

Model pro predikci půdní diverzity bakterií a archaeí.

Autoři: Kopecný J.¹, Omelka M.², Tejnecký V.³, Vaníček J.⁴, Marečková, M.^{1,4}

Afilace autorů:

¹ Ekologie a epidemiologie mikroorganismů, Výzkumný ústav rostlinné výroby, v. v. i., Drnovská 507/73, 161 00 Praha 6 – Ruzyně

² Katedra pravděpodobnosti a matematické statistiky, Univerzita Karlova, Matematicko-fyzikální fakulta, Ke Karlovu 3, 121 16 Praha 2

³ Katedra pedologie a ochrany půd, Česká zemědělská univerzita v Praze, Kamýcká 129, 165 00 Praha – Suchdol

⁴ Katedra mikrobiologie, výživy a dietetiky, Česká zemědělská univerzita v Praze, Kamýcká 129, 165 00 Praha – Suchdol

Popis modelu

Model se snaží o zjednodušení hodnocení půdní biodiverzity pro účely ochrany přírody. Půdní biodiverzita je při ochraně opomíjena, i když má velký význam a jedním z důvodů je i relativní obtížnost jejího hodnocení. Model toto hodnocení zjednodušuje. Model bude aktualizován z dat roku 2024, kdy odebereme další lokality a doměříme některé proměnné, chemické i biologické.

Sestavení modelu (Doc. Ing. Marek Omelka, Ph.D. MFF UK)

Jako odezvu (vysvětlovanou proměnnou) jsme si v této prvotní fázi zvolili počet "druhů" resp. sekvenčních variant prokaryotních organismů nalezených v daném vzorku (dále značeno "P_sobs"). Jako vysvětlující (predikující) proměnné jsme pak měli k dispozici.

Kategorické proměnné

1. Půdní horizont - s možnostmi A, B, FH.
2. Typ místa – cesta ("c"), hospodářský les ("h") a starobylý les ("s").

Numerické proměnné:

3. pH půdy měřeno třemi různými způsoby – pH(H₂O), pH(CaCl₂), pH(BaCl₂)
4. Koncentrace kationtů - Al³⁺, Ca²⁺, Fe³⁺, K⁺, Mg²⁺, Mn²⁺, Na⁺, H⁺
5. Souhrnné ukazatele kationtů - BC, EA, CEC, BS
6. Množství organického uhlíku v % C_{ox}

K predikci jsme použili mnohorozměrnou lineární regresi (Zvára, 2008), která se pro svou univerzalitu a jednoduchost dobře hodí pro tyto druhy problémů, kdy nemáme představu o funkčním vztahu mezi odezvou a vysvětlujícími proměnnými. Výpočty jsme pak prováděli ve statistickém výpočetním prostředí R (R Core Team, 2023).

Hledali jsme model, který je na jednu stranu co nejjednodušší, tj. neobsahuje vysvětlující proměnné, které statisticky významně nepříspěvají k vysvětlení odezvy. Na druhou stranu však má co nejlepší prediktivní schopnost měřeno modifikovaným R², které lze interpretovat jako procento vysvětlené variability odezvy.

Při hledání modelu jsme zjistili následující:

1. Veličiny měřící pH, tj. $\text{pH}(\text{H}_2\text{O})$, $\text{pH}(\text{CaCl}_2)$ a $\text{pH}(\text{BaCl}_2)$ nesou velmi podobnou informaci a v modelu nám stačí pouze jedna z těchto veličin. Kvalita modelu je přitom velmi podobná, ať je v ní kterákoliv z těchto veličin.

2. Podobně veličiny měřící celkové množství kationtů BC, EA, CEC, BS nesou velmi podobnou informaci a v modelu lze použít téměř kteroukoliv z těchto veličin, přičemž kvalita modelu je srovnatelná.

3. Veličiny měřící pH a veličiny měřící celkové množství kationtů jsou silně korelované. Zahrnutí např. BS v okamžiku, kdy už je v modelu např. $\text{pH}(\text{H}_2\text{O})$, zlepšuje model již jen o pár procent (viz níže).

4. Veličiny měřící celkové množství kationtů BC, EA, CEC, BS jsou lepšími prediktory než jednotlivé kationty. Navíc se neukazuje jako přínosné zahrnutí některých jednotlivých kationtů v okamžiku, kdy tam již máme některou z proměnných BC, EA, CEC nebo BS.

5. Význam veličiny " C_{ox} " je ve srovnání s pH a souhrnnými ukazateli kationtů. Přínos je malý, nicméně data naznačují, že pro vyšší rozsah výběru by jeho přínos k vysvětlení odezvy již mohl být statisticky významný.

Vývoj analýzy dat

I. Model využívající pouze pH

Jelikož kvůli stále běžícím laboratorním analýzám máme relativně více chybějících dat pro veličiny $\text{pH}(\text{BaCl}_2)$, BC, EA, CEC, BS, tak jsme se nejdříve rozhodli pro model pouze s proměnnými horizont, typ místa a $\text{pH}_{\text{H}_2\text{O}}$ (použití $\text{pH}_{\text{CaCl}_2}$ místo $\text{pH}_{\text{H}_2\text{O}}$ by dalo v podstatě stejně dobrý model). Jelikož se ukázalo, že vliv $\text{pH}_{\text{H}_2\text{O}}$ je statisticky významně různý pro různé kombinace dvojic horizont a typ místa, tak pro predikci P obs (jsme dospěli k modelu

$$\text{predikce P obs} = a + b * \text{pH}(\text{H}_2\text{O})$$

kde koeficienty a (absolutní člen) a b (směrnice) lze nalézt v následující tabulce. Řádek přitom udává horizont a typ místa.

	a	b
A-c	1529.4	320.24
A-h	-654.8	681.66
A-p	-589.9	685.56
B-c	1552.3	179.15
B-h	1068.4	281.44
B-p	1189.6	253.20
FH-c	-1042.5	895.97
FH-h	165.7	499.13
FH-p	-1891.7	999.15

Tento model vysvětluje cca 54 % variability prokaryotní diverzity P_sobs.

II. Model využívající jak pH tak koncentrace kationtů

Níže popsaný model vysvětluje cca 58% variability, což je drobné zlepšení oproti modelu I. Důvodem je to, že veličiny měřící pH a veličiny BC, EA, CEC, BS jsou silně korelované a nesou do značné míry podobnou informaci.

predikce $P_{sobs} = b_0 + b_1 * pH_{BaCl_2} + b_2 * EA$,

kde koeficienty a (absolutní člen) a b (směrnice) lze nalézt v následující tabulce. Řádek přitom udává horizont a typ místa.

	b0	b1	b2
A-c	3317.2	67.72	-6.64
A-h	1114.9	443.86	-6.64
A-p	970.7	505.32	-6.64
B-c	2410.1	67.72	0.35
B-h	672.5	443.86	0.35
B-p	463.7	505.32	0.35
FH-h	1355.8	443.86	-9.51
FH-p	1552.0	505.32	-9.51

Model v podstatě stejné kvality (ve smyslu vysvětlené variability odezvy) by se získal, pokud by se místo $pH(BaCl_2)$ uvažovalo $pH(H_2O)$ nebo $pH(CaCl_2)$ a místo EA pak BS, BC nebo i CEC.

Podrobnější analýzy naznačují, že v budoucnu asi půjde model vylepšit zahrnutím koncentrace konkrétní kationtů. K tomu však bude zapotřebí větší rozsah dat.

III. Model pro predikci prokaryotní diverzity s využitím proměnné horizont

Jednoduchý model, který nevyužívá proměnné BC, CES a BS by byl:

predikce $P_{sobs} = a + b * pH(H_2O)$

kde koeficienty a (absolutní člen) a b (směrnice) lze nalézt v následující tabulce. Radek přitom udává horizont a site.

	a	b
A-c	1529.4	320.24
A-h	-654.8	681.66
A-p	-589.9	685.56
B-c	1552.3	179.15
B-h	1068.4	281.44
B-p	1189.6	253.20
FH-c	-1042.5	895.97
FH-h	165.7	499.13
FH-p	-1891.7	999.15

Vysvětlená variabilita je cca 54 %.

Možná by mohlo být zajímavé, že nejmenší "b" jsou pro horizont B. Tj. v tomto horizontu je P_{sobs} nejméně citlivá na $pH(H_2O)$

IV. Model pro predikci prokaryotní diverzity s využitím BC, CEC, BS

Níže popsaný model vysvětluje cca 57% variability, což oproti modelu pouze s $pH(H_2O)$ zase není žádné zásadní zlepšení.

predikce $P_{\text{sobs}} = b_0 + b_1 * \text{pH}(\text{H}_2\text{O}) + b_2 * \text{BS}$,

	b0	b1	b2
A-c	2683.5	132.75	1.22
A-h	166.1	490.85	1.22
A-p	-453.3	648.85	1.22
B-c	2433.2	132.75	-5.91
B-h	396.1	490.85	-5.91
B-p	-328.2	648.85	-5.91
FH-h	-671.9	490.85	10.70
FH-p	-914.4	648.85	10.70

Tady možná stojí za pozornost, že u horizontu B se vliv $\text{pH}(\text{H}_2\text{O})$ a BS navzájem do určité míry vylučuje.

Poznámky:

* Místo BS lze použít BC či CEC. Rovnice by byly trochu jiné, ale kvalita modelu ve smyslu vysvětlené variability podobná.

* Místo $\text{pH}(\text{H}_2\text{O})$ lze použít $\text{pH}(\text{BaCl}_2)$ nebo $\text{pH}(\text{CaCl}_2)$. Rovnice by byly trochu jiné, ale kvalita modelu ve smyslu vysvětlené variability podobná.

* Přidávání samostatných prvků jako např. Mg^{2+} , Fe^{3+} ,... model významně nezlepšuje.

Použitá literatura:

R Core Team (2023). R : A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Zvára, K. (2008). Regrese. Matfyzpress, Praha. ISBN 978-80-7378-041-8.