču tās vairāk vai mazāk vienmēr atgriežas pie (aptuveni) trīs populārākajām pieejām: ordinācijas, VIF, korelācijas koeficienti.

Tā kā ir zināms, ka modelis ar visiem EGV parasti būs gandrīz nelietojams, bieži zinātnieki cenšas atlasīt atbilstošākos EGV pirms modeļa izveides. Šim nolūkam tiek lietoti dažādi statistiskie rādītāji un analīzes veidi kā ordinācija, VIF, klāsteranalīze u.c. (Dormann *et al.*, 2013). No ordinācijas metodēm biežāk tiek lietota PCA, bet pastāv vēl citi veidi (Zhang *et al.*, 2023). Tā kā visbiežāk tiek lietots PCA, kā arī tā ir atzīta kā viena no efektīvākajām metodēm šāda veida atlasei (Zhang *et al.*, 2023), rindkopā izskatīšu tikai to. Viena opcija ir skatīties eigenvērtības, kas tiek salīdzinātas pret izvēlēto sliekšņa līmeni (nereti 0.7-1) (King & Jackson, 1999). Šādā gadījumā no sākuma veic analīzi, lietojot datus no visiem EGV, ko varētu izmantot sugas modelēšanai, kā arī ar klātbūtnes datiem. Ja eigenvērtības ir zemākas par sliekšņa līmeni, tad šie EGV tiek atmesti (King & Jackson, 1999). Otra opcija ir skatīties uz komponenšu skaidroto varianci. Tieki analizēti komponenti, kuru kopējā variance skaidro “pietiekami daudz” variances, t.i., 70-95% (De Marco & Nóbrega, 2018). Katrs komponents sastāv no visiem analīzē iekļautajiem EGV. Tieki paturēti tie EGV, kuriem ir vislielākais svars atlasītajos komponentos (De Marco & Nóbrega, 2018).

Populāra pieeja ir izmantot pārveida korelācijas koeficientus, VIF, vai arī citas statistikas vērtības kā multikolinearitātes indikatorus (Dormann *et al.*, 2013). Ja izlemts darboties ar VIF, tad parasti tiek aprēķināts VIF katram no izvēlētajiem EGV. Ja VIF ir lielāks par desmit, tas tiek atmests (Dormann *et al.*, 2013). Tomēr, jāizmanto piesardzība, kad tiek atmesti mainīgie, jo atbrīvošanās no viena izmaina visu pārējo VIF. Tādēļ nav iesakāms atbrīvoties no visiem EGV, kuru VIF pārsniedz desmit vienā reizē, bet gan pakāpeniski, aprēķinot VIF ikreizi, kad tiek atmests mainīgais, līdz vairs nav EGV ar VIF lielāku par desmit (Naimi & Araújo, 2016). Darbojoties ar pārveida korelācijas koeficientiem, tiek aprēķināta

korelācija starp katru mainīgo pāri. Ja korelācijas koeficients pārī ir lielāks nekā izvēlētais sliekšņa līmenis, tad vienu no mainīgajiem atmet, līdz vairs nav pāru, starp kuriem novērojama stipra korelācija (Naimi & Araújo, 2016). Visbiežākais izmantotais sliekšņa līmenis, kuru pārsniedzot ir jāatmet mainīgais, ir $r > 0.7$ (Dormann *et al.*, 2013). Izmantojot pārveida korelāciju koeficientus, var būt sarežģīti izšķirties, kuru EGV vajadzētu neiekļaut. Dažreiz šo var atrisināt, domājot, kurš no diviem EGV modelējamajai sugai ir ekoloģiski nozīmīgāks (Dormann *et al.*, 2013).

Pastāv arī pieeja, kura savā ziņā apvieno jau iepriekš minētās pieejas automatizēta EGV atlase. Ja iepriekš minētajās metodēs tomēr liels uzsvars likts arī uz ekoloģiskajām zināšanām, tad automatizētā atlasē tas gluži nav iespējams. Tomēr automatizēta atlase nereti balstās uz jau minētajām metodēm PCA, atlasē, kas balstīta uz VIF vai korelācijas koeficientiem. Piemēram, R pakotnē “sdm” piedāvā pieeju, kas apvieno gan VIF, gan pārveida korelācijas koeficientu aprēķināšanu. No sākuma aprēķinot katru mainīgo pāru korelācijas koeficientus, tiek skatīts EGV pāris ar augstāko korelācijas koeficientu, un tad tiek atmests mainīgais, kuram no pāra ir vislielākais VIF. Procedūra tiek atkārtota, līdz vairs nepastāv pāri ar korelācijas koeficientu augstāku par sliekšņa līmeni (Naimi & Araújo, 2016).

Vēl viena pieeja, ar ko projekta ietvaros jau esam saskārušies, ir EGV izvēle, balstoties uz ekspertu zināšanām par sugas ekoloģiju. Kā jau mēs (projekta dalībnieki) zinām, šajā pieejā mainīgos atlasa ekologi, balstoties uz zināšanām par sugai nepieciešamajiem resursiem un apstākļiem. Ir novērots, ka modeļi, kas izmanto ekspertu izvēlētos EGV, ir pielīdzināmi (Harris *et al.*, 2013; Lin & Chiu, 2020) un arī var pārspēt modeļus, kuros EGV tika izvēlti, lietojot algoritmus vai citas automatizētas/matemātiskas pieejas, modeļu kvalitātes metrikās (Mod *et al.*, 2016; Romero *et al.*, 2023). Manuprāt, šis ir skaidrojams ar to, ka ekspertiem piemīt intuīcija, kuras datoriem vienkārši nav. Eksperti parasti zina arī ekoloģiju, un ir spējīgi aizpildīt informācijas tukšumus par specifiskām sugas vajadzībām ar zināšanām, kas nāk no teorijas, pieredzes dabā un darba ar citām sugām. Šī īpašība ir īpaši aktuāla, ja nepieciešams modelēt retu un mazpētītu sugu izplatību (Romero *et al.*, 2023). MaxEnt SDM kontekstā ekspertu zināšanas varētu būt vēl svarīgākas, jo MaxEnt lieto tikai novērojumu klātbūtnes datus (Phillips *et al.*, 2006). Tādēļ ir nepieciešams kaut kā definēt arī to daļu vides, kur suga tiešām nebūs sastopama, ko automatizēta mainīgo izvēle nav spējīga veikt uzticami.

Tomēr nodaļas beigās vēlos teikt, ka neviena pieeja neeksistē vakuumā. Lietojot vienu pieeju, nav jau izmantot arī vēl kādu pieeju, lai pārliecīnātos, ka modelī tiek likti piemērotākie

EGV. Iepazīstoties ar literatūru, reti kura publikācija izmantoja tikai vienu pieeju. Bieži tās tika lietotas tandēmā, piemēram, ekspertiem tika sniegts saraksts ar EGV, kas atlasīti lietojot PCA, tad ekspertu izvēles atkārtoti analizētas ar PCA (Tourne *et al.*, 2019).

Izskaņa un pārdomas

Projekta sanāksmēs īpaša uzmanība tika pievērsta EGV mērogam vai izvēlamies šūnas, 500 m, 1250 m, 3000 m vai arī 10 000 m mērogu. Tādēļ liktos, ka šīs idejas būtu plaši apskatītas un izvērtētas literatūrā, taču pārsteidzošā kārtā referāta rakstīšanas laikā, kamēr specifiski nemeklēju, es nesaskāros ar šādu parādību ka viens EGV ir pieejams vairāk kā vienā mērogā. Mani arī pārsteidza, ka citos pētījumos lielākais EGV skaits pirms to atlases bija aptuveni 150, un reti tas bija vairāk par ~50. Projektā ir pieejami aptuveni 540 datu slāņi, kas tā pat ir krietni vairāk mainīgo nekā lērumā publikāciju, pat nēmot vērā, ka liela daļa no projekta mainīgajiem ir viena un tā paša mainīgā iterācijas dažādos mērogos. Manuprāt, šis piešķir kontekstu tam, ka mēs (projekta dalībnieki) esam ļoti unikālā situācijā, šķietami, tādā par ko citi SDM pētnieki varētu tikai sapņot (vai arī murgot).

Lai arī nozīmīgākā pieeja EGV atlasei ir ekoloģijā balstīta pieeja, bija gandrīz neispējami uztvert kādēļ tika izvēlēti tādi EGV, kādi tika izvēlēti. Es varbūt naivi sagaidīju, ka doma būs skaidri aprakstīta kā piemēram: “izvēlēts mainīgais Y, jo sugai nepieciešams y”, bet pat publikācijas, kas koncentrējās uz sugu ekspertu nozīmi SDM kontekstā īsti neapraksta šo ekspertu domu gaitu. Tas ir, manuprāt, skumji, jo tādā veidā nav iespējams smelties iedvesmu no citiem zinātniekiem, kā arī tādā veidā (ne)aprakstot šo speciālistu ekspertīzi, tā paliek pie viņiem, kas ir īpaši problemātiski, nēmot vērā, ka dabas ekspertu ir ļoti maz.

Kopsavilkumā, EGV izvēlei primāri ir jābūt balstītai nevis statistikā, bet gan ekoloģijā. Kad ir veikta EGV pirmatlase, nēmot vērā sugars ekoloģiju, tikai tad var pievērsties statistikas metodēm.

IZMANTOTĀ LITERATŪRA

- Anderson, R.P., Martinez-Meyer, E., 2004. Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. *Biol Conserv* 116, 167–179. [https://doi.org/10.1016/S0006-3207\(03\)00187-3](https://doi.org/10.1016/S0006-3207(03)00187-3)
- Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. *J Biogeogr* 33, 1677–1688. <https://doi.org/10.1111/j.1365-2699.2006.01584.x>
- Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol Modell* 157, 101–118. [https://doi.org/10.1016/S0304-3800\(02\)00205-3](https://doi.org/10.1016/S0304-3800(02)00205-3)
- De Marco, P., Nóbrega, C.C., 2018. Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *PLoS One* 13, e0202403. <https://doi.org/10.1371/journal.pone.0202403>
- Dormann, C.F., 2007. Promising the future? Global change projections of species distributions. *Basic Appl Ecol* 8, 387–397. <https://doi.org/10.1016/j.baae.2006.11.001>
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Fournier, A., Barbet-Massin, M., Rome, Q., Courchamp, F., 2017. Predicting species distribution combining multi-scale drivers. *Glob Ecol Conserv* 12, 215–226. <https://doi.org/10.1016/j.gecco.2017.11.002>
- Harris, R.M.B., Porfirio, L.L., Hugh, S., Lee, G., Bindoff, N.L., Mackey, B., Beeton, N.J., 2013. To Be Or Not to Be? Variable selection can change the projected fate of a threatened species under future climate. *Ecological Management & Restoration* 14, 230–234. <https://doi.org/10.1111/emr.12055>
- King, J.R., Jackson, D.A., 1999. Variable selection in large environmental data sets using principal components analysis. *Environmetrics* 10, 67–77.

[https://doi.org/10.1002/\(SICI\)1099-095X\(199901/02\)10:1<67::AID-ENV336>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1099-095X(199901/02)10:1<67::AID-ENV336>3.0.CO;2-0)

Lin, C.-T., Chiu, C.-A., 2020. COMPARISON OF PREDICTOR SELECTION PROCEDURES IN SPECIES DISTRIBUTION MODELING: A CASE STUDY OF *Fagus hayatae*. *CERNE* 26, 172–182. <https://doi.org/10.1590/01047760202026022657>

Mackey, B.G., Lindenmayer, D.B., 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. *J Biogeogr* 28, 1147–1166. <https://doi.org/10.1046/j.1365-2699.2001.00626.x>

Mod, H.K., Scherrer, D., Luoto, M., Guisan, A., 2016. What we use is not what we know: environmental predictors in plant distribution models. *Journal of Vegetation Science* 27, 1308–1322. <https://doi.org/10.1111/jvs.12444>

Naimi, B., Araújo, M.B., 2016. sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography* 39, 368–375. <https://doi.org/10.1111/ecog.01881>

Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* 12, 361–371. <https://doi.org/10.1046/j.1466-822X.2003.00042.x>

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol Modell* 190, 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>

Romero, D., Maneyro, R., Guerrero, J.C., Real, R., 2023. Using fuzzy logic to compare species distribution models developed on the basis of expert knowledge and sampling records: Expert knowledge versus sampling in species distribution modelling. *Front Zool* 20. <https://doi.org/10.1186/s12983-023-00515-x>

Tourne, D.C.M., Ballester, M.V.R., James, P.M.A., Martorano, L.G., Guedes, M.C., Thomas, E., 2019. Strategies to optimize modeling habitat suitability of *Bertholletia excelsa* in the Pan-Amazonia. *Ecol Evol* 9, 12623–12638. <https://doi.org/10.1002/ece3.5726>

Zhang, H., Guo, W., Wang, W., 2023. The dimensionality reductions of environmental variables have a significant effect on the performance of species distribution models. *Ecol Evol* 13. <https://doi.org/10.1002/ece3.10747>