



# VUMOP

VÝZKUMNÝ ÚSTAV MONITORINGU  
A OCHRANY PŮDY, v.v.i.

Jan Skála

**Představení interaktivního znalostního portálu  
SoilPass pro informace o hygienickém stavu  
zemědělských půd v ČR  
I. část – prediktivní modely**

# Motivace tvorby nástroje

- z hlediska obsahu rizikových látek v půdě je nejlepším opatřením pro budoucnost zabránit dalším nadměrným vstupům rizikových látek do půdy
- zároveň současná odpadová politika v Evropském prostoru podporuje minimalizaci vzniku odpadů, což s rostoucím deficitem přirozených materiálových zdrojů v agro-ekosystémech vede ke změnám v oblasti opětovného využívání alternativních zdrojů – kaly, komposty, sedimenty
- ČR má nebývalé množství informací o hygienickém stavu půdy i komplexní legislativní nástroje ochrany zemědělské půdy - chybí volně dostupná datová základna pro komplexní informování vlastníků, hospodařících subjektů, rozhodovacích orgánů ochrany půdy a aktérů managementu půdy



# SS03010364 Systém na podporu rozhodování při hodnocení kvality půdy z hlediska obsahu rizikových látek v zemědělských půdách České republiky

Program Prostředí pro život 3  
2021 - 2023

## Program **Prostředí pro život**

Výzkum byl podpořen projektem č. SS03010364 financovaným se státní podporou Technologické agentury ČR a Ministerstva životního prostředí v programu Prostředí pro život.



- cílem projektu je poskytnout **otevřený nástroj** pro nezávislou metodickou a informační podporu k hodnocení hygienického stavu půd – tj. nabídnout zdroj věrohodných nezávislých informací o obsazích rizikových látek jako podklad pro přiměřenou ochranu zemědělské půdy v návaznosti na existující legislativní rámec ochrany ZPF

# Informační rozcestník – SoilPAss

- k naplnění cíle projektu byl spuštěn v roce 2024 interaktivní nástroj „SoilPAss = „Soil Pollution Assesment“ (<https://soilpass.vumop.cz/>), který bude součástí geoportálu Výzkumného ústavu monitoringu a ochrany půdy - <https://geoportal.vumop.cz/>
- systém zahrnuje interaktivní mapovou aplikaci, která může sloužit nejen k identifikaci oblastí se zvýšenými obsahy, ale také oblastí nezatížených či s deficitem biogenních stopových prvků
- živý systém – doplňování informací o aktuálních změnách regulací, nových poznatcích výzkumu či metodické návody nejen k použití aplikace, ale obecně k hodnocení hygienického stavu půdy
- validace podkladových map pomocí nových aktuálních dat ve spolupráci s Ministerstvem zemědělství

# Ústřední myšlenka projektu

Využít kombinace:

- BAD – pro modely využít „*best available data*“ - maximálně možný rozsah relevantních záznamů získaných v termínu řešení projektu
- BAT – využít „*best available techniques*“ - nasazení progresivních nástrojů digitálního mapování půd pomocí algoritmů strojového učení + zapojení široké palety doplňkových proměnných pro prediktivní modely

# BAD + BAT = velká potřeba datových vstupů

- kombinování dat pocházejících z různých (geo)chemických vzorkování může být velmi užitečné pro prediktivní modelování, neboť umožňuje zvýšit nejen územní detail, ale také pokrytí znakového prostoru, neboť datově-založené metody strojového učení predikují spolehlivě uvnitř znakové prostoru vysvětlujících proměnných – tj. při jeho dobrém pokrytí dostatečným počtem vzorků, což se může zdát obtížné při prediktivním modelování (geo)chemických obsahů prvků/látek v půdě, kde se potkávají dvě nepříznivé okolnosti: jednak analytická a ekonomická náročnost stanovení a jednak potřeba pokrytí celé řady prediktorů charakterizující predispozice půdy ke znečištění – zranitelnost k sorpci, faktory expozice ke zdrojům (přírozeným i antropogenním) a faktory ovlivňující depozici ze zdrojů či post-depoziční změny

# BAD

- získání relevantních dat = shromáždění dat z různých zdrojů + harmonizace na jednotný analytický normativ dle platné Vyhlášky č. 153/2016 Sb.) tj. „pseudototální“ obsah po rozkladu lučavkou královskou
- harmonizace = robustní regresní řešení pro přepočty mezi různými extrakty používanými ve zdrojích dat



! Děkujeme Ministerstvu životního prostředí ČR a Ministerstvu zemědělství ČR za poskytnutí resortních dat a také jejich resortním organizacím České geologické službě, resp. Ústřednímu kontrolnímu a zkušebnímu ústavu zemědělskému za dlouhodobé úsilí při pořízení i správě těchto dat.

# BAD – rizikové prvky – proč harmonizace?

- Z hlediska vztahu k platné legislativě lze datové vstupy použité pro RP v projektu rozdělit na dvě obecné skupiny:
  - 1) data, která se svým analytickým postupem shodují s legislativně danými postupy pro hodnocení obsahu prvků v zemědělských půdách dle platné Vyhlášky č. 153/2016 Sb. - tj. jsou výsledkem rozkladu vzorku (tj. převedení do roztoku) pomocí mineralizace lučavkou královskou
  - 2) data, které nejsou výsledkem extrakčního postupu shodného s legislativními požadavky – tj. mají odlišný analytický způsob stanovení obsahu RP



# BAD – rizikové prvky – data 1)

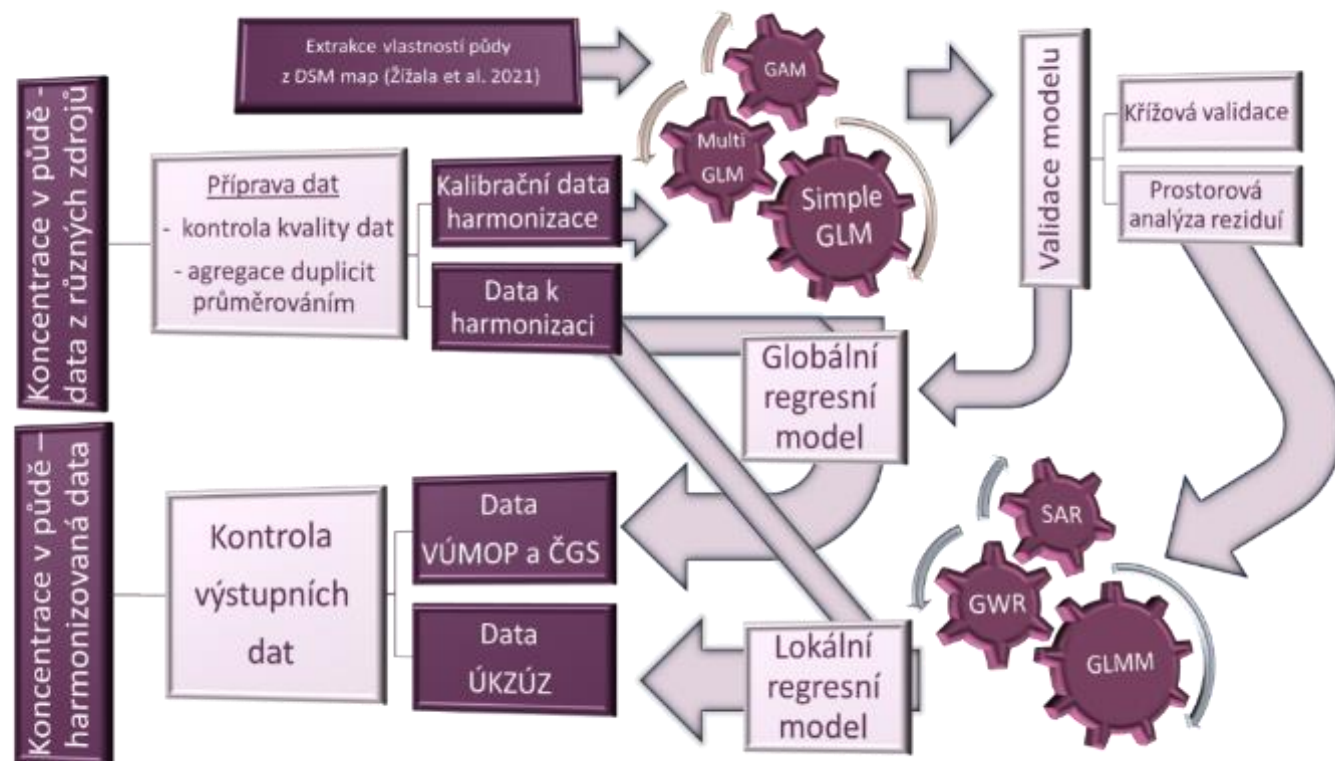
- data, která se svým analytickým postupem shodují s legislativně danými postupy pro hodnocení obsahu prvků v zemědělských půdách dle platné Vyhlášky č. 153/2016 Sb. - tj. jsou výsledkem rozkladu vzorku (tj. převedení do roztoku) pomocí mineralizace lučavkou královskou
  - výsledky Ústředního kontrolního a zkušebního ústavu zemědělského (ÚKZÚZ) z Bazálního monitoringu půd (BMP) a Registru kontaminovaných ploch (RKP) od roku 1998
  - výsledky z přeshraničního projektu – Rizikové látky v půdě ve vztahu k životnímu prostředí – přeshraniční základy ochrany půdy (Bavorsko – Česká republika)
  - výsledky z Mezinárodních evropských projektů FOREGS (SALMINEN et al. 1998) a GEMAS (REIMANN et al. 2014).
  - archivní data z výzkumné činnosti VÚMOP
  - data z analytických stanovení na referenčních plochách při Inventarizaci úložných míst těžebního odpadu v rámci Operačního programu Životní prostředí (OPŽP) v letech 2010–2012.
  - cílená data pořízená v rámci projektu TAČR - analytické srovnání různých extrakčních metod na výběru ploch v ČR

# BAD – rizikové prvky – data 2)

- data, které nejsou výsledkem extrakčního postupu shodného s legislativním požadavkem – tj. mají odlišný analytický způsob stanovení obsahu RP a byla nutná harmonizace dat pomocí regresního přepočtu
  - data Ústředního kontrolního a zkušebního ústavu zemědělského (ÚKZÚZ) z Registru kontaminovaných ploch (RKP) do roku 2009 - parciálního rozklad půdní matrice chladnou 2 mol/l  $\text{HNO}_3$  dle ZBÍRAL et al. (2004)
  - data z litogeochemické databáze (ČGS 1997) pro půdy – archivní analytické údaje o chemickém složení vzorků půd odebraných v rámci různých regionálních geochemických průzkumů, geologického mapování a výzkumu České geologické služby – výsledky klasickou celkovou (tříkyselinovou) mokrou extrakci prvků s analytickou koncovkou ICP nebo AAS, nebo výsledky z nedestruktivní analýzy rentgenovou fluorescenční analýzou (RFA).
  - data z Monitoringu cizorodých látek (MCL) - kombinace celkového tříkyselinové extrakce pomocí  $\text{HNO}_3$ ,  $\text{HClO}_4$ , HF (ISO 14869-1:2001) a dále parciálního rozkladu půdní matrice chladnou 2 mol/L  $\text{HNO}_3$  (ZBÍRAL et al. 2004).

# BAD – rizikové prvky – data 2)

- pracovní postup pro odvození spolehlivých přepočtových rovnic mezi jednotlivými extrakty zahrnoval testování celé řady regresních modelů (obecný lineární model - GML, zobecněné aditivní modely - GAM, geograficky vážené regresní modely – GWR), přičemž jako závislé proměnné byly pro přechodové modely, kromě obsahů z relační extrakce, vybrány další doplňkové půdní parametry – pH,  $C_{ox}$ , zrnitost půdy



# BAD – rizikové prvky – data 2)

- Data ÚKZÚZ – přepočítání mezi extrakty kyseliny dusičné a lučavky královské - nejspolehlivější výsledky byly dosaženy pomocí lokálních přepočtových modelů pomocí tzv. geograficky vážená regrese - GWR (BRUNSDON et al. 1998)

	Data kalibrace	Simple R <sup>2</sup>	Multi R <sup>2</sup>	GAM R <sup>2</sup>	GWR R <sup>2</sup>	Simple – reziduální Moran I*	Multi – reziduální Moran I*	GAM – reziduální Moran I*	GWR – reziduální Moran *I	Data přepočítání
As	7 142	0.51	0.52	0.59	0.81	0.25	0.23	0.24	-0.007	5 235
Be	7 142	0.62	0.62	0.64	0.85	0.23	0.23	0.21	0.07	21 926
Cd	7 142	0.63	0.64	0.68	0.79	0.15	0.14	0.15	-0.008	46 077
Co	7 142	0.84	0.85	0.87	0.91	0.24	0.19	0.13	-0.006	27 779
Cr	7 142	0.71	0.76	0.78	0.86	0.23	0.14	0.14	-0.007	46 068
Cu	7 142	0.71	0.72	0.75	0.87	0,33	0,28	0.22	-0.006	41 827
Ni	7 142	0.72	0.75	0.78	0.88	0.33	0.21	0.17	-0,008	40 741
Pb	7 142	0.84	0.84	0.84	0.89	0.10	0.10	0.09	0.008	46 095
V	7 142	0.68	0.73	0.75	0.85	0.29	0.18	0.18	-0.006	25 717
Zn	7 142	0.61	0.62	0.67	0.82	0.30	0.28	0.21	-0.005	41 875

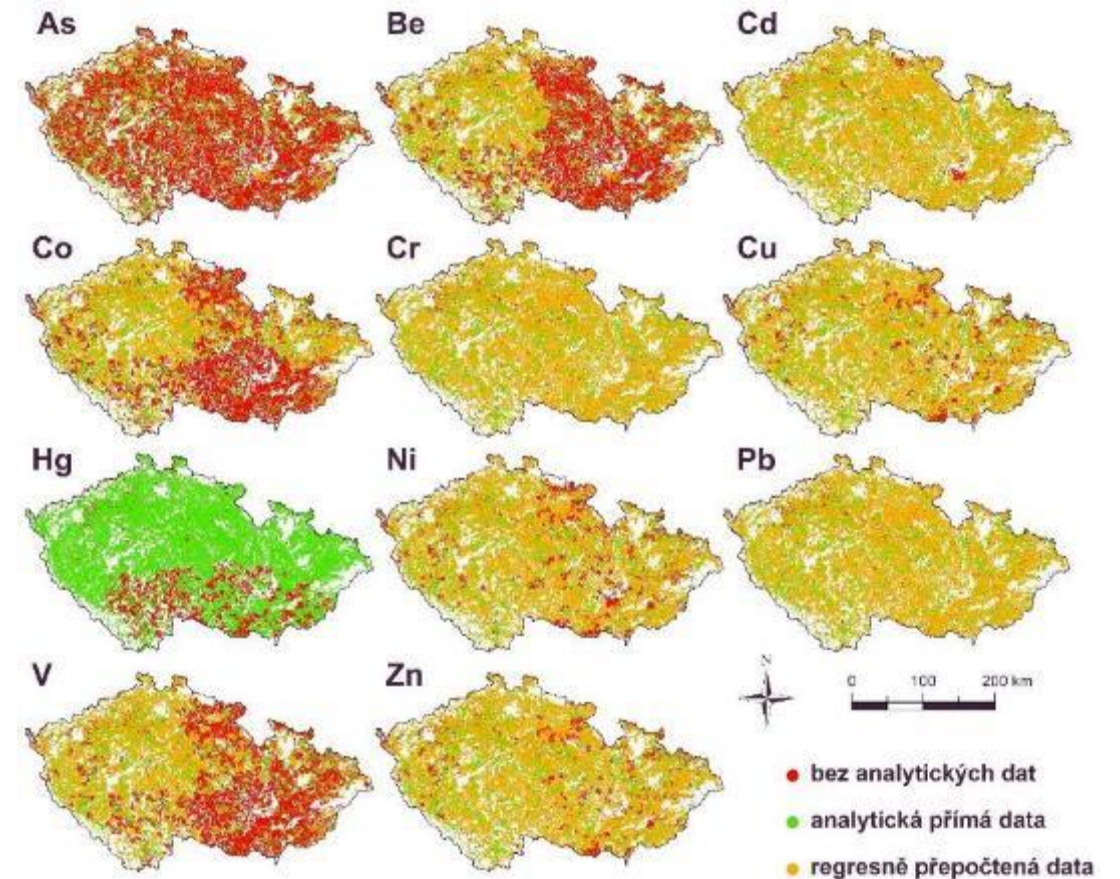
# BAD – rizikové prvky – data 2)

- Data ČGS a MCL (VÚMOP) – regresní přepočítání na základě kombinace extrakce kyselinou dusičnou) a celkového 3-kyselinového rozkladu – jednoduché lineární modely byly dostatečně robustní

	<i>Data kalibrace</i>	<i>Simple R<sup>2</sup></i>	<i>Multi R<sup>2</sup></i>	<i>GAM R<sup>2</sup></i>	<i>Výsledný počet dat po harmonizaci</i>
As	160	0.72	0.72	0.74	3 979
Be	160	0.71	0.69	0.70	3 775
Cd	160	0.74	0.72	0.74	3 542
Co	160	0.95	0.95	0.96	3 086
Cr	160	0.88	0.88	0.93	3 707
Cu	160	0.98	0.97	0.98	3 707
Ni	160	0.91	0.91	0.92	3 441
Pb	160	0.90	0.90	0.91	3 730
V	160	0.85	0.86	0.88	3 413
Zn	160	0.97	0.97	0.97	3 707

# BAD – harmonizovaná data

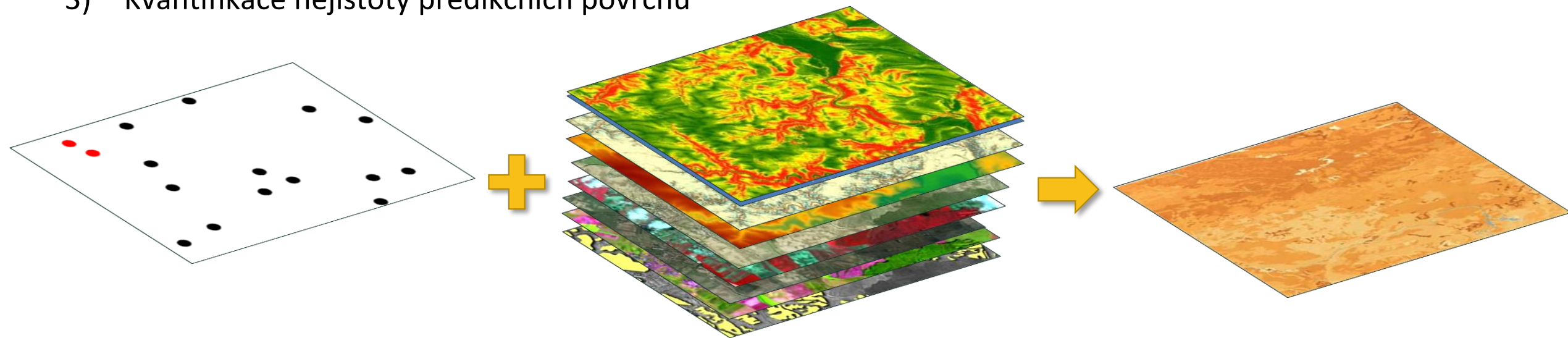
- výsledkem harmonizace dat pro rizikové prvky jsou dva typy dat
  - 1) data, která byla přímo analyticky určena mineralizací vzorku lučavkou královskou
  - 2) data, která byla na cílový obsah v lučavce královské regresně přepočítaná z jiného extraktu (či kombinace extraktů) – data zatížena oproti 1) nejistotou přepočtu, proto jsou následně v modelu penalizována (model dává větší váhu datům typu 1, přičemž míra spolehlivosti přepočtu vyjádřená koeficientem determinace regresního modelu určuje váhy vzorku pro prediktivní model)



# BAT

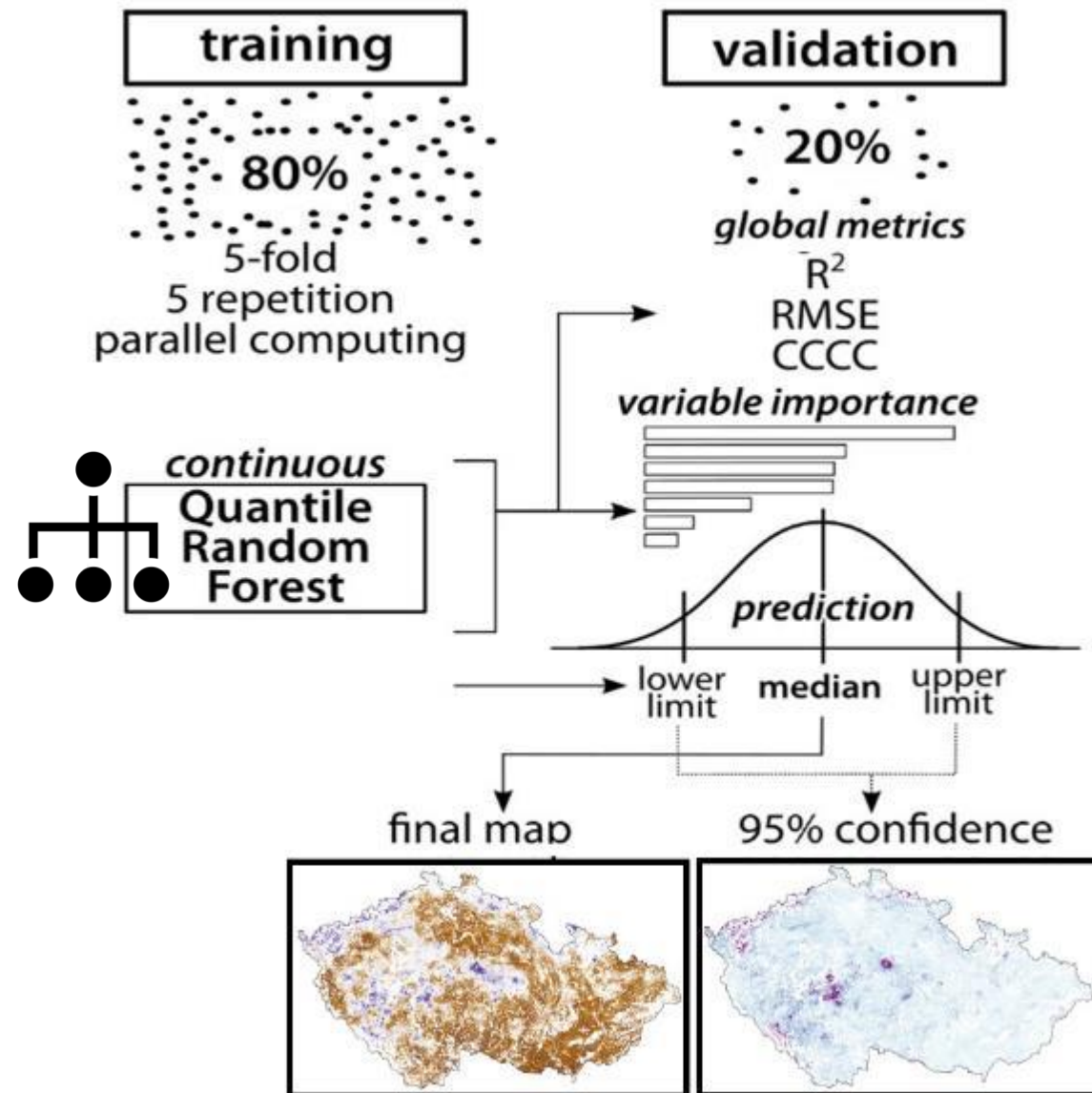
## Principy Digitálního mapování půd / Prediktivní geochemické mapování

- půdní vzorky – BAD (harmonizovaná data) + pomocné proměnné, tzv. kovariáty (prediktory)
- 1) Kalibrace modelu – nalezení nelineárních vztahů mezi vzorky a prediktory
  - 2) Predikce v neměřených bodech a tvorba predikčních povrchů
  - 3) Kvantifikace nejistoty predikčních povrchů



# BAT

- pro modelování vztahů mezi prediktory a cílovými proměnnými použit algoritmus kvantilové formy náhodných lesů – „*Quantile Regression Forest*“ (QRF) (MEINSHAUSEN 2006)
- kvantilová verze náhodných lesů - umožňuje lokální hodnocení spolehlivosti modelu pomocí šířky predikčního intervalu, která doplňuje celkovou míru přesnosti modelu kvantifikovaného z rozdílů predikovaných a skutečně naměřených hodnot v souboru dat odejmutých z primárního vzorku před vlastním trénováním modelu
- trénování vážené verze QRF z důvodu zohlednění spolehlivosti dat při kombinování regresně přepočtených i přímo naměřených dat
- kvantifikace vlivu důležitosti kovariantních proměnných (tzv. „*variable importance*“) –tj. jaké prediktory jsou informačně hodnotné pro přesnou predikci





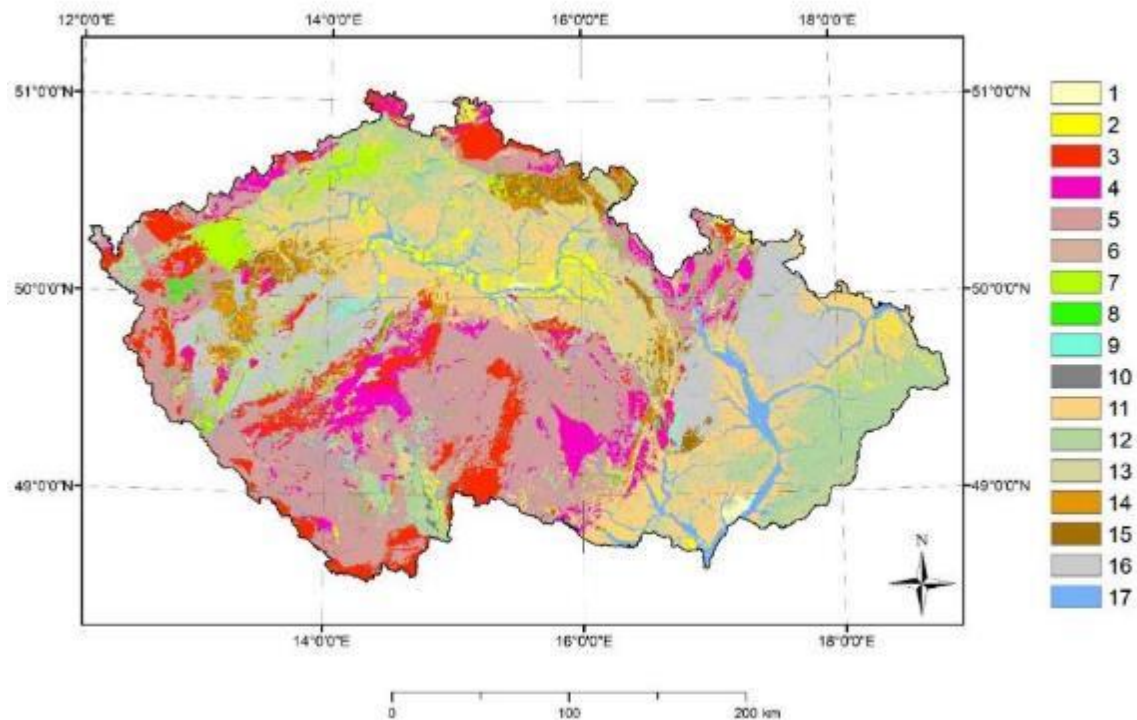
# BAT - Prediktory použité pro tvorbu modelů

- použitý model náhodných lesů je založen na kvantifikování vztahů mezi modelovanou proměnnou a prediktory ve známých bodech, které pak slouží k predikci modelované proměnné v bodech, kde cílovou informací o modelované proměnné neznáme.
- z hlediska pochopení prostorové distribuce znečišťujících látek jsou důležité zejména interakce mezi faktory: zdroje kontaminace, distribuční mechanismy, vlastnosti kontaminantů a sorpční vlastnosti půdy
- ~ 30 prediktorů prostředí + 20 prostorových prediktorů (geografické obalové vzdálenosti pro popis prostorových autokorelačních vztahů)
  - přirozené zdroje prvků v půdách (půdní vlastnosti, vlastnosti geologického podloží)
  - antropogenní zdroje (koncentrace znečištění v ovzduší, urbanizace, hustota dopravní sítě, hustota zaniklé těžby)
  - terénní vlastnosti
  - klimatické charakteristiky

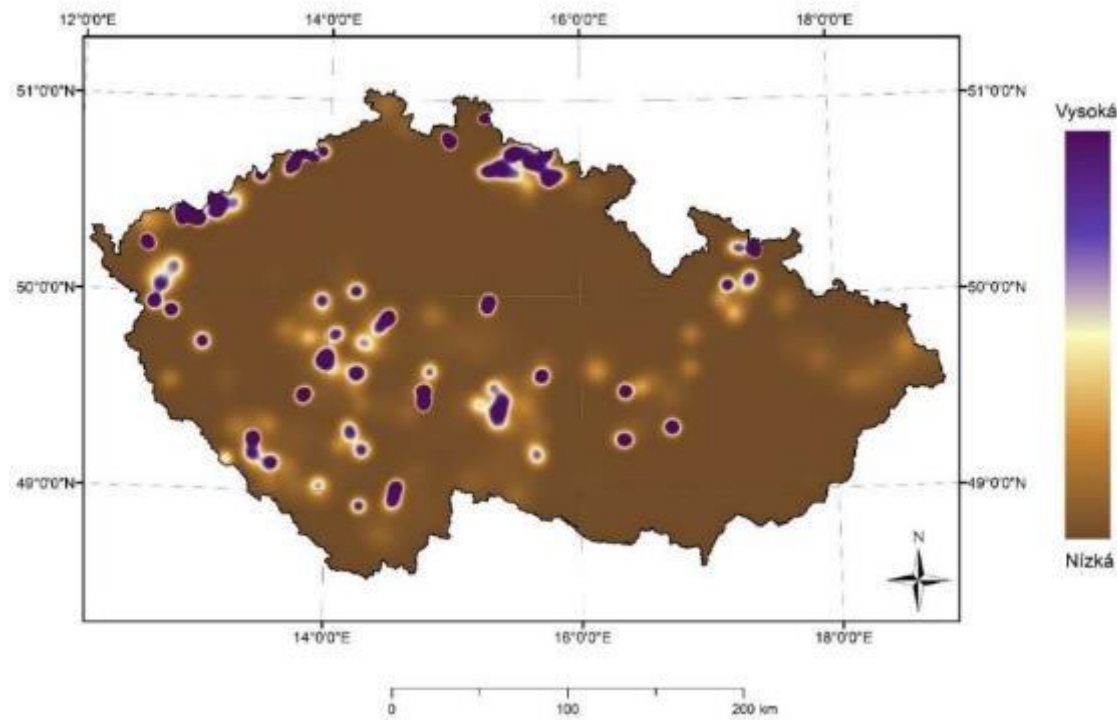
# BAT – prediktory pro přirozené zdroje prvků

- pro potřeby prediktivního mapování byly vytvořeny rastrové podklady agregovaných skupin půdotvorných substrátů - kombinace několika zdrojů z geologického mapování (nezakryté geologické mapy ČR - © Česká geologická služba) a pedologického mapování (mapování půdotvorných substrátů dle KPP VÚMOP digitalizované pro MZe ČR)
- pro lepší parametrizaci litologické variability podloží bylo využito doplňkových numerických proměnných pro potencionální zpřesnění litologie pokrývající geofyzikální anomálie pro rozlišení nehomogenity horninového prostředí - anomálie tíhového zrychlení, magnetické anomálie a anomálie radioaktivity hornin v podobě úhrnné aktivity gama z letecké radiometrie (© Česká geologická služba)
- pro lepší parametrizaci geochemické variability podloží byly připraveny odvozené rastrové podklady pro vzdálenost k nejbližší tektonické poruše a jejich hustota, dále hustota úložišť těžebního odpadu (dle databáze © Česká geologická služba) pro odhad výskytu přirozených metalogenních zón

*Rastr klasifikace podloží do 17 agregovaných jednotek půdotvorných substrátů v ČR*



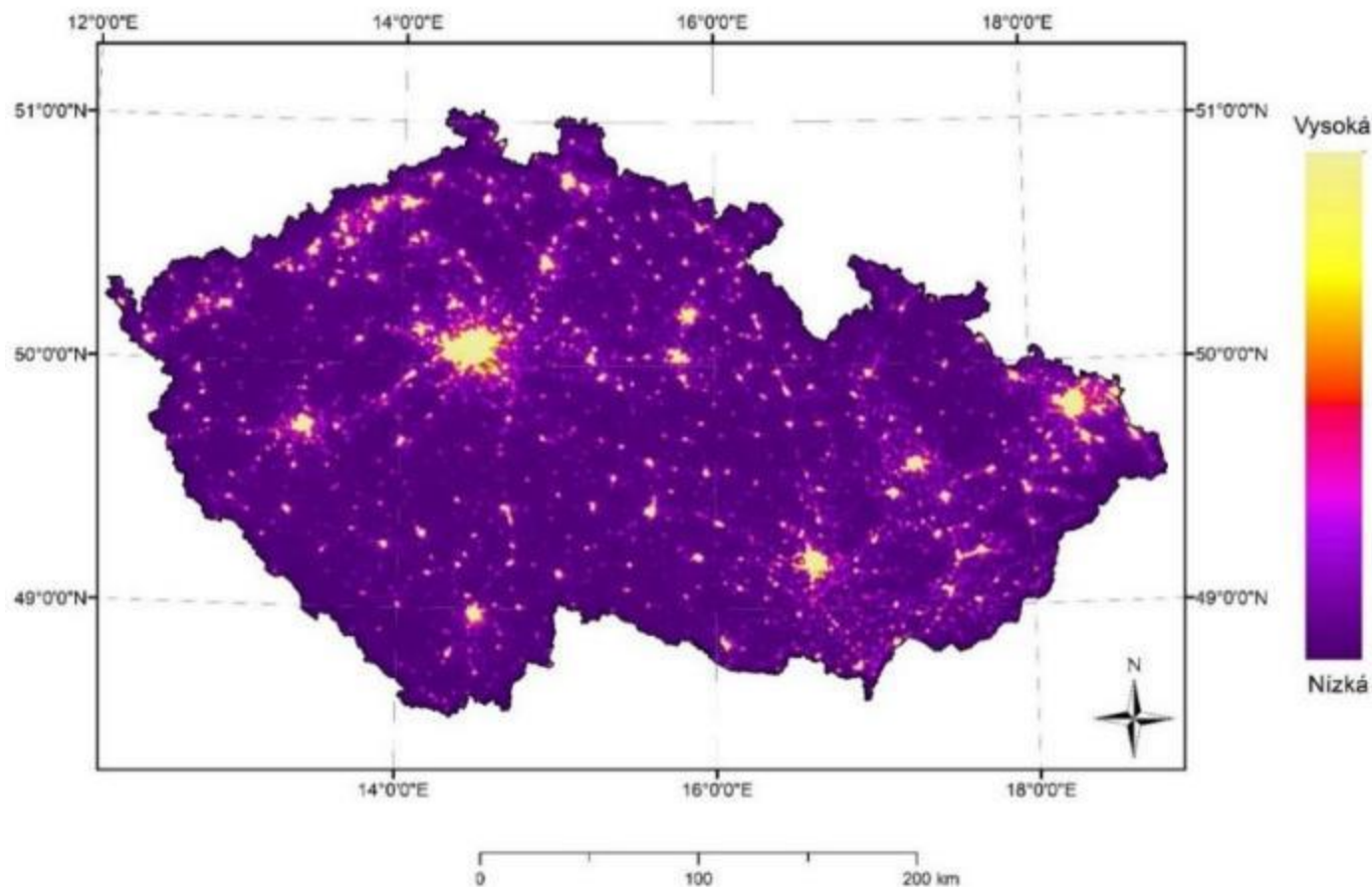
*Hustota těžebních aktivit, která vznikla na základě jádrové odhadu hustoty z bodových dat úložných míst těžebního odpadu (databáze © Česká geologická služba)*



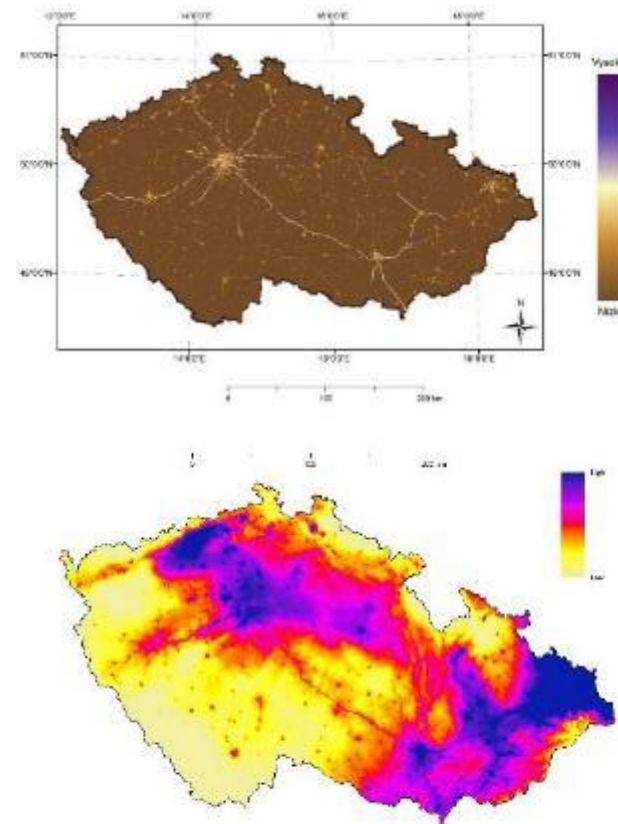
# BAT – prediktory pro antropogenní zdroje prvků

- pro parametrizaci antropogenně podmíněných atmosférických vstupů rizikových látek/prvků do půdy bylo využíváno kombinace – parametrů reliéfu (expozice) s využitím klimatických (srážkových) dat a výsledků monitorování znečišťujících látek v ovzduší v rámci Informačního systému kvality ovzduší (ISKO) (BARTOŇOVÁ 2004) - přímé informace o imisních vstupem pro některé látky (benzo(a)pyren, As, Cd, Ni, Pb), pro ostatní typy látek byly využity pětileté průměry parametru PM10, tj. suspendované částice, které jsou dobrým indikátorem potenciální atmosférické zátěže
- kromě stacionárních emisních zdrojů se na zvýšených zátěžích mohou podílet i mobilní spalovací zdroje – zejména doprava – na základě vektorových podkladů silniční a dálniční sítě byly pomocí jádrového vyhlazení s vážením pro intenzitu dopravy (třídy silnic) vytvořeno hustotní pole intenzity vlivu dopravy
- pravděpodobnost antropogenních zátěží, zejména těch průmyslových, roste v urbánních a periurbánních oblastech - míru urbanizace lze dobře parametrizovat pomocí výsledků dálkového průzkumu Země pomocí intenzity umělého jasů noční oblohy z osvětlení (ELVIDGE et al. 2017)
- pro zjednodušený odhad dlouhodobého vlivu hospodaření na zemědělské půdě byla na základě archivních dat LPIS (1996–2020) provedena kategorizace 4 typů využití půd - plochy, které byly v jednotlivých letech vedeny vždy jako orná půda, resp. jako travní porost, dále plochy, které v daném období byly alespoň jednou vedeny jako chmelnice či vinice, a dále zbytek ploch, kde se kultury v různých letech střídaly (kromě uvedených ploch chmelnic a vinic).

*Světelná intenzita noční oblohy jako vhodný parametr pro vymezení míry urbanizace území – zdroj dle (ELVIDGE et al. 2017)*



*Odvozený rastr intenzity dopravní zátěže z hustoty dopravních sítí vážených mírou průměrné intenzity v jednotlivých kategoriích pozemních komunikací pomocí jádrového vyhlazení liniový prvků silniční sítě (zdroj databáze silniční sítě - © ZABAGED)*



*Rastr ročních průměrných koncentrací BaP (víceleté průměry), © Český hydrometeorologický ústav  
Informační systém kvality ovzduší (ISKO)*

# Výsledné mapy

- statické mapy vs. interaktivní povrchy
- prediktivní povrch koncentrací – 11 x (RP) + 7 x (POPs)
  - Q5 – dolní hranice predikce
  - Q50 – průměrná predikce
  - Q95 – horní hranice predikce
- prediktivní povrch pravděpodobnosti překročení limitní hodnoty - 11 x (RP) + 7 x (POPs)
  - Lehké půdy + Běžné půdy = kombinovaný povrch
- prediktivní povrch pravděpodobnosti výskytu lehké půdy (< 0,01 mm) = pravděpodobnost platnosti přísnějších limitních hodnot - 1 x
- prediktivní povrch pravděpodobnosti překročení alespoň jedné limitní hodnoty – 2 x
  - rizikové prvky
  - organické polutanty

