## **Úvod**

Téma předložené studie vychází z výsledků dizertačního projektu autorky (Hlavatá 2017), jehož předmětem bylo zpracování rozsáhlých archeobotanických souborů z doby římské. Hlavním záměrem zmíněného projektu byla rekonstrukce vybraných aspektů ekonomických poměrů obyvatelstva v daném období, která se opírala o vyhodnocení archeobotanických dat pocházejících z různých chronologických a kulturních kontextů v regionu středního Podunají. Analyzované soubory byly získány během terénních výzkumů realizovaných zejména v poslední třetině 20. a na počátku 21. století, přičemž pro jejich získání nebyly aplikovány jednotné metody archeobotanického vzorkování. To způsobilo, že i formální informace byly velmi různorodé a použití vzorků, bylo z toho důvodu, omezené. Ke sledovaným formálním vlastnostem zkoumaného souboru patří zejména objem odebraného sedimentu, počet a prostorová stratifikace vzorků (kontext/uloženina). Tato skutečnost byla impulzem pro prozkoumání otázky, zda a do jaké míry ovlivňuje metodika odběru vzorku výpovědní hodnotu zkoumaného souboru dat a jakým způsobem je třeba při archeobotanickém vyhodnocení k těmto souborům přistupovat, aby bylo možné z nich vytěžit co nejvíce validních informací. Proto, hlavním záměrem této studie není vyhodnocení vzorků z hlediska archeobotanické interpretace, nýbrž z hlediska vlivu jejich formálních vlastností na množství zachycených rostlinných makrozbytků, resp. identifikovaných druhů.

Odběr archeobotanických vzorků je provázen množstvím různých faktorů, jedná se hlavně o:

1. zvolený způsob odběru vzorků během výzkumu,
2. archeologická a archeobotanická terénní dokumentace, rozpoznání archeologických struktur a (post)depozičních procesů, uskladnění vzorků před extrakcí,
3. proplavování, mikroskopická analýza,
4. konstrukce vhodné databáze a správný výběr dat,
5. využití statistických metod a interpretace jejích výsledků.

Data, se kterými archeobotanici (i archeozoologové) většinou pracují, nejsou ideální a v porovnání s jinými archeologickými prameny více podléhají redukci již během terénního výzkumu. Neznamená to však, že tyto data jsou vesměs nevyužitelná, navzdory tomu mají velký potenciál a nenahraditelnou informační hodnotu. Za důležité však považujeme poznat všechna, nebo co možná nejvíce jejich omezení, s cílem lépe interpretovat minulou skutečnost, kterou data odrážejí. Pomocí statistických analýz a modelů bude popsán vliv zvolené metody vzorkování na makrozbytkovou skladbu archeobotanických vzorků a tudíž i na archeobotanickou či archeologickou interpretaci souborů. Tento faktor bývá totiž během realizace terénních výzkumů často – ať již vědomě či okolnostmi vynuceně – opomíjen.

## **Základní soubor**

Zájmovým regionem studie je území severně od Dunaje, omezené převedším na jihozápadní Slovensko, jižní Moravu a střední Čechy (Mapa 1). Základní vymezení je dané existencí archeobotanického materiálu, který tvořil výchozí databázi dizertačního projektu (Hlavatá 2017, 19–21, 23–26).

Z hlediska chronologie soubor náleží do období od zlomu letopočtu do konce 4. století, resp. na začátek 5. století n. l., což v chronologii daného regionu odpovídá době římské až počátku stěhování národů. Jelikož byly informace o podrobnější chronologii kontextů, ze kterých pochází archeobotanické vzorky, často nejednoznačné či vágní, byla bližší datace do jednotlivých stupňů vyloučena z analýz (Hlavatá 2017, 21, 22, 60).

Celkový počet lokalit, ze kterých pocházely archeobotanické vzorky, byl 54 (podle polohy lokality), resp. 84 (podle datování). Základní soubor tvořilo 1000 archeobotanických vzorků, ve kterých bylo dohromady identifikováno 70032 rostlinných makrozbytků (bez uhlíků), čítající 309 botanických druhů (Hlavatá 2017, 63, 64, Príloha Mapa 1–3, Katalóg). Z tohoto počtu vzorků byl pouze u 686 zaznamenán objem (celkem 13757,8 litrů odebraného sedimentu). Případy bez záznamu o objemu vzorku byly z analýz vyloučeny a počet lokalit tak klesl na 25, resp. 47 (Mapa 1). Rovněž i počet nalezených makrozbytků klesl na 36460 a 238 botanických druhů.

Archeologické lokality jmenovitě neuvádíme a nespecifikujeme. Celý soubor je zde jako modelový, tudíž z hlediska jeho analýzy nejsou důležité konkrétní lokalizace, ale vlastnosti vzorků z nich pocházejících.

## **Archeobotanické vzorky**

Na rozdíl od metodiky odběru vzorků doporučované v archeobotanické literatuře (srov. Barker 1975; Hajnalová 2012, 33; M. Jones 1991, 53–62; Klečka 1934, 98 Kočár, Dreslerová 2010, 205; Lee 2012; Orton 2000, Pearsall 2000, 66–76; Smith 2014; van der Veen 1982; 1984; 1992, 21), v archeologické praxi, zejména při výzkumech staršího data, byl sediment z uloženin často odebírán subjektivně selektivním způsobem, často bez předem stanoveného teoretického či metodického konceptu. Tato situace se projevuje i ve zkoumaném souboru. Navzdory tomu se mezi analyzovanými soubory nacházelo též určité množství systematicky odebraných vzorků. Vstupní informace o původu vzorků byly tudíž značně heterogenní a mnohokrát i výslovně torzovité.

Soubor vykazuje největší variabilitu právě ve způsobu vzorkování a extrakce rostlinného materiálu z odebraných vzorků. Ty pocházejí z různorodých typů archeologických výzkumů. Zároveň se projevuje i rozdílnost v celkové archeologické metodě v rámci jediného výzkumu (např. různé sezóny výzkumu v jedné poloze, jednotlivé sondy, různé výzkumné záměry). Výpovědní hodnotu ovlivňuje též nekonstantní objem jednotlivých vzorků, resp. objem vůbec. Samostatnou kategorií je vyzdvihování ojedinělých volně viditelných nálezů semen (pro přímou analýzu, většinou bez proplavování, tedy bez sedimentu). Pro archeobotanické analýzy, jako například výpočet hustoty rostlinných makrozbytků, jsou takto odebrané vzorky velmi těžko použitelné. Nejde-li o celou navzorkovanou výplň zkoumané uloženiny, hovoříme o vzorku sedimentu, nikoli o vzorku archeologické uloženiny (to je možné pouze v případě totálního vzorkování). Nelze tedy zaměňovat pojmy „objem vzorku/sedimentu“ a „objem uloženiny“.V případě absence informace o objemu vzorku/sedimentu je tento faktor z analýzy nutno vyloučit a případně nahradit jiným ukazatelem, například celkovým počtem makrozbytků a/nebo například indexem tzv. čistoty vzorku (sample purity, např. Simpsonův index D, citováno podle Wallace et al. 2019, 452). Avšak jak uvádí M. Wallace et al. (2019, 452), počet makrozbytků závisí nejen na hustotě vzorku, ale i na velkosti vzorku.

V rámci souboru byly archeobotanické vzorky odebrané systematicky (intervalově) na čtyřech lokalitách. Kombinovaným způsobem – systematicky i subjektivně (Hlavatá 2017, 28, 29) byly ovzorkované dvě lokality. V případě některých lokalit, ze kterých byly vzorky získány v neproplaveném/nezpracovaném stavu (Hlavatá 2017, 30), byly jedinou informací o způsobu odběru vzorků popisné informace ze štítků. Štítky většinou obsahovaly jenom jednoduchou informaci o čísle objektu, v některých případech o vrstvě, jindy jen pořadové číslo odebraného vzorku bez dalších informací. Takový způsob popisu vzorků, kdy s archeobotanickým materiálem není dodán ani podrobný přírůstkový katalog vzorků, ztěžuje analýzu a interpretaci archeobotanického materiálu.

## **Vstupní archeologické informace**

Dalším problémem při zpracování zkoumaného souboru byl nedostatek vstupních archeologických informací. Nebylo možné rozdělit, resp. porovnávat výsledky archeobotanických analýz podle typů objektů a uloženin. Krom toho se projevil zřejmý metodologický a terminologický rozdíl mezi vzorky ze slovenských, moravských a českých lokalit.

Nejnižší úroveň archeologické informace představoval neurčený kontext/uloženina, kdy byla popisná informace velmi vágně uvedena jako „objekt“, „vrstva“ nebo „sektor“ bez bližšího popisu. Na nejvyšší úrovni se nacházelo určení kontextu/uloženiny podle čísla, typu, funkce a části (např. objekt č. 1, jáma, zásobní jáma, západní část/polovina apod.), následovalo označení a specifikace uloženiny (např. K8, destrukční vrstva apod.) a zároveň byl archeologický objekt situován prostorově v rámci plochy/sondy výzkumu (např. plocha P1, sonda IV, sektor 7/b apod.). Rovněž skromné byly i prostorové informace o samotných vzorcích jako hloubka vyzdvižení, část objektu/kontextu apod.

## **Selekce, standardizace a úprava dat**

Do analýzy byly zařazeny všechny získané zuhelnatěné rostlinné makrozbytky, kromě nálezů uhlíků a dřev. Pod pojmem „jeden rostlinný makrozbytek“ (RMZ) se rozumí každé celé zrno a/nebo semeno (sensu lato), pokud si zachovalo celistvý tvar, i v případě že bylo velmi destruované. Zlomky byly přepočítané standardním způsobem (Hlavatá 2017, 46–48), tak aby je bylo možné započítat do celkových součtů makrozbytků (z NISP na MNI). Fragmenty slámy, rákosu a zuhelnatěné organické hmoty byly ponechány v jednotkách fragmentů (NISP). Pod pojmem „druh“ se rozumí botanický druh (taxon).

Analyzovaná data byla upravena do čtyř datových tabulek m × n, kdy m reprezentuje pozorování (vzorky, objekty/uloženiny, lokality) a n představuje proměnné (druhy, RMZ, objem, hustota).

Pro statistické analýzy se archeobotanická data obvykle upravují. Nejčastěji se eliminují vzorky s malým počtem nálezů a botanické druhy (taxony), které se vyskytují v nízkém počtu vzorků (srov. Bogaard 2004; van der Veen 1992; souhrnně Reed 2016, 209nn, s další literaturou). Většinou se použijí jenom ty vzorky, které obsahují 10 a více nálezů a  taxony, které se objevují ve třech a více vzorcích, případně se použijí vzorky, které obsahují minimálně 50/100 nálezů semen plodin a minimálně 10 nálezů semen plevelů (v smyslu Bogaard 2004, 60–64, s další literaturou; Hajnalová 2012, 122; Hajnalová/Varsik 2010, 210; G. Jones 1991, 67–69; Látková 2017; van der Veen 1992). M. Kuna a kolektiv argumentují (2013, 130), že takovou úpravou archeobotanických dat přicházíme o velké spektrum nálezů. Stručně řečeno, standardizace dat upravuje datový soubor ve prospěch bohatých/bohatších vzorků, chudší vzorky bývají z analýzy vyloučeny, přičemž mnohokrát se může jednat o významově hodnotné nálezy a/nebo informace. Větší množství chudších vzorků může být reprezentativnější než malé množství bohatých vzorků (Kuna et al. 2013, 130; Orton 2000, 167). Taková forma zpracování datového souboru má větší opodstatnění vzhledem k ucelenější interpretaci archeobotanických dat a může odhalit, kromě jiného, tafonomické procesy či rozdíly v managementu pěstovaných plodin (Kuna et al. 2013, 130).

Spojování vzorků podle samostatných vzorkovaných elementů (sensu stricto Lee 2012, 650 – elements/units of population; srov. Orton 2000, 30–39, 165–168), nebo stratigraficky a kontextuálně nerozlišitelných vzorků, je pro další archeobotanické analýzy potřebné, jelikož tyto netvoří samostatné entity (G. Jones 1991, 67–69; Smith 2014, 182, s další literaturou; srov. Kuna et al. 2013; Látková 2017). Takto však bylo možné spojit jenom pět vzorků, které obsahovaly dostatek archeologických informací a zároveň pocházely ze stejné uloženiny, anebo její části.

Pokud se jedná o selekci objektů a jejich typů, tyto jsme zahrnuli do analýzy tak aby odrážely skutečný stav v našem souboru. To znamená, že byly analyzovány vzorky z určených uloženin nebo objektů, stejně jako vzorky se všeobecním určením (např. „vrstva“ bez konkrétnějšího popisu apod.).

Vzhledem k řešené problematice byla pro analýzu použita neredukovaná data, tedy všechny vzorky, které disponovaly informací o objemu odebraného sedimentu. Analyzovány byly rovněž i vzorky, které neobsahovaly žádné nálezy makrozbytků, anebo obsahovaly jen malé množství RMZ (např. 1 ks apod.). Cílem bylo postřehnout v rámci materiálu případné elementy, které by se „tradiční“ standardizací do statistického hodnocení nedostaly, anebo by mohly uniknout při interpretaci výsledků. Vyloučeny z analýzy byly pouze tři vzorky hromadných nálezů, které vykazovaly velmi odlehlé hodnoty (více jak 4000 makrozbytků ve vzorku s malým objemem).

**Statistické modely**

Před tvorbou statistických modelů byla data podrobena deskriptivní statistice, s cílem odhalení jejich základních vlastností. Pokud je absolutní hodnota šikmosti a špičatosti (vydělena dvěma standardními odchylkami) vyšší než 1, tak jsou data na hladině významnosti 0,5 %  statisticky významná (Field/Miles/Field 2012, 17). Jak je vidět (Tab. 1, skew.2SE, kurt.2SE), hodnoty jsou vysoko nad 1 a tedy šikmost i špičatost jasně dominují. Dalšími indikátory jsou vysoké rozdíly mezi průměry a rozptyly, resp. standardními odchylkami (Tab. 1, var, std.dev). Především u RMZ tento rozdíl poukazuje na tzv. nadměrný rozptyl (*overdispersion*) – častý jev hlavně v ekologických a environmetálních datech, který může mít rozdílné příčiny (O’Hara/Kotze 2010; Zuur et al. 2009, 270-271; Martin et al. 2005; souhrnně Hilbe 2011, kapitola 4). Stejně tak hodnota *Shapiro-Wilk*testu (Tab. 1, Shapiro-Wilk ) a její p-value potvrzují „ne-normální“ distribuci pro druhy a RMZ.

Hodnoty makrozbytků (RMZ) vykazovaly největší rozptyl a často způsobovaly nadměrný nebo nedostatečný rozptyl (underdispersion). To se týká dat i po odstranění odlehlých hodnot (*outliers*). Řešením by bylo odstranit přibližně 10 % hodnot shora i zdola, nebo data transformovat. Tím by však došlo ke ztrátě velké části informací, rovněž by byly vyloučeny nulové hodnoty. Vzhledem k faktu, že informace o absenci makrozbytků ve vzorku je relevantní, nebyla data redukována. Podstatnou částí postupu je totiž sledovat surová data v původním stavu a s nejmenšími možnými zásahy, vzhledem k předpokladu, že takováto data věrněji odrážejí reálný způsob vzorkování.

Graf 1 zobrazuje surová data, v nichž bychom normální distribuci hledali marně. Diagonálně jsou zobrazeny histogramy a jejich distribuce.[[1]](#footnote-1) Data objemů jednotlivých vzorků mají rysy bimodální distribuce, zatímco data druhů jsou viditelně zešikmené doleva. Distribuce RMZ je zas hodně poznačena větším množstvím nízkých a nulových hodnot, v kombinaci s extrémně vysokými hodnotami (srov. Tab. 1). Ty roztahují osu X a stlačují hodnoty do jednoho intervalu (bin).[[2]](#footnote-2) V levé dolní části grafu 1 jsou zobrazeny korelační diagramy (scatterplots) spolu s trendovou spojnicí.[[3]](#footnote-3) Na diagramech objemu s RMZ (Graf 1, X) a objemu s druhy (Graf 1, Y) nelze pozorovat téměř žádnou závislost. Naproti tomu vztah mezi RMZ a druhy (Graf 1, Z), odhliadnuc od odľhahlých hodnôt, vykazuje na první pohled závislost vysokou. Pravá horní část grafu 1 zobrazuje míry asociace podle Spearmanova korelačního koeficientu.[[4]](#footnote-4) Zdá se být logické, že mezi RMZ a druhy (Graf 1, Zx) je vysoká pozitivní korelace, tedy když stoupá počet extrahovaných RMZ, stoupá i počet identifikovaných druhů. Toto tvrzení ovšem není kauzálně podmíněné. Jako kontraintuitivní se jeví téměř nulová korelace mezi objemem a RMZ/druhy (Graf 1, XY), což koresponduje i s výsledky v korelačních diagramech. To by naznačovalo, že množství nalezených RMZ a druhů nezávisí na množství odebraného sedimentu. Tomuto problému bude pozornost věnována níže.

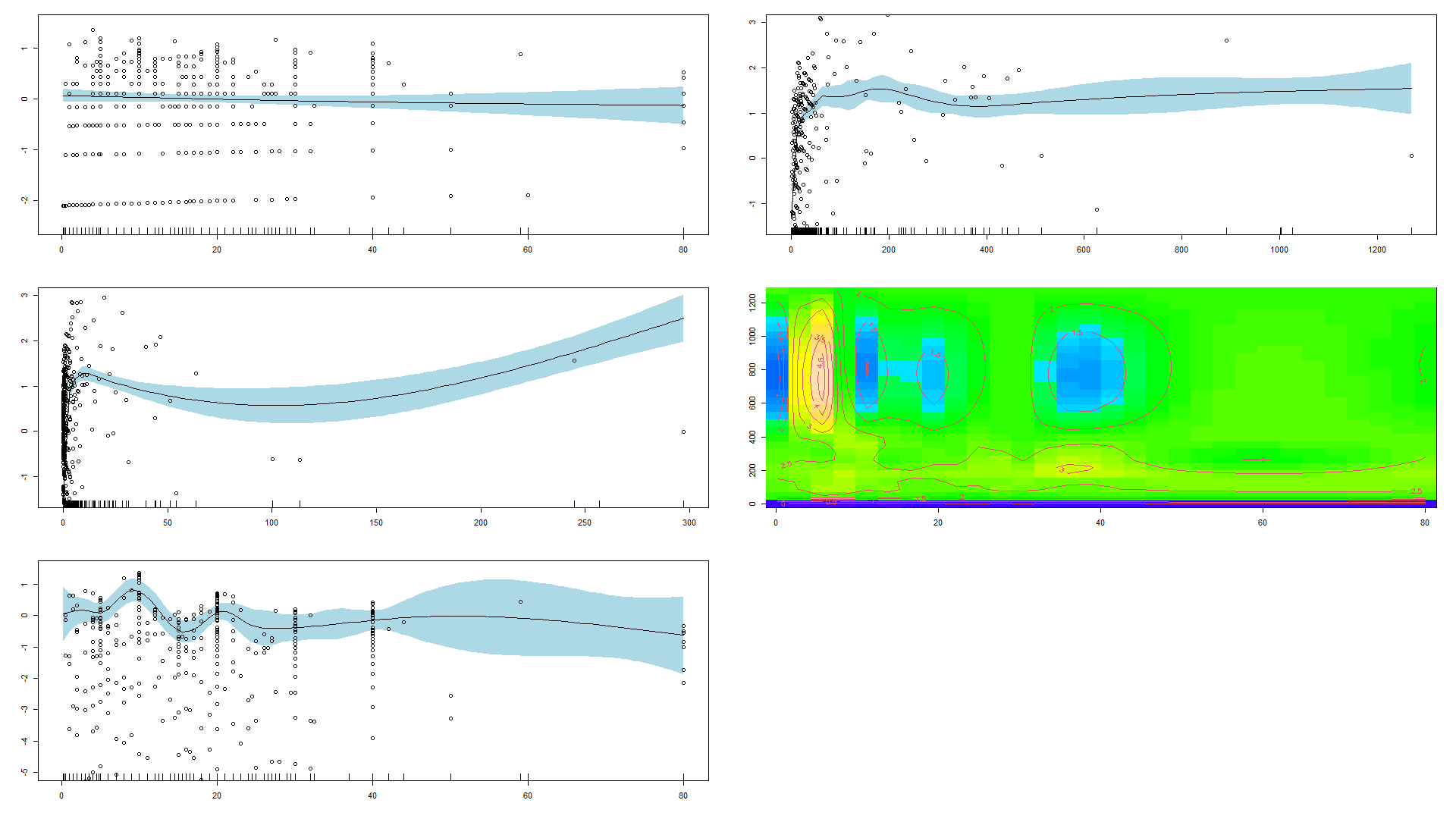
Najprv sme potrebovali nájsť matematický model (formu), ktorý by najlepšie reprezentoval spracovávané dáta (matice), aby mohli byť ďalej štatisticky vyhodnotené (Orton 2000, 9-10; Orton 1980, 20-21). Keďže skúmané premenné (response variables) nespĺňali predpoklady normálnej distribúcie, využili sme tzv. negatívne-binomálne rozdelenie (*negative binomial distribution*), ktoré sa najviac približovalo naším dátam. Ide vlastne o rozšírenie Poissonovho rozdelenia a obsahuje extra parameter pre nadmerný rozptyl (Hilbe 2011, 253) a často sa použiva  ekologických  enviromentálnych štúdiách (napr. Zuur et al. 2009, 225-230).

Pri modelovaní sme primárne použili tzv. Generalized additive models (GAM)[[5]](#footnote-5), kde v podstate o rošírené Generalized linear models (GLM), avšak nepredpokladajú normálnu distribúciu ani linearitu, resp. parametrickú formu dát (Crawley 2007, 611). Hlavný rozdiel oproti lineárnym modelom je, že GAM namiesto *β* parametrov využívajú tzv. neparametrické smooth funkcie, v zjednodušenej forme zapísané ako Yi = α + f (Xi) + εi kde εi ∼ N(0,σ 2), pričom Yi je vysveľovaná (response) premenná a f (Xi) je smooth funkcia (krivka) pre vysvetľujúcu premmenú (explanatory). K GAM pozri S. Wood (2017),  Alain. F. Zuur et. al. (200?), M.J Crawley (2007) a takisto množstvo užitočných informácii a príkladpv je možné dohľadať na internete (<https://rdrr.io/cran/mgcv/> - toto pd čiaru). GLM[[6]](#footnote-6) sme použili pre tvorbu podporných, resp. porovnácich modelov v prípade, keď skúmaný vzťah vykazoval známky linearity (závislosti).

Dohromady sme vytvoili 13 jednoduchých modelov, pričom každý modul je reprezentovaný jedným diagramom s príslúchajúcimi výslednými hodnotami v tabuľke 2. Dôvody, pre pomerne vysoký počet modelov, boli dva: a) chceli sme do určitej miery ilustrovať analytický postup a pozrieť sa jednotilivé premenné samostatne; b) RMZ kvôli povahe dát často spôsobovali v komibnácii s inými premenným nadmerný alebo naopak nedostatočný rozptyl. Tam kde to bolo možné a zmysluplné, sme do modelu zaviedli viac premenných. Pri opise často zamieňame slová vplyv a efekt, ktoré je do istej miery možné brať ako synonymá, pričom efekt odkazuje skôr na štatistickú stránku modelu, zatiaľ čo vplyv na archeobotanickú. Výsledné hodnoty modelov sú predstavené v tabuľke 2.

V případe modelů, které byli počítaní pomocí tzv. tenzorového produktu interakce[[7]](#footnote-7) (vícebarevné diagramy) sme znížili penalizaci pomocí gamma funkce z (v literatuře) odporoučeny hodnoty 1.4 na vychodzí hodnotu 1, aby sme a mohli se podívat na menší nuanse dat v modelu (citácia).

Modely 1 až 5



Při analýze modelov 1 až 5 byla použita matice m × n, kdy m = počet vzorků (683) a n = proměnné: objem vzorku, počet identifikovaných makrozbytků (RMZ) ve vzorku, počet botanických druhů nalezených ve vzorku, hustota makrozbytků v litru vzorku.

**Model 1** (diagram 1, A)zobrazuje vztah anebo efekt botanických druhů a objemem vzorků, který ale není ze štastistického hladiska významný, jeho efekt na množství druhů je zanedbatelný a vysvětluje pouze 0,24% celkové variability. Abychom to ověřili, vzorky sme ještě rozdelili do čtyř objemových kategorii podle litrů (A: 1 - 9, B: 10 - 29, C: 30 – 40, D: 41 – 80). Jako statisticky nejsignifikantnejší a s největším efektem (edf 2,22) se projevila kategorie A, teda vzorky s nejmeším množstvím litrů. Zvyšné kateogorie neměli v podstate žádny efekt. Z hlediska modelu to značí, že z hlediska všech vzorků objem nehrá téměř žádnou rolu, kolik druhů se ve vzorce najde, avšak z hlediska objemových kategórii je viditelný efekt menších vzorků na zachycenou druhovou skladbu. Samozřejmně archeobotanická, resp. archeologická interpretace nemusí byt tak přímočára a pozěji se k tomu ještě vrátime.

**V modelu 2** (diagramu 1 B) sme se podívali na vztah botanických druhů a počtů RMZ. Jak už bylo zmíněno, RMZ sou z hlediska jejich hodnot problematické. V tomto případě se to projevilo především nedostatečnou disperzí, což bývá často způsobeno shlukováním nízkych hodnot anebo přizpůsobením modelu vícerým extrémním hodnotám (Hilbe 2014, 136; Zuur et al. 2009, 225). V našem případě můžeme uvažovat o kombinaci těchto faktorů. To ostaně je dobře zobrazeno na diagramu, kde počet druhů vykazuje vztah především s nižšími počty makrozvyšků a někde při množství 50 až 100 makrozvyškú je možné vidět výrazný zlom, spůsobený odlehlými hodnotami. Prváve tito hodnoty spůsobili tento výrazný efekt a taktéž štatisticky významný efekt. Při vyšších počtech makrozvyšků se pak jejich ich vztah k druhům jeví jako „plochý“ a nemůžeme tak hovořit o jejich nejaké vzájemné závislosti v zmyslu, žeby větší objemy přinášeli i bohatší archeobotanické soubory.

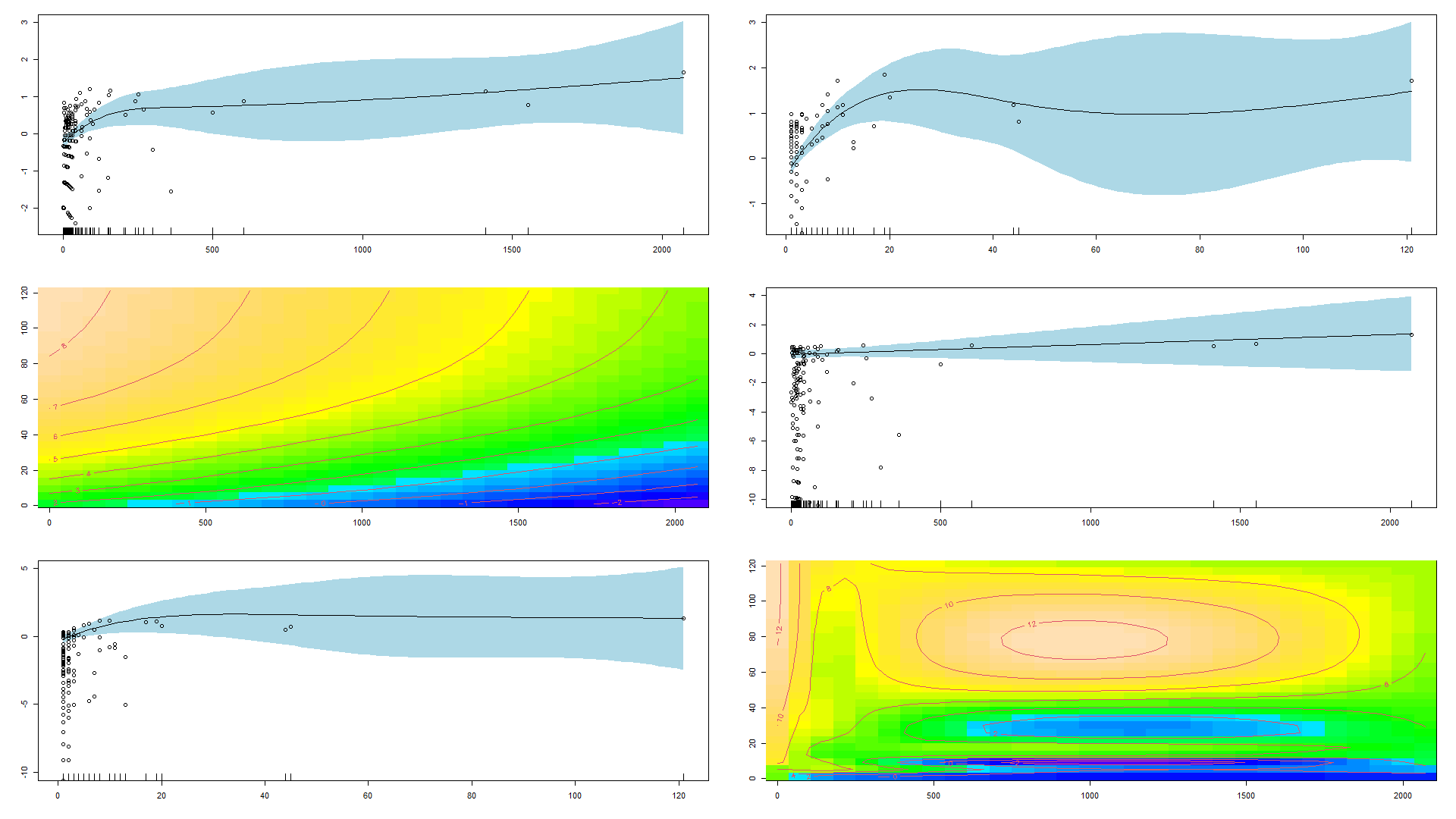
**Model 3** (diagram sme zde uvedli především pro ilustraci vzahu počtem druhů a hustotou, a především jednak kvůli charakteru hustoty, která se jednak počíta jako poměr RMZ a objemu a jednak kvůli jeji častýmu využití v archeobotanických studiích. Když si porovnáme modely 2 a 3 vidíme značnou podobnost jak po stránce tvaru funkce (vykazuje velmi podobné chováni jako RMZ), tak z hlediska hodnot (tab. 2). Hustota je teda už implicitne obsáhnutá v RMZ a proto ji dál nebudume používat.

**Model 4** zobrazuje velmi podobný stejný vztah jako hustota (RMZ) avšak s tím rozdílem, že tady je počítan pomocí tzv. tenzorového produktu interakce, který modeluje vzájemnou „čistou“ interakci medzi skoumánymi proměnnými, kdy nebere v úvahu hlavní efekt jednotlivých proměnných. Tito proměnné taktéž nemusí být zaznamenné v stejných jednotkách. Navyše je diagram lépe čitelný, protože osy x a y jsou zobrazený v původních jednotkách. Jak je vidět, model 4 kondenzuje výsledky modelů 2 a 3 , kdy největší efekt pozorujeme při nižších objemech a vyšších počtů makrozvyšků. O něco menší efekt je pak viditelný v spodní části diagramu, který odráži skutečnost, kdy se především odebrali větší vzorky, v kterých se identifikovalo i větší množství různych druhů.

V **modelu 5** sme blíže analyzovali vztah počtů makrozvyšků vzhledem k objemům vzorků. Tenhle model by pro změnu poznačený nadrozměrnou disperzí kvůli spomínanym vlastnostnem RMZ. Vypočítaná funkce (křivka) je však staticky významná a doležená vysokým efektem objemu na množství získaných markozvyšků. Tento efekt, neboli vztah je však značne nelineárny. To znamená, že i v tomhle případe nemůžeme jednoduše predikovat, že s vyššími objemy získame i vyšší množství RMZ a naopak. Model přitom vysvětluje jen pouhých 6,5% variability a teda objem sám o sobě nestačí na to, aby odhalil další skryté struktury v datech. O něco lepší informace sme získali při modelováni pomocí kategorií objemů. Vysoký efekt (edf 6.3 ) a statistickou významnost projevila kategorie B (10 -29 litrů), která má ze statistického hlediska nejväčší potenciál při získávaní většího počtu RMZ.

V **modelech 1 až 5** sme si přiblížili vzájemné vztahy medzi objemem, počtem botanických druhů, RMZ a hustoty z hlediska všech vzorek, kde se tito údaje dali sledovat. Po proskoumáni vztahu RMZ a hustoty sme hustotu nakonec vyloučili z dalších analýz, protože v podstate kopíruje chováni RMZ a mohlo by hrozit, že bude negatívne ovplvňovat modely. Namísto sme využili při skoumáni jejich vzájemnej vztahu funkci tenzorový produkt interakce. Co se týče vztahu objemu k počtům druhů a RMZ, nemá objem, z hlediska celého souboru, téměř žádny vliv na ich množství. Z hlediska kategorií objemu se však ukazuje, že menší (do 10 litrů) a středně velký vzorky (do 29 litrů) jsou ze statistického hladiska signifikantješí a vykazují vyšší efekt, neboli poskytují o něco vyšší šanci k tomu, abychom z nich získali i větší a taky reprezentatívnejši množství nálezu. Samozřejmně toto tvrzení nemůže byt postaveno pouze na množství odebraného sedimentu ve vzorce a paušalizováno na jiné vzorky, objekty či lokality. V tomto případe ilustruje jen malou část problematiky vzorkováni a spíše představuje první stupeň k dalším analýzam a interpretacím.

**Modely 6 až 11**

V **modelech 6 až 11** sme skoumali stejné typy vztahů jako v předchozích modelech, ale tentokrát z hlediska objektů/kontextů. To přináší o něco intuitívnejší pohled na vplyv vzorkování archeologických situací. Taktéž sme mohli sledovat uvedné vztahy i pomocí počtů vzorků. Předtím je však nutno uvést, že až z 80% objektů se vyzvedlo nejvíce 2 vzorky. Z archeobotanického hlediska se jedná málo repreznatívní soubor, ale ze statistického hladiska se na to můžeme dívat jako na relatívne „rovnomerně“ ovzorkovaný soubor i když do velké míry poznačený nejednotnou velikosti odebraných litrů sedimentu z objektů.

**Model 7** (diagram – A) sleduje vztah medzi objemem a počtem druhů. Ti se soustředí především do levé části grafu v oblasti „menších“ objemů, přičemž funkce (křivka) nejvíce roste přibližne do 250 litrů a pak se stáčí k odelhlým hodnotám. Podle údajů v tabulce 1 (m6) má objem celkem jasný efekt na počet druhů v objektě, především při nižších objemech (do cca 250 litrů).

Na **modeli** 8 (diagram - B)vidíme ještě výraznejší efekt počtů vzorku na počet druhů, kde funkce (křivka) dosahuje vrchol někde u 20 vzorek a pak zde opět úradují odlehlé hodnoty. Počty vzorků vysvětlují dvakrát tolik variabilty (21,1%) jako objem a taky jde o lepší model (tabulka – AIC). Větší množství odebraných vzorků má teda jednozančne signifikantejší přínos z hlediska získaní bohatší skladby druhů v objektech jako objem samotný.

**Model 8** představuje vzájemnou interakci mezi objemem a počtem vzorků, která však není vůbec statisticky signifikantná (tab. M8), jekoliž hlavní efekt zde představují opět počty vzorků (edf 3,03) „na úkor“ objemu (edf 1)[[8]](#footnote-8). To je vidět i v procentu vysvětlené variablity, která je téměř stejná jako v modelu 8, rovněž jako v diagramu C, kde počet druhů silně inklinuje k vyšším počtům vzorků.

**Model 9** řeší vztah mezi RMZ a objemem a podle výsledku (tabuka 2, m9) a diagramu - D nemá objem žádny efekt na množství získaných makrozvyšků a navyše není ani statisticky signifikantný. Z hlediska souboru objektů jde o model s nejhorším výsledkem (porov. AIC hodnoty). To se shoduje s modelem 1, kde byl síce efekt objemu značný, ale do velké míry poznačený odlehlými hodnotami, přičemž šlo také o daleko nejslabší model ze všech. Je vcelku zajímavé, že objemy mají při vysvětlovaní RMZ takovu slabou výpovědní hodnotu

**Model 10** zobrazuje RMZ vysvětlované počtami vzorků a jak je vidět na diagramu – E, v jeho levé části (přibližne do 20 vzorků) múžeme sledovat nepatrný efekt, avšak na hranici signifikantnosti a, podobně jako při modeli 8, jsou výsledky poznačení nadměrným rozptylem.

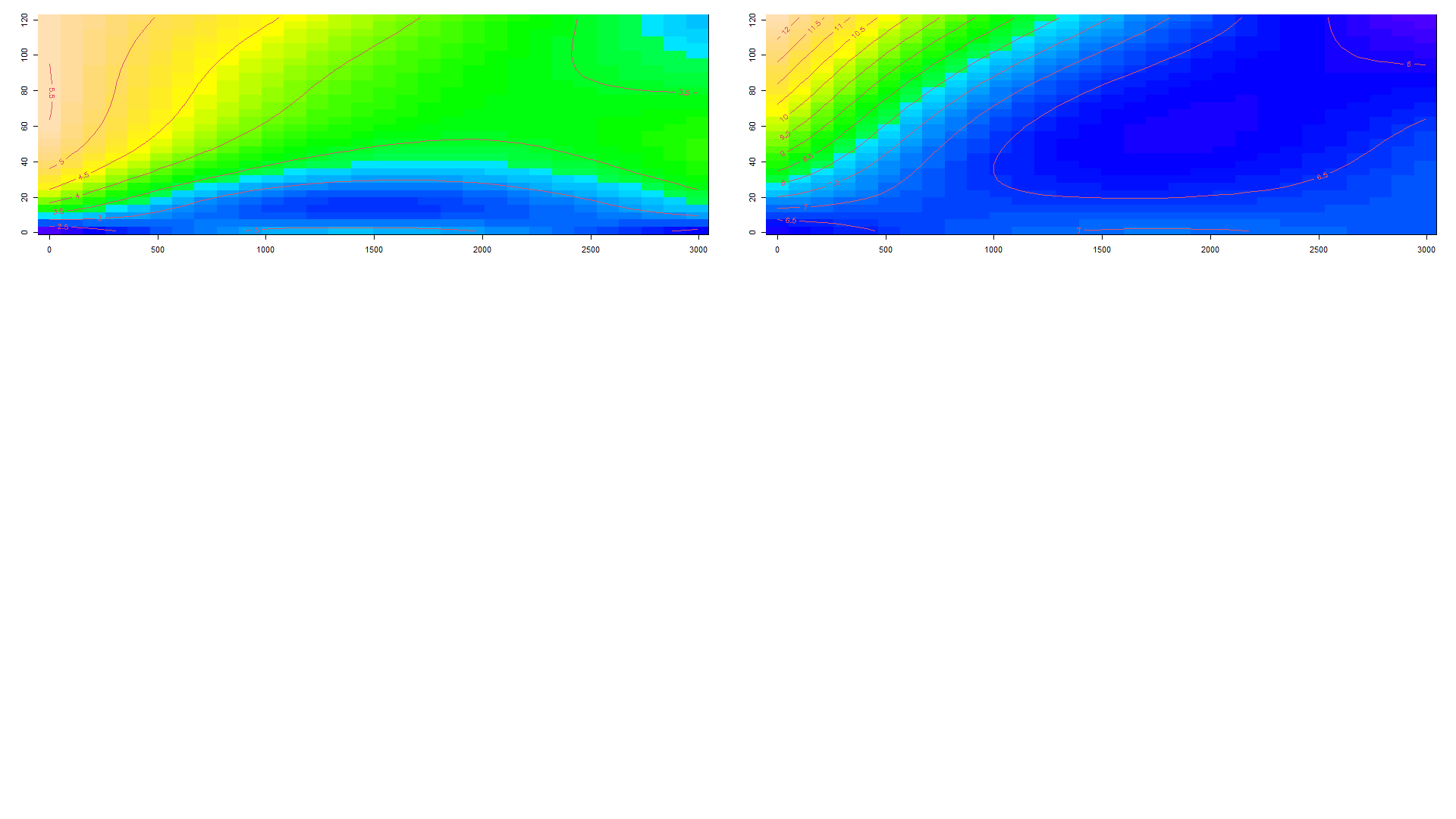
O něco zajímavější je situace u **modelu 11** (diagram 2 – F), kde je zobrazená vzájemná interakce RMZ a počtů vzorků. Na rozdíl od modelu 9 a 10 je tu vidět značný efekt (tabulka, m11) a vysokou statistickou signifikantost a z hlediska RMZ jde o nejlepší model s nejvyšším procentem vysvětlené variability. Na diagramu 2 F jstou patrné dva hlavní efekty. První v levé části poukazuje na úzky vztah RMZ k vyšším množstvím vzorků a menším objemům, zatímco druhý souvisí s vyšším objemem (ca mezi 300 až 1800) a též vyššími počtami vzorků (ca mezi 40 až 110 vzorků). Model 11 reflektuje především spůsob vzorkování objektů (viz výše), ale také pokazuje na to, že při získaní většího množství RMZ z archeologické situacie je nutno zvolit vhodnou rovnováhu mezi dostatečným počtem odebraných vzorků a jejích objemem.

**Modely 6-11** o něco víc osvětlili situaci vplyvu vzorkováni na výslední archeobotanický soubor. Především začlenení proměnné počet vzorků do modelů přineslo nové informace o tom, jak může být skladba botanických druhů ovplyvněná vyššími vzorky, resp. vice versa. Více odebraných vzorků se tady jeví jako tá lepší alternativa. Co se týče množství vyzdviženého sedimentu v rámci objektů, výsledky částečne korespondují s těmi z modelů 1 až 5 a teda objem sám o sobě není určujíci efekt pro získaní hojnějšího množstvi nálezů. Právě vhodně zvolená kombinace více menších vzorků by mohla přinést spolehlivé výsledky. Co je to však to menší vzorka? To se samozřjemně nedá rozhodnout v rámci zde prezentovaných dat a do veliké míry zo závisí od cíle výskumu. Každopádne je možné domnívat se, že například místo jedné 50 litrové vzorky je lepší vyzdvihnout pět 10 litrových, nebo deset 5 litrových vzorek prostorově pokrývajíci větší plochu skoumané archeologické jednotky. Takovéto vzorky se vždy dají v rámci analýzy spojit do jednoho elementu (in sensu Lee 2012). Bohužel, prostorové informace ohledem jednolivých vzorek, či archeologických jednotek byli nedostatečné a proto sme je nezačlenili do modelů, přičemž prostorová informace by s velikou pravdepodobností pomohla vysvětlit větší procento variability a poodhalit další skryté struktury v datech.

ORTON + DRENNAN – dvakrát viac neznamená aj dvakrát viac informácii.

s spůsobem je metoda ovplyvňováno

Nízky počet „velkoobjemných“ vzorků se v tomto případe neukazuje jako idealni přístup pro dosáhnuti takového cíle.



Při posledních dvou modelech 12 a 13 sme sledovali vztah botanických druhů (diagram 3 -A) a RMZ (diagram 3 – B) vzhledem k objemu a počtů vzorků na upravené matici m x n, kde m jsou lokality a n botanické druhy, RMZ, objemy v rámci lokality a celkové počty odebraných vzorku z lokality. V modelech sme sledovali hlavní efekty jakož i jejich interakci. Diagrami a údaje v tabulce 5 podtrhují to, co jse postupne kristalizovalo v předchozích modelech a teda, že počet vzorků opět sehráva určujíci úlohu v tom, jak reprezentatívni materiál jsme schopni získat z lokality.

výstupních dat (tab 2, m1), mají všechny prediktory vliv na výsledný model a jsou, kromě RMZ, statisticky signifikantní. Soubor je do určité míry postižen nedostatečným rozptylem (underdispersion), což bývá způsobeno shlukováním dat (Hilbe 2014, 136) či možným přizpůsobením se smooth funkce vícerým extrémním hodnotám (Zuur et al. 2009, 225). Z celkové variability modelu (81 %) RMZ vysvětlují až 76 %. Z toho vyplývá, že model by bylo možné redukovat o data RMZ a ponechat je na separátní analýzu. I přesto byla v modelu ponechána, protože jde o základní kritérium archeobotanických analýz.

Objem vzorku (graf 2, A) má určitý efekt na počet druhů ve vzorku a jejich vzájemný vztah je nelineární. Zlom nastává přibližně při 10litrových hodnotách a dále má už jen slabě stoupající tendenci, provázenou vyšší mírou nespolehlivosti odhadu funkce (interval spolehlivosti na hladině 95 % – modrá barva). Vliv velkého objemu (50–80 litrů) na vyšší počet zachycených druhů ve vzorku je neprůkazný. Samotný objem (bez ostatních prediktorů) přitom vysvětluje jen 2,2 % variability a jeho celková vypovídací hodnota tak zůstává pouze orientační. Vzhledem k extrémním hodnotám jsou makrozbytky statisticky nevýznamné, ovšem situace na grafu 2b naznačuje, že vyšší počty makrozbytků ve vzorku mohou mít mírný vliv na nižší počty druhů ve vzorku. Naopak hustota makrozbytků ve vzorku vykazuje poměrně silný vztah s počtem druhů a je statisticky významná. Na grafu 2, c křivka hustoty výrazně (téměř lineárně) stoupá až po hodnotu cca 25 RMZ na litr sedimentu. Následný výrazný zlom je způsoben přítomností extrémních hodnot. Vyšší hustota makrozbytků ve zkoumaných vzorcích není jednoznačným indikátorem většího počtu druhů. S tím korespondují výsledky grafu 2, d, kde je pomocí „ti“ funkcezobrazena interakce objemu vzorku a RMZ z hlediska počtu druhů ve vzorku. Zde se vyprofilovaly tři hlavní efekty (žlutá barva). Největší efekt je vidět při menších objemech s vysokými počty makrozbytků, druhý efekt při vyšších objemech a vyšších počtech makrozbytků a poslední efekt při vysokých objemech a nižších počtech makrozbytků. Není viditelný žádný zjevný trend nebo vzorec upřednostňující jediný prediktor.

Hlavním výsledkem modelu 1 je, že objem, makrozbytky a hustoty mají slabou vypovídací hodnotu z hlediska predikce počtu druhů v jednotlivých vzorcích. To značí, že vztah mezi počty druhů a prediktory je nelineární a nemůžeme předpokládat, že s vyšším objemem vzorku, nebo při větším množství makrozbytků bychom získali i vyšší počet druhů. Do jisté míry by tento závěr mohl reprezentovat povahu archeobotanických dat obecně. Například v případě vzorků reprezentujících zásoby plodin (ve smyslu tzv. single event contexts) existuje předpoklad, že nebudou složeny z většího množství druhů, ale budou pravděpodobněji obsahovat malou druhovou rozmanitost při velkém množství makrozbytků CITACIA

Model 2

V případě modelu 2 došlo k úpravě dat s cílem blíže prozkoumat vztahy rostlinných druhů, makrozbytků a objemu vzorků v jednotlivých objektech nebo uloženinách. Matice obsahuje 210 řádků (objekty nebo uloženiny) a 4 sloupce (druhy, RMZ, objem, počet vzorků). Tentokrát byly použity všechny proměnné společně, včetně jejich vybraných interakcí (viz níže). Hodnota gamma byla ponechána na doporučené hladině 1,14 (Wood 2017, STRANA; Zuur et al. 2009, STRANA) a byla tak snížena penalizace pro funkci smooth, za účelem postihnutí hlavního trendu.

Graf 3 zobrazuje vztah nebo efekt počtu druhů s jednotlivými proměnnými (objem, RMZ, počet vzorků, hustota, interakce objem + RMZ a objem + hustota). Model vysvětluje 85 % variability avšak hodnoty v tabulce 2, m6 naznačují, že zkoumané proměnné se na tom podílejí různou měrou. Objem (graf 3, a) je sice statisticky signifikantní, ale jeho vliv na celkový výsledek je v podstatě nulový. Jeho vztah k počtu botanických druhů v jednotlivých objektech je lineární (myšleno „plochý“ – flat). Klesající trend je způsoben především odlehlými hodnotami[[9]](#footnote-9) a s nimi stoupá i míra nespolehlivosti. Sám o sobě by objem vysvětloval přibližně 11,5 % variability s menším signifikantním a mírně nelineárním efektem a tak je celkově hodnocen jako slabý prediktor.

Počty makrozbytků v jednotlivých objektech (graf 3, b) vykazují úzký vztah k počtům druhů v nich nalezených (tab. 2, m2). To je důležitý výsledek hodnocení vzorků podle objektů. Znamená to, že počet makrozbytků nalezených v objektu nebo uloženině má vliv na celkový počet nalezených druhů v dané uloženině. U tohoto modelu jsou počty vzorků statisticky významné a jsou zodpovědné za 21,4 % variability. Na grafu 3, c sledujeme mírný efekt, kde vyšší počet vzorků může přinést vyšší počet druhů. Hustota v tomto případě není příliš signifikantní a koresponduje s dalšími pozorováními – konkrétně s interakcí makrozbytků s objemem (graf 3, XY) a počtu vzorků s objemem (graf 3, YZ). Jejich společný efekt na množství druhů je zanedbatelný a nesignifikantní. Hlavní vliv mají spíše jednotlivé proměnné (resp. hlavní smooth funkce) než jejich interakce. V obou případech se s vyšším objemem postupně vytrácí jeho efekt na množství druhů nebo makrozbytků. Nejlépe je to sledovatelné na grafu 3, f, kde množství druhů inklinuje k počtům odebraných vzorků (žlutá) a nikoliv k jejich objemu. V konečném důsledku to naznačuje, že v objektech nebo uloženinách v zkoumaném souboru, způsobilo odebrání většího množství vzorků (ne jenom vzorků s větším objemem) nalezení většího množství makrozbytků o více botanických druzích. To může být způsobeno dvěma faktory. Prvním je velký rozdíl ve vzorkování mezi jednotlivými lokalitami, přičemž převážně byly odebírány vzorky s nižšími objemy než ty velkoobjemové (50–80 litrů na vzorek). Z některých objektů nebo uloženin byl odebrán malý počet velkých vzorků (někdy dokonce pouze jeden vzorek) a nikoli více menších vzorků. Ve druhém faktoru může jít o rozdíl mezi vzorkovanými uloženinami. Uloženiny, ze kterých pocházelo nejvíce velkoobjemových vzorků, se mohli archeologicky lišit od uloženin, které byly vzorkovány menšími vzorky.

Modely 3, 4, 5 a 6

Pro model 3 (graf 4, a) byla použita stejná matice jako u modelu 1. Předmětem zkoumání byl vztah mezi objemem a počtem makrozbytků v jednotlivých vzorcích. Statisticky významný objem má efekt na počet makrozbytků ve vzorku a podtrhuje jejich nelineární vztah. Z grafu je možno vyčíst tři hlavní vrcholy – menší vzorky (cca 5 litrů), cca 10litrové a cca 20litrové vzorky. U posledně jmenovaných se projevuje největší efekt objemu. S vyššími objemy stoupá i míra nespolehlivosti. Podobně jako u počtu druhů, ani tady nelze předpokládat, že s vyšším objemem vzorku bude stoupat i množství makrozbytků.

Modely 4, 5 a 6 vycházely ze stejné matice jako model 2. Hodnota gamma byla snížena na 0,1 a zvýšila se tak „wiggliness“ funkce, pro možnost graficky zachytit co nejmenší nuance modelu, které by jinak unikly pozornosti. Model 4 (graf 4, b; tab. 2) zachycuje vztah objemu s makrozbytky v jednotlivých objektech, dokládá poměrně vysoký efekt a nelineární vztah. Křivka makrozbytků je nejvíc ovlivněna menšími objemy, odlehlé hodnoty (vysoké objemy) zvyšují míru nespolehlivosti. Podobná situace je i u modelu 5 (graf 4, c), který zohledňuje počet makrozbytků z hlediska počtu vzorků odebraných z jednotlivých objektů. Model je však ovlivněn tím, že z celkového počtu 209 objektů nebo uloženin, byl ze 131 odebrán pouze jediný vzorek. Navzdory, nebo právě kvůli tomu model potvrzuje silnou nelinearitu.

Poslední model (graf 4, d) prezentuje interakci objemu s počtem odebraných vzorků vzhledem k makrozbytkům. Tato interakce má signifikantní efekt na makrozbytky a potvrzuje výsledky modelů 4 a 5. Na první pohled vyšší počty vzorků odebraných z objektu nemají žádný vliv na množství RMZ (modrá zóna). To však zapříčiňuje zmíněný vysoký počet objektů, resp. uloženin s jediným odebraným vzorkem.

1. pomocí jádrového odhadu hustot – kernel density estimation, k tomu např. Baxter 2015, 37-49 [↑](#footnote-ref-1)
2. Při vytvoření jednoduchého lineárního modelu tato data nesplňují ani další předpoklady jakými jsou homogenita a zčásti i nezávislost [↑](#footnote-ref-2)
3. LOESS křivka, k tomu např. Zuur et al. 2009, 39-42 [↑](#footnote-ref-3)
4. Velmi podobné výsledky dosahuje i použití koeficientu Kendallova tau. [↑](#footnote-ref-4)
5. Za použitia funkcií z balíka mgcv (Wood 2017). [↑](#footnote-ref-5)
6. za použitia funkcií z balíka MASS (Venables, Ripley 2002). [↑](#footnote-ref-6)
7. Jedná se o funkci „tensor product smooth interaction“ (srov. Wood 2017, 187, 326; https://rdrr.io/cran/mgcv/man/gam.html). [↑](#footnote-ref-7)
8. Není zahrnuto v tabulce 2 [↑](#footnote-ref-8)
9. Pro porovnání, u modelování pomocí tzv. generalized linear models s využitím negativní binominální regrese (funkce glm.nb z balíku MASS), při odebrání o jednu jednotku objemu navíc existuje šance, že počet druhů se zvýší průměrně o 1%. [↑](#footnote-ref-9)