

Posudek na diplomovou práci Jakuba Macha: Dlouhodobé změny jarní a podzimní fenologie ptáků na území ČR

Diplomová práce J. Macha se věnuje atraktivnímu a módnímu tématu korelací mezi proměnlivostí klimatu a ptačí fenologií. Práce je psána přehledně, ale až příliš stručně, což je zvláště patrné v Diskuzi. K práci mám následující konkrétní připomínky, dotazy a doporučení:

Český abstrakt: "... časování návratu ... je průkazně ($p < 0,05$) spojeno ..." Statistická průkaznost sama o sobě není nijak relevantní – závisí např. na externích faktorech nesouvisejících se studovaným jevem, např. velikostí vzorku (tj. s dostatečně velkým vzorkem najdeme signifikantní rozdíl či trend prakticky ve všem). To, že se něco v přírodě "mění", je triviální – mění se vše a pořád. A co si čtenář vezme se sdělení, že "časování návratu ... je spojeno s NAO indexem" pokud není uvedeno jak (pozitivně/negativně)? Shrnutí a sečteno: zajímavá a důležitá je (1) velikost efektu (jak moc se mění přílety/odlety?) a (2) směr efektu (přilétají/odlétají ptáci dříve nebo později?). Obě tyto informace však v Abstraktu chybí, resp. se objevují pouze pro některé studované odletové vztahy až na jeho konci.

Anglický abstrakt: text přesně neodpovídá české verzi (viz např. předposlední věta druhého odstavce). Používání trpného rodu je nevhodné, používejte rod činný. Krkolomnost trpného rodu je patrná např. na s. 12: "Data budou porovnána s domácí i zahraniční literaturou" – kým? Kdy? Nehledě na fakt, že srovnání s literaturou je nedílnou součástí jakékoli odborné práce – říká se tomu Diskuze. To nemusíte hlásit:)

s. 13: Dle Metodiky autor použil první resp. poslední příletová data. To je nevhodné, poněvadž následná analýza je nevyhnutelně zkreslena odlehlými hodnotami. Doporučuji pracovat s percentily, např. "odříznout" 5% krajních hodnot apod.

s. 13: Z popisu statistiky není úplně jasné, jak autor postupoval: použil "forward selection" nebo "backward elimination"? Co kontrola pro počet pozorování v daném roce, tj. pro reprezentativnost dat? Přílety i odlety by mohly záviset na nadmořské výšce – proč to autor nevzal v práci v úvahu? Nakonec dvě nejdůležitější věci:

- (1) Zredukování dat na roční průměry a mediány je špatná statistická praxe – ztrácí se většina informace, která byla horko těžko nasbírána, mizí např. informace o variabilitě v datech uvnitř let.
- (2) Pozorování pocházejí z opakovaně sledovaných stanic; data tedy mají strukturu, kterou je třeba zohlednit, tj. identita stanice jako tzv. náhodný efekt. Zvláště problematické může být nezahrnutí této proměnné, pokud nebyly všechny stanice obsazeny po všechny roky sledování. Pravidelnost sběru dat se mi u tak časo-prostorově rozsáhlého souboru dat jeví značně nepravděpodobná. Obrázky v příloze navíc ukazují mezery v datech, takže data jsou heterogenně "poshlukovaná" i v čase; to by možná měly analýzy také zohlednit.

s. 14, Tab. 1: V popisku tabulky měl autor vysvětlit, jaká data ukazuje, tj. jak si má čtenář interpretovat např. údaj "83,5" (správně to uvedl v Obr. 1–16). Škoda, že autor nepřevodl údaje na srozumitelnější kalendářní data.

s. 18: Z čísta jasna se zde objevuje sdělení, že "posun v datech po devadesátém roce ... [může] souviset se změnou skupiny pozorovatelů"! Z popisu Metod na s. 13 to jasně nevyplývá. Jde o závažný problém: metodika vždy ovlivňuje výsledky. Uvedená metodická změna měla být buď autorem vzata v úvahu ve statistickém zpracování, nebo měl jinak sebraná data analyzovat odděleně, případně novější data (s malým rozptylem časové řady) neanalyzovat vůbec.

Diskuze je *extrémně* stručná a chybí závěrečný "take-home message" odstavec.

Příloha 1, Tabulka 2:

- (1) Popisky jsou zmatené. Předpokládám, že "Průměr" a "Medián" odkazují na průměr a medián přiletu ptáků, zatímco v témže sloupci uvedené "NAO průměr" a "NAO medián" odkazují na průměr a medián NAO – první termíny tedy odkazují na *vysvětlovanou* odpověď, zatímco druhé termíny odkazují na *vysvětlující* prediktory!
- (2) V tabulce dále chybí statistiky pro – takový dojem to aspoň dělá – nevýznamné členy vyššího řádu (např. v prvním řádku je testová statistika a P-hodnoty pro "rok" a "rok²", ale chybí pro "rok³" – k dobrému statistickému vychování patří uvádět i nevýznamné statistiky a P-hodnoty, viz Nakagawa S. & Cuthill I. C. 2007: Effect size, confidence interval and statistical significance: a practical guide for biologists. Biological Reviews 82: 591–605.
- (3) Nevím, proč jsou tučně vyznačeny P-hodnoty pro nižší členy v modelech, kde jsou významné i členy vyššího řádu (např. první řádek – lineární člen je tučně v modelu, kde je významný nadřazený kvadratický člen) a naopak, tedy když nižší členy nejsou významné, tak nejsou vyznačeny tučně. V modelech s významnými vyššími členy je významnost nižších členů, kterou dodá statistický program, irelevantní – zůstávají v modelu bez ohledu na tuto významnost; jinými slovy: pokud je "rok²" významný, tak i "rok" je automaticky významný – vždyť je obsažen uvnitř prediktoru "rok²". Viz Grafen & Hails 2002: Modern statistics for the life sciences. Oxford: Oxford University Press (strana 193) a také Engqvist L. 2005: The mistreatment of covariate interaction terms in linear model analyses of behavioural and evolutionary ecology studies. Anim. Behav. 70: 967–971.

Příloha 2: Všechny obrázky mají extrémně malé popisky (písmo mělo být cca pětkrát větší!)

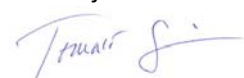
Obrázky ukazují korelaci mezi rokem (x) a daty příletů a odletů (y). To ovšem není biologicky nijak zajímavé a relevantní. "Rok" je v tomto typu práce abstraktní fyzikální veličina, která samozřejmě nemůže vykonávat – a to ani v principu – žádný kauzální vliv na chování ptáků. To, co naopak konat vliv může a co ostatně sám autor uvádí jinde v práci, je NAO index – změny počasí se *mohou* promítat do změn příletů a odletů; roky samozřejmě nemají jak. Výsledky takto prezentované (x = čas) nemají biologický význam, resp. jen částečně proto, že čas je korelátem NAO. Proto by bylo zajímavé uvést grafy kde x = NAO index (tj. ne rok).

Závěrem:

V Diskuzi by měl být autor opatrný – analyzovaná data jsou korelativní, což neumožňuje určit příčinu poměrně dosti heterogenních výsledků. Autor tak jistě nemůže kategoricky tvrdit, že "tyto časové posuny ve fenologii jsou reakcí především na probíhající změnu klimatu". Snad ano. Pokud však nejsou vzaty v úvahu další faktory, např. populační dynamika sledovaných druhů, vliv člověka na zimovištích či tahových cestách apod., nelze s jistotou říci nic.

Nakonec nejzávažnější připomínka: co je v této práci nulová hypotéza? Je evidentní, že pokud budu analyzovat dlouhé časové řady, navíc na mnoha druzích, vždy nevyhnutelně zjistím nějaké statisticky významné trendy. Věc začíná být zajímavá až ve chvíli, kdy (1) detekuji více než náhodné množství trendů, které jsou (2) konzistentní s biologicky smysluplným/i prediktorem/y a mají (3) velikost efektu, který je biologicky významný. Výsledky této práce – zjištěné trendy, jejich směry a velikosti – na mě dělají dojem náhodného šumu. Prosím o názor diplomanta k těmto připomínkám.

Přes uvedené připomínky doporučuji DP J. Macha k obhajobě a jako klasifikační stupeň navrhuji



Tomáš Grim
Katedra Zoologie UP