

# Prostorové vyhlazování rizika v pojišťovnictví

Od PSČ k přesné rizikové mapě

Tobiáš Veska

28.11.2025

# Agenda

1. Úvod – geografický efekt, limity administrativního členění (PSČ)
2. Teorie kredibility – tradiční stabilizace dat
3. Principy prostorového vyhlazování
4. Jádrové vyhlazování
5. Zobecněné aditivní modely (GAM)
6. Aplikace a dopady

# 1. Úvod



# Geografický efekt v pojištění

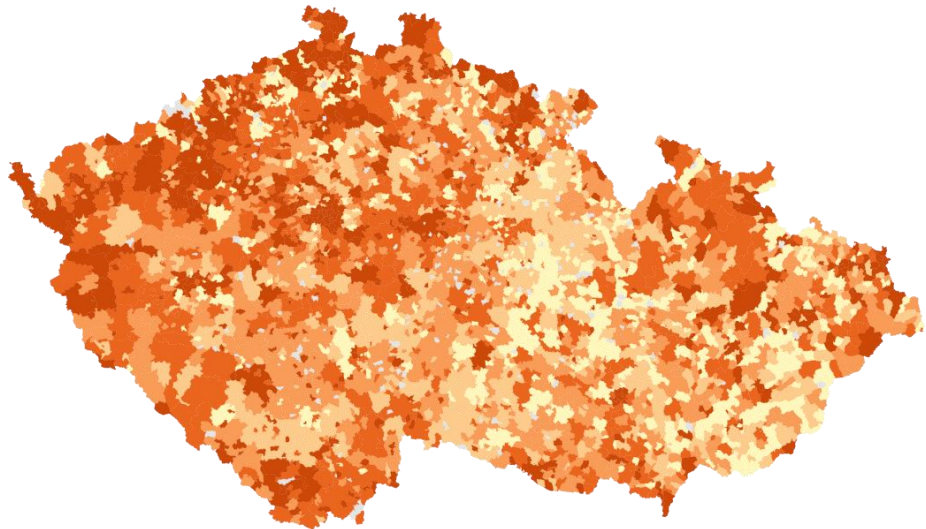
- Rizikovost je prostorově heterogenní
  - Silný vliv faktorů jako doprava, klima, kriminalita...
- Vzniká potřeba zachytit prostorové rozdíly s omezenými daty:
  - Nerovnoměrná data (města vs. venkov)
  - Agregovaná data (PSČ, okresy)
  - Volatilita kvůli nízké expozici (příliš malé objemy v některých segmentech)
- Cílem je vytvořit hladkou, stabilní a predikčně silnou geografickou mapu rizika, která:
  - Odpovídá realitě
  - Je stabilní i při nízkých objemech dat
  - Robustnost vůči odlehlým hodnotám (outliers)
  - Implementovatelnost do ratingové struktury
- Limitace administrativního členění
  - Běžné administrativní jednotky (PSČ, okresy) neodpovídají skutečné geografii rizika → vznikají skoky, extrémy a nestabilita

**"Vše souvisí se vším, ale blízké věci spolu souvisí více než ty vzdálené."**

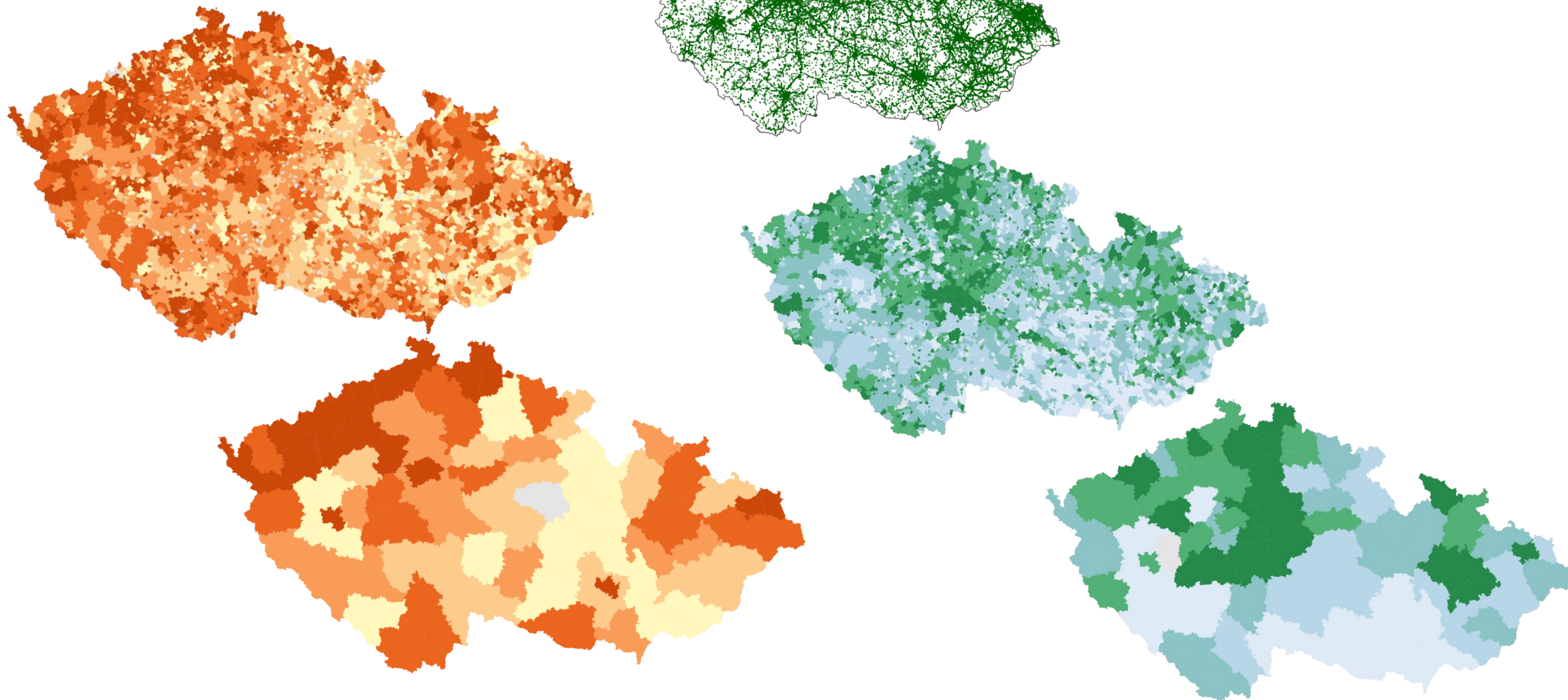
*Waldo Tobler*  
*First Law of Geography*

# Data

- Pro prezentaci byla použita volně dostupná data:
  - Kriminalita - [kriminalita.policie.gov.cz](http://kriminalita.policie.gov.cz)
  - Dopravní nehody - [nehody.policie.gov.cz/](http://nehody.policie.gov.cz/)
  - Počet obyvatel - [csu.gov.cz](http://csu.gov.cz)
- Kriminalita –
  - frekvence, vážená počtem obyvatel (okres/PSČ)
- Pro demonstraci, mapy a grafy bylo použito R:
  - Speciálně pro shapefily a údaje o rozdělení ČR pak byl použit balíček RCzechia - [github.com/jlacko/RCzechia](https://github.com/jlacko/RCzechia)
- Dopravní nehody
  - Původní data s přesnou lokací
  - frekvence, vážená počtem obyvatel (okres/PSČ)

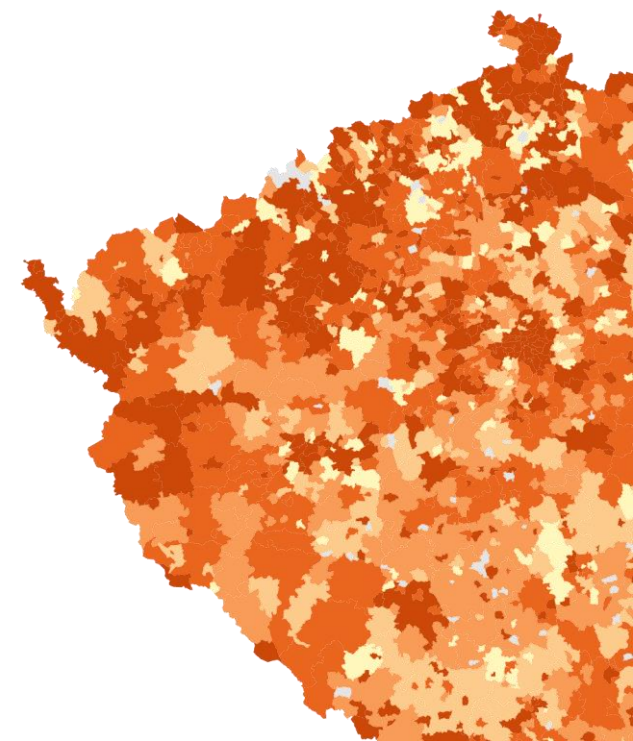


# Data



# Proč administrativní hranice selhávají v moderním pricingu?

- Administrativní vs riziková geografie: Nehody nerespektují hranice segmentů. Rizikovost se mění spojitě (krajina, provoz), nikoli skokově
- Past malých čísel (Small Area Problem): Většina segmentů má příliš malou expozici. Rozptyl odhadu roste
- Důsledek: Nízká robustnost - jedna škoda neúměrně vychýlí odhad. Nejedná se o signál, ale o šum
- Nekonzistence: Sousední vesnice s podobným charakterem mají dramaticky odlišné sazby
- Obchodní problém: Klienti na hranicích zón čelí cenovým skokům, které neumíme vysvětlit ničím jiným než šumem v datech
- Cíl: Najít hladký trend, nikoli „zubatou“ křivku



# Od izolované informace k souvislostem

- Princip půjčování informace: Když jedna oblast nemá dostatek dat, část informace si „půjčí“ (borrowing strength) od svého okolí
- Odhad (intuitivně):
  - $\hat{\mu} = \text{vlastní data} \oplus \text{informace z okolí}$
- Hledáme kompromis:
  - Snížit rozptyl (šum) za cenu mírného vychýlení (přimíchání okolí)
- Malé objemy  $\rightarrow$  vysoký rozptyl odhadu
  - Odhad je silně ovlivněn náhodou
    - $\bar{X}_j = \mu_j + \epsilon_j, \epsilon_j \sim N(0, \frac{\sigma^2}{n_j})$
  - Spolehlivost roste s více daty – klesá rozptyl
    - $\text{Var}(\bar{X}_j) = \frac{\sigma^2}{n_j}$ , kde  $\bar{X}_j = \frac{1}{n_j} \sum_i X_{ij}$

$j$  – index regionu

$n_j$  – expozice

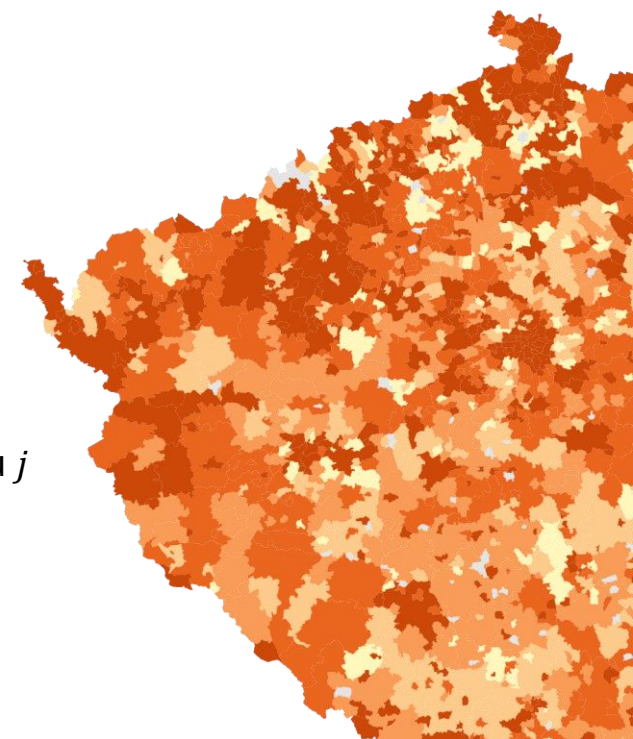
$X_{ij}$  – jednotlivá pozorování v regionu  $j$

$\bar{X}_j$  – výběrový průměr region  $j$

$\mu_j$  – neznámá hodnota parametru

$\epsilon_j$  – náhodná chyba

$\sigma^2$  – rozptyl



# Spojitosť vs. Diskrétnosť

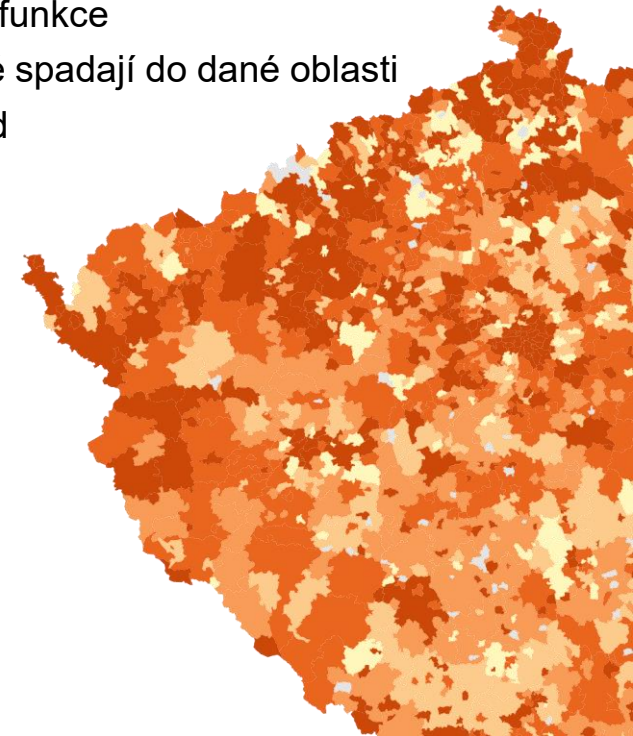
- Faktory ovlivňujúci riziko se nemění skokově
  - Skutečné riziko je hladká funkce v prostoru
    - $\mu(x, y)$
- Administrativní hranice vytváří blokově konstantní odhad
  - $\hat{\mu}_{PS\check{C}_j}(x, y) = \bar{X}_j = \frac{1}{n_j} \sum_i X_{ij}, (x, y) \in PS\check{C}_j$
- Výsledkem je vysoká chyba diskretizace
  - $\mathbb{E} \left[ \left( \mu(x, y) - \hat{\mu}_{PS\check{C}}(x, y) \right)^2 \right] \gg 0$
- Důsledek: nahrazujeme detailní profil hrubým průměrem
  - Vzniká rozdíl mezi realitou a modelem (Bias), který nelze odstranit větším množstvím dat, pouze změnou metody
    - $\mu(x, y) \not\approx \hat{\mu}_{PS\check{C}}(x, y)$

$(x, y)$  – souřadnice (oblasti, škody)

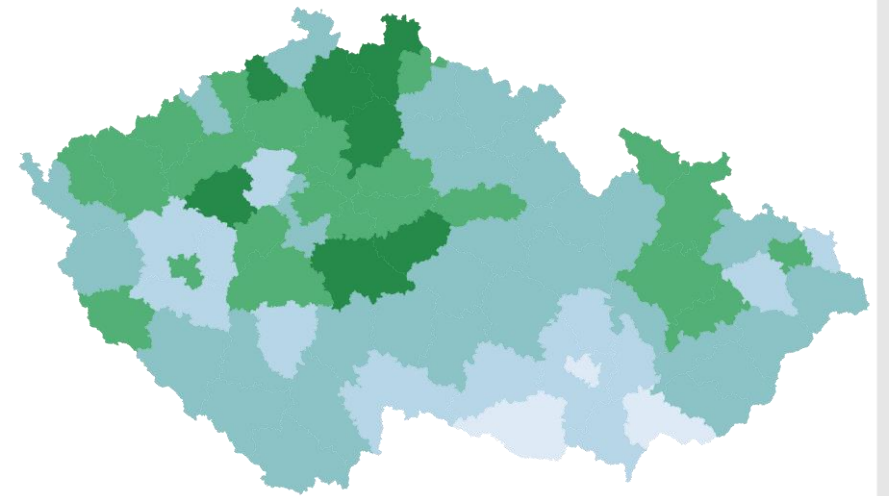
$\mu(x, y)$  - skutečná riziková funkce

$PS\check{C}_j$  - množina bodů, které spadají do dané oblasti

$\hat{\mu}_{PS\check{C}_j}(x, y)$  - diskrétní odhad



## 2. Theorie kredibility



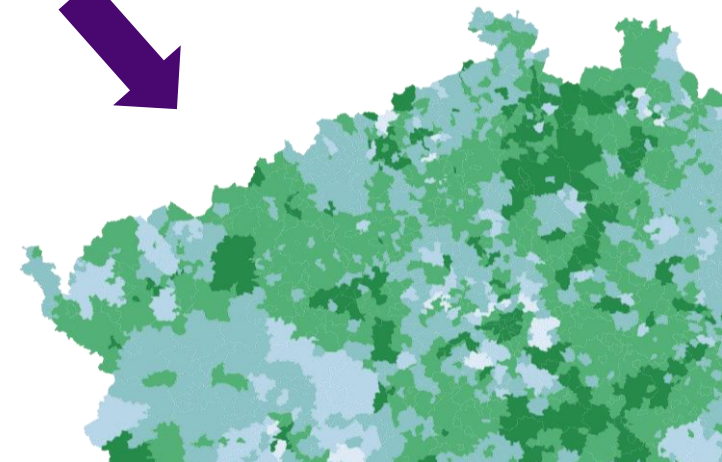
# Proč kredibilita?

- Malé segmenty mají vysoký rozptyl odhadu
  - Lokální průměr je užitečný, ale nestabilní
- Globální průměr je stabilní, ale necitlivý
  - Ignoruje lokální strukturu
- Kredibilita kombinuje obě informace
  - $\hat{\mu}_j^{kred} = Z_j \cdot \bar{X}_j + (1 - Z_j) \cdot M, Z_j \in [0,1]$
- Cílem je stabilizace odhadů, ne nalezení ideálního modelu
- Kredibilita je v pojišťovnictví všudypřítomná  
ohodnocení rizikovosti flotil, systém bonus–malus, rezervování..

$\hat{\mu}_j^{kred}$  – kredibilní odhad pro region  $j$

$Z_j$  - kredibilní faktor

$M$  – globální průměr



# Kredibilní faktor Z

- Z vyjadřuje spolehlivost lokálního odhadu
  - Roste s objemem expozice  $n$ 
    - $\hat{\mu}_j^{kred} = Z_j \cdot \bar{X}_j + (1 - Z_j) \cdot M, Z_j \in [0,1]$
- Je nutné najít správnou váhu Z:
  - Lze použít obecný model Bühlmann–Straub:

$$Z_j = \frac{n_j}{n_j + k}$$

- Parametr  $k$  je tzv. kredibilní konstanta

$$k = \frac{\hat{\sigma}^2}{\hat{\tau}^2}$$

$$\hat{\sigma}^2 = \frac{1}{J} \sum_{j=1}^J s_j^2, kde s_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^J (X_{ij} - \bar{X}_j)^2$$

$$\hat{\tau}^2 \approx \frac{1}{J - 1} \sum_{j=1}^J (\bar{X}_j - M)^2 - \frac{\hat{\sigma}^2}{\bar{n}}$$

$\hat{\mu}_j^{kred}$  – kredibilní odhad pro region  $j$

$Z_j$  - kredibilní faktor

$M$  – globální průměr

$\hat{\sigma}^2$  - odhad vnitroskupinového rozptylu, vnitřní variabilita v segmentech

$\hat{\tau}^2$  - odhad meziskupinového rozptylu, variabilita mezi segmenty

$\bar{n}$  - průměrná expozice v regionu

$J$  – celkový počet segmentů

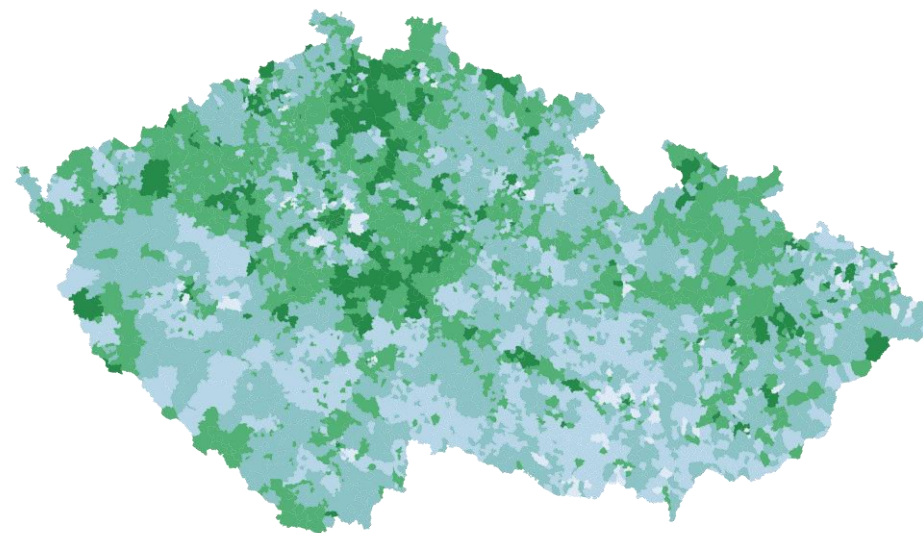
# Jak kredibilita ovlivňuje chybu odhadu

- Celková chyba odhadu = rozptyl + systematická odchylka

$$MSE(\hat{\theta}) = \underbrace{Var(\hat{\theta})}_{\text{náhoda}} + \underbrace{(\mathbb{E}[\hat{\theta}] - \theta)^2}_{\text{syst. odchylka}}$$

- Kredibilita
  - Snižuje rozptyl (stabilizuje lokální odhady)
  - Neřeší problém diskretizace (hranice PSČ)
  - Máme stabilnější, ale stále zubatou mapu
  - Systematická odchylka zůstává

$$\mathbb{E}[\hat{\mu}_{PS\check{C}}^{kred}(x, y)] \approx \mu(x, y)$$



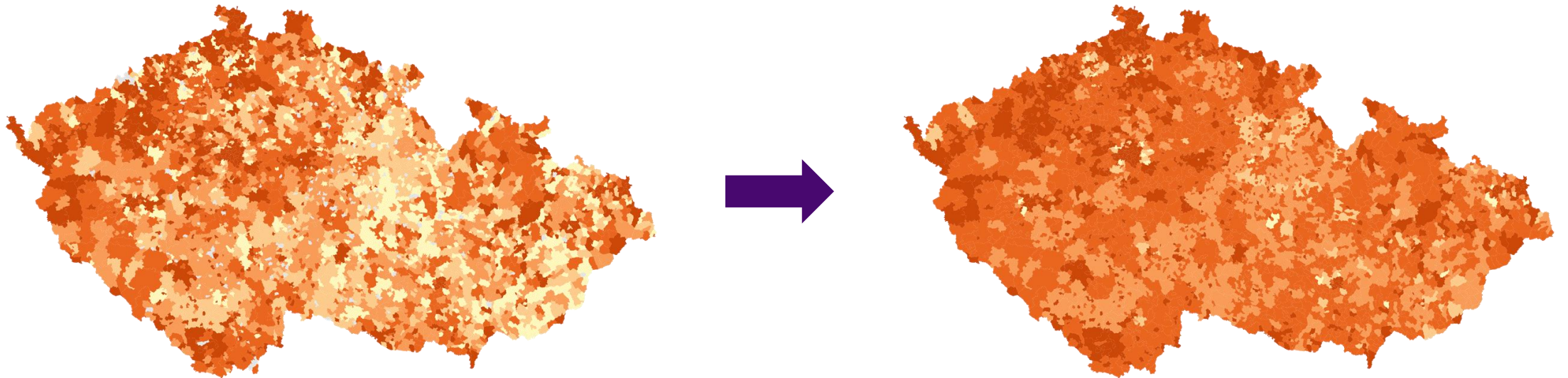
# Nedostatky teorie kredibility

- Řeší pouze rozptyl odhadu, ne strukturu rizika
- Nesnižuje systematickou odchylku způsobenou PSČ
- Stále jde o kusovou aproximaci prostoru
- Kredibilita stahuje odhad ke globálnímu průměru ( $M$ ), nikoliv k sousedům
- Ignoruje prostorovou souvislost: Sousední oblasti se navzájem neovlivňují
- Je to baseline, nikoliv finální model



# Nedostatky teorie kredibility

- Řeší pouze rozptyl odhadu, ne strukturu rizika
- Nesnižuje systematickou odchylku způsobenou PSČ
- Stále jde o kusovou aproximaci prostoru
- Kredibilita stahuje odhad ke globálnímu průměru ( $M$ ), nikoliv k sousedům
- Ignoruje prostorovou souvislost: Sousední oblasti se navzájem neovlivňují
- Je to baseline, nikoliv finální model



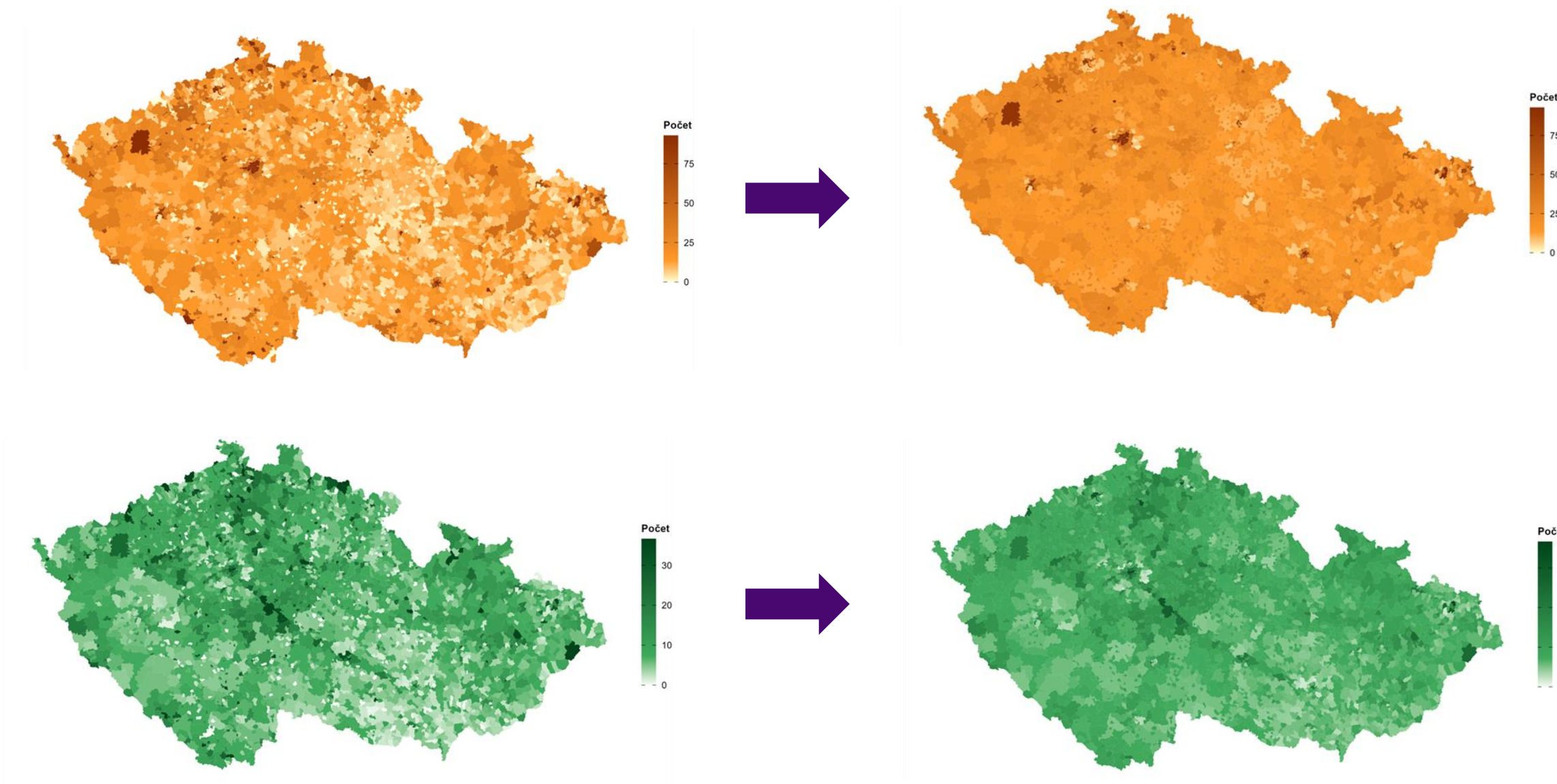
# Pokročilé varianty kredibility a jejich limity

- Hierarchická kredibilita (Hierarchical Credibility)
  - Využívá vnořenou strukturu regionů (PSČ  $\subset$  Okres  $\subset$  Kraj  $\subset$  ČR)
  - Princip: Pokud PSČ nemá data, půjčí si od okresu, ten případně od kraje
  - Matematicky: Rekurzivní aplikace kredibilního vzorce pro více úrovní
  - Limit: Stále jde o diskrétní bloky. Hranice kraje vytvoří skok v sazbě, i když jsou obce sousedé.
- Kredibilita s prostorovou kovariancí (Spatial Credibility)
  - Opuštění předpokladu nezávislosti mezi regiony
  - Zavádí korelaci na základě vzdálenosti regionů  $d_{jk}$

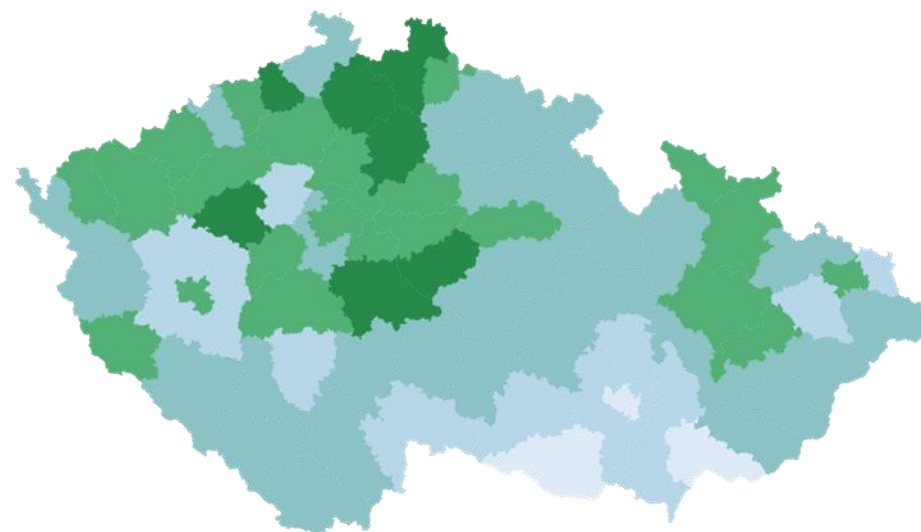
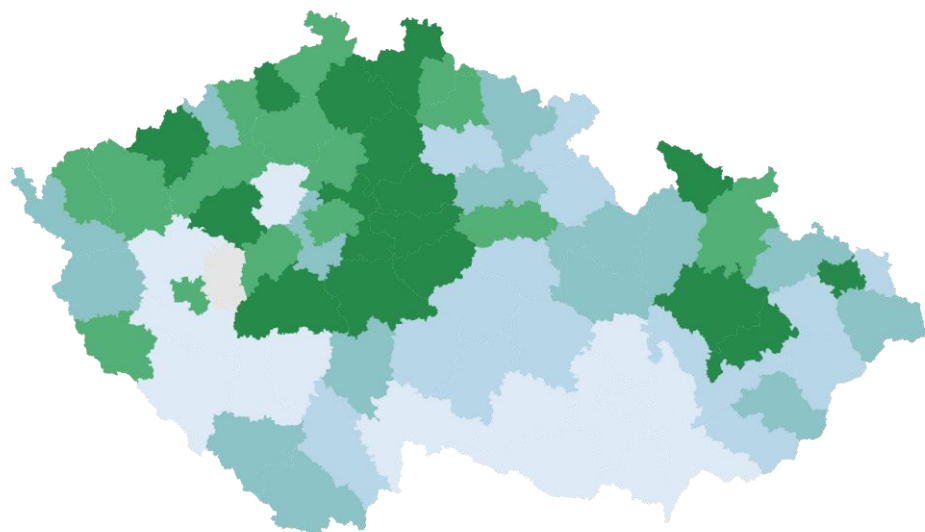
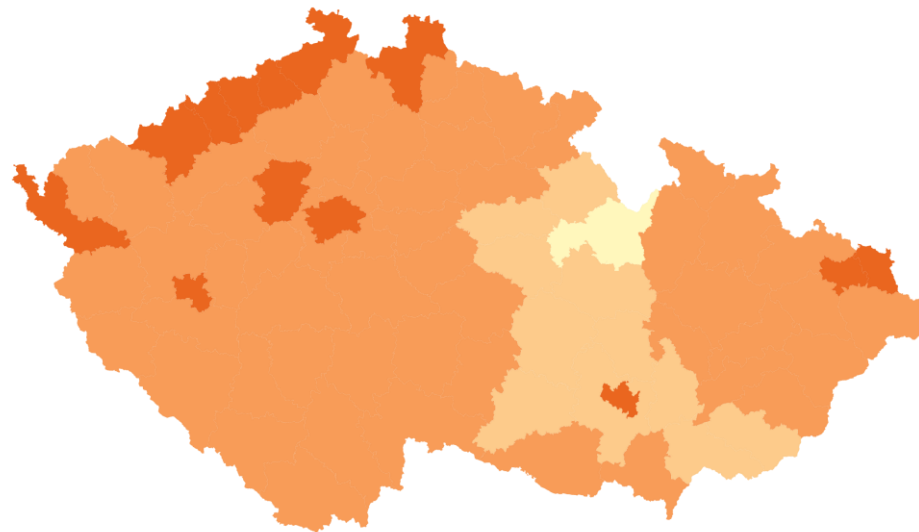
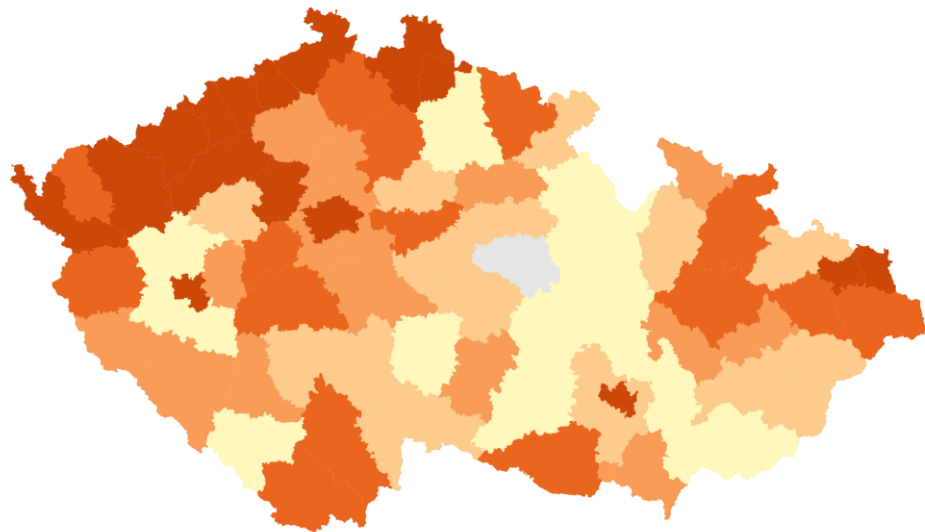
$$\text{Cov}(X_j, X_k) \neq 0$$

- Náročné na odhad parametrů a výpočet

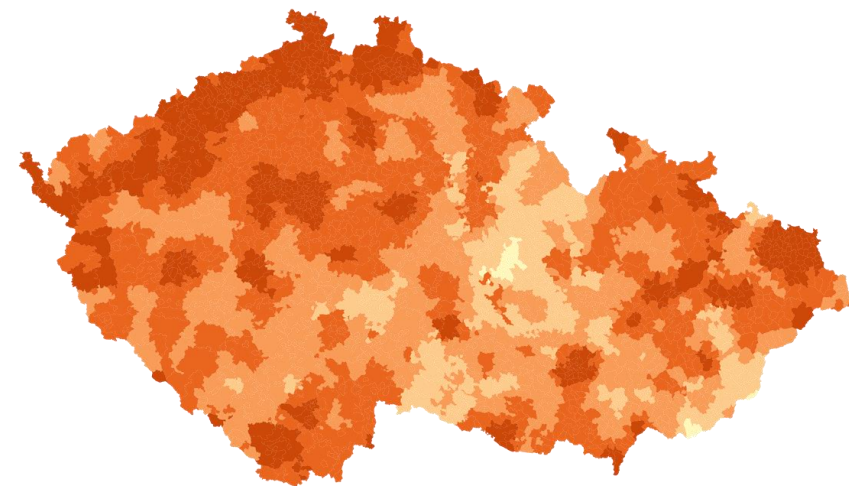
# Kredibilita – PSČ



# Kredibilita – Okres

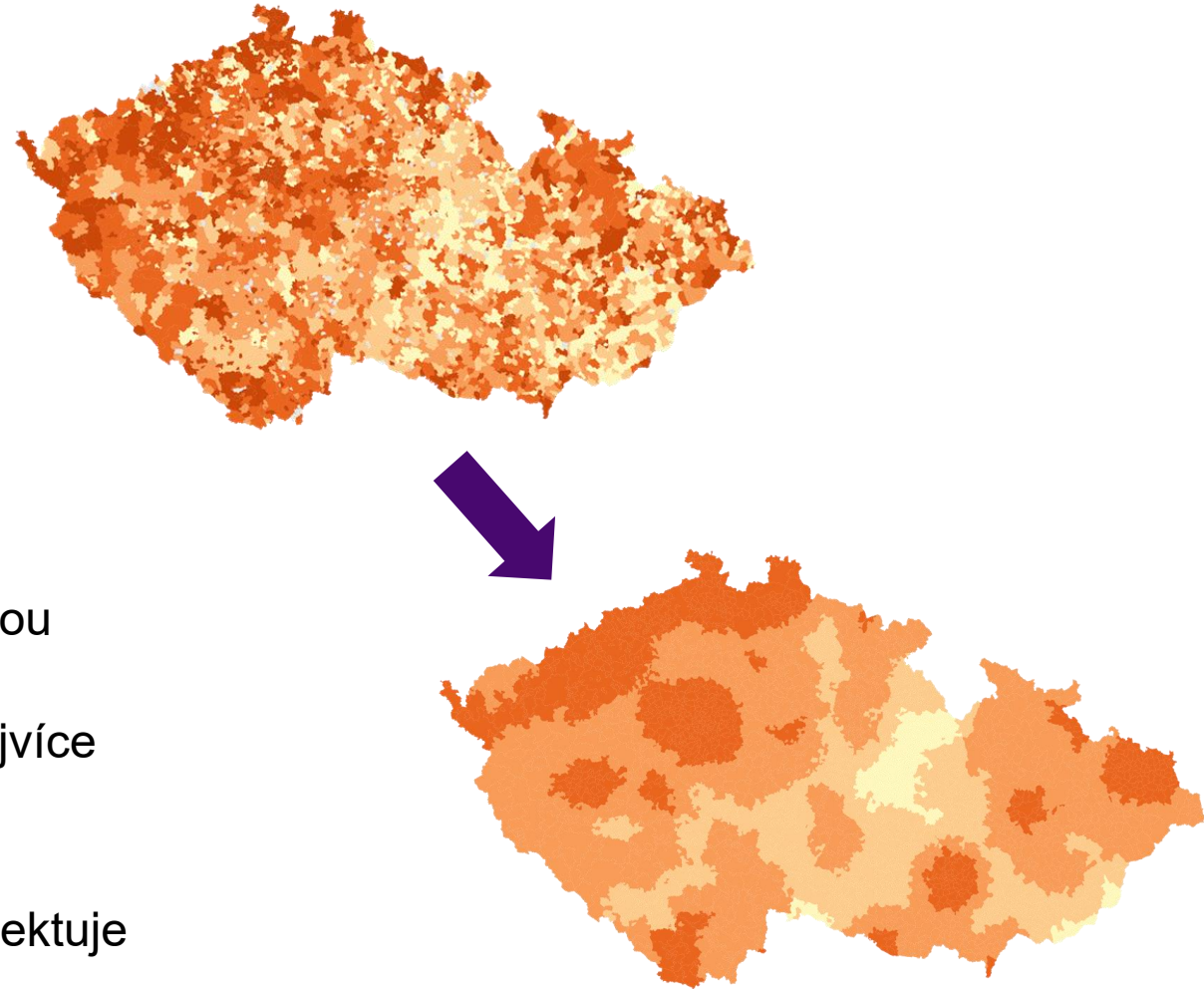


### 3. Prostorové vyhlazování



# Proč potřebujeme prostorové vyhlazování?

- Povaha rizika je spojitá, nikoliv skoková
  - Faktory rizika se mění plynule
- Formalizace – proto modelujeme riziko jako spojitou funkci v prostoru
  - $\mu: \mathbb{R}^2 \rightarrow \mathbb{R}$
- PSČ tvoří pouze po částech konstantní aproximaci
  - $\hat{\mu}_{PSČ}(x, y) \not\approx \mu(x, y)$
- Kredibilita stabilizuje odhady, ale nevyužívá geografickou blízkost
- Využití prostorové autokorelace: Blízké body nesou nejvíce informace
  - mají podobné riziko → přispívají k odhadu
- Cílem je stabilní a realistická riziková mapa, která respektuje spojitost prostoru
  - $\hat{\mu}(x, y) \approx \mu(x, y)$



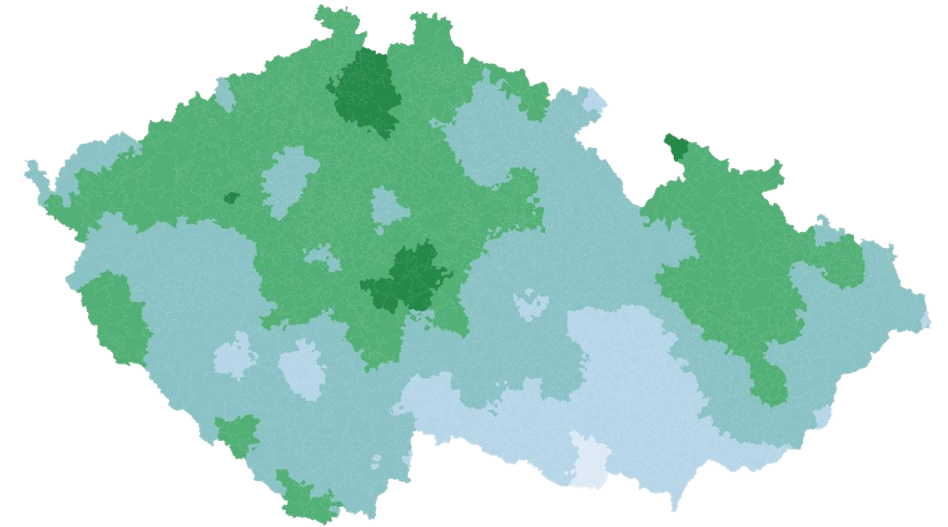
# Dvě cesty

- Oba přístupy mají stejný cíl – odhadnout hladkou funkci rizika v prostoru
- Jádrové vyhlazování (Kernel smoothing)
  - Filosofie – „lokální průměrování“
  - Výpočet váženého průměru okolí

$$\hat{\mu}(x, y) = \frac{\sum_i K_h(d_i) \cdot Y_i}{\sum_i K_h(d_i)}$$

- Zobecněné aditivní modely (GAM)
  - Filosofie – „Globální optimalizace“
  - Hledání kompromisu mezi věrností datům a hladkostí

$$\hat{\mu} = \underset{\mu}{\operatorname{argmin}} \left[ \sum_i (Y_i - \mu(x_i, y_i))^2 + \lambda J(\mu) \right]$$



$d_i$  - Eukleidovská vzdálenost mezi cílovým bodem  $(x, y)$  a daty  $(x_i, y_i)$

$K_h$  - jádrová funkce

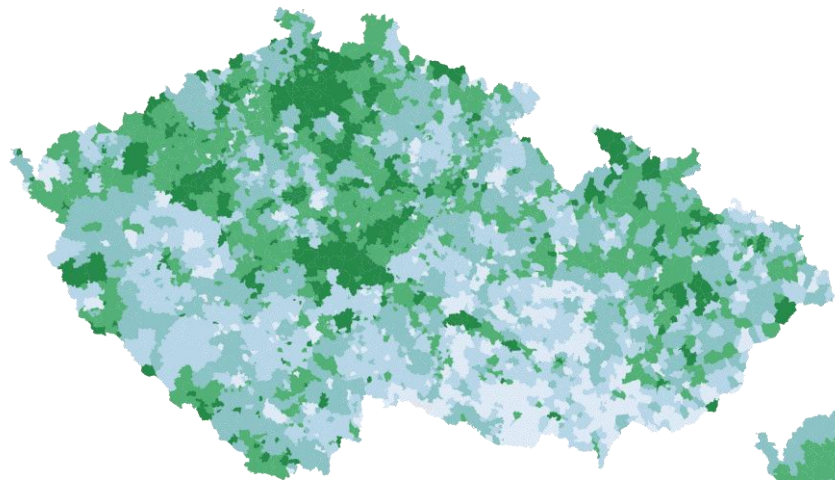
$h$  - parametr šířky pásma (Bandwidth)

$\lambda$  - vyhlazovací parametr

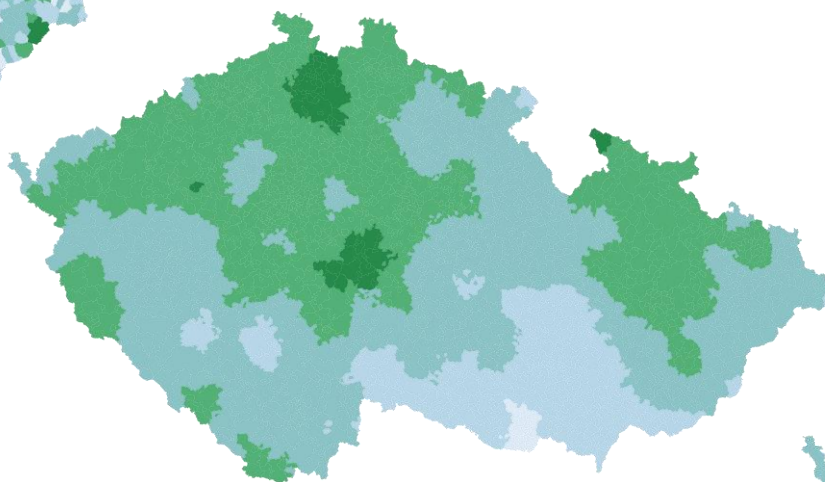
$J(\mu)$  - funkce penalizace „vlnitosti“

$Y_i$  - hodnota rizika v daném segmentu  $i$

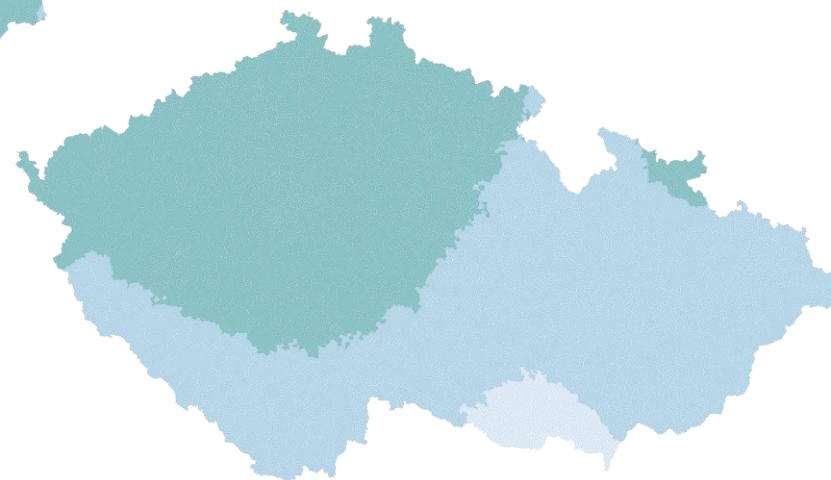
# Mapy, volba parametru



$h = 2km$

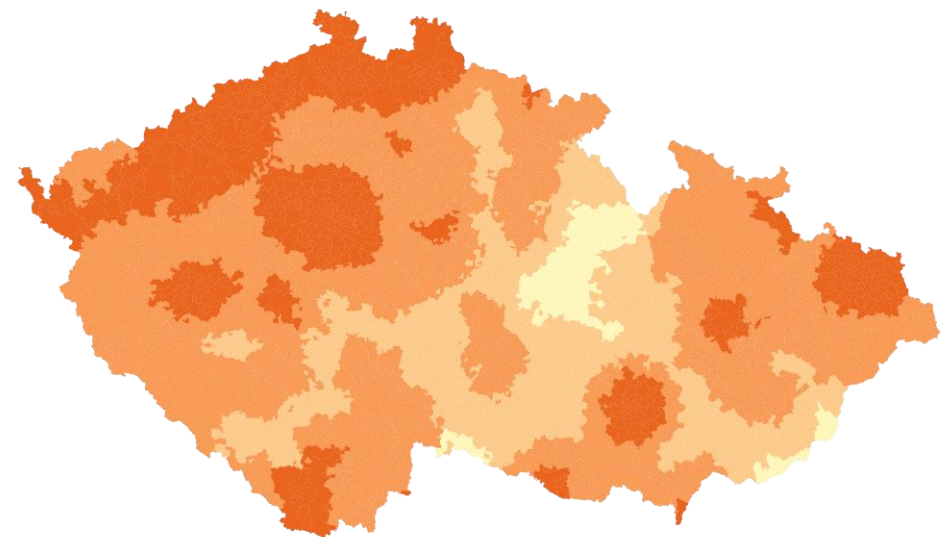


$h = 10km$



$h = 50km$

## 4. Jádrové vyhlazování



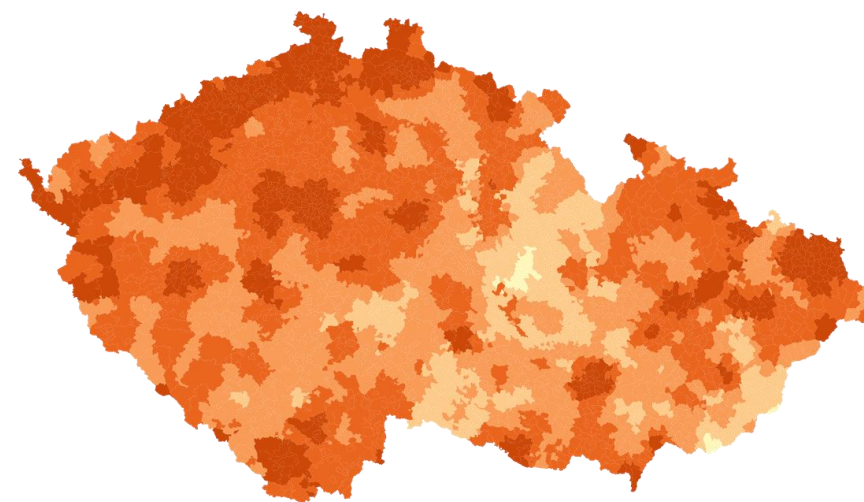
$h = 10km$

# Jádrové vyhlazování (kernel smoothing)

- Nejpřímější forma prostorového vyhlazování
- Odhad v bodě = lokální vážený průměr
  - Pro každý bod  $(x,y)$  se díváme na blízka pozorování
  - Vypočítáme vážený průměr okolních škod
  - Váhy (typicky) klesají s rostoucí vzdáleností
  - Výsledkem je lokálně průměrovaná mapa
- Šířku okolí řídí parametr  $h$  (bandwidth)
  - Intuitivní zápis -  $\hat{\mu}(x, y) = \text{vážený průměr dat v okolí } (x, y)$
  - Váhy

$$w_i = K\left(\frac{d((x, y), (x_i, y_i))}{h}\right) = K\left(\frac{d_i}{h}\right)$$

- $w_i$  – váha dat  $i$
- $K$  – jádrová funkce (kernel)
- $d$  – vzdálenost bodů
- $h$  – šířka vyhlazovacího okna (bandwidth)



# Jádrové vyhlazování (kernel smoothing)

- Formálně je výpočet rizika definován

$$\hat{\mu}(x, y) = \frac{\sum_i K\left(\frac{d_i}{h}\right) Y_i}{\sum_i K\left(\frac{d_i}{h}\right)}$$

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}$$

$$t = \frac{d}{h}$$

- Váhy určuje funkce  $K(t)$  (kernel)
  - $K(t) \geq 0$
  - $K(t)$  nerostoucí pro  $t > 0$
  - $K(t) = K(-t)$

- Kernel  $K(t)$  určuje, jak rychle klesá váha pozorování s rostoucí vzdáleností – určuje pouze tvar
- Existuje více možností, jak kernel zvolit:
  - Gaussovský
    - Vzdálené body mají malou, ale nenulovou váhu

$$K(t) = \exp\left(-\frac{t^2}{2}\right)$$

- Epanechnikov
  - Váha je nulová mimo daný interval, v intervalu je parabolická

$$K(t) = \max(0, 1 - t^2)$$

- Rovnoměrná
  - Všechna pozorování v okruhu mají stejnou váhu

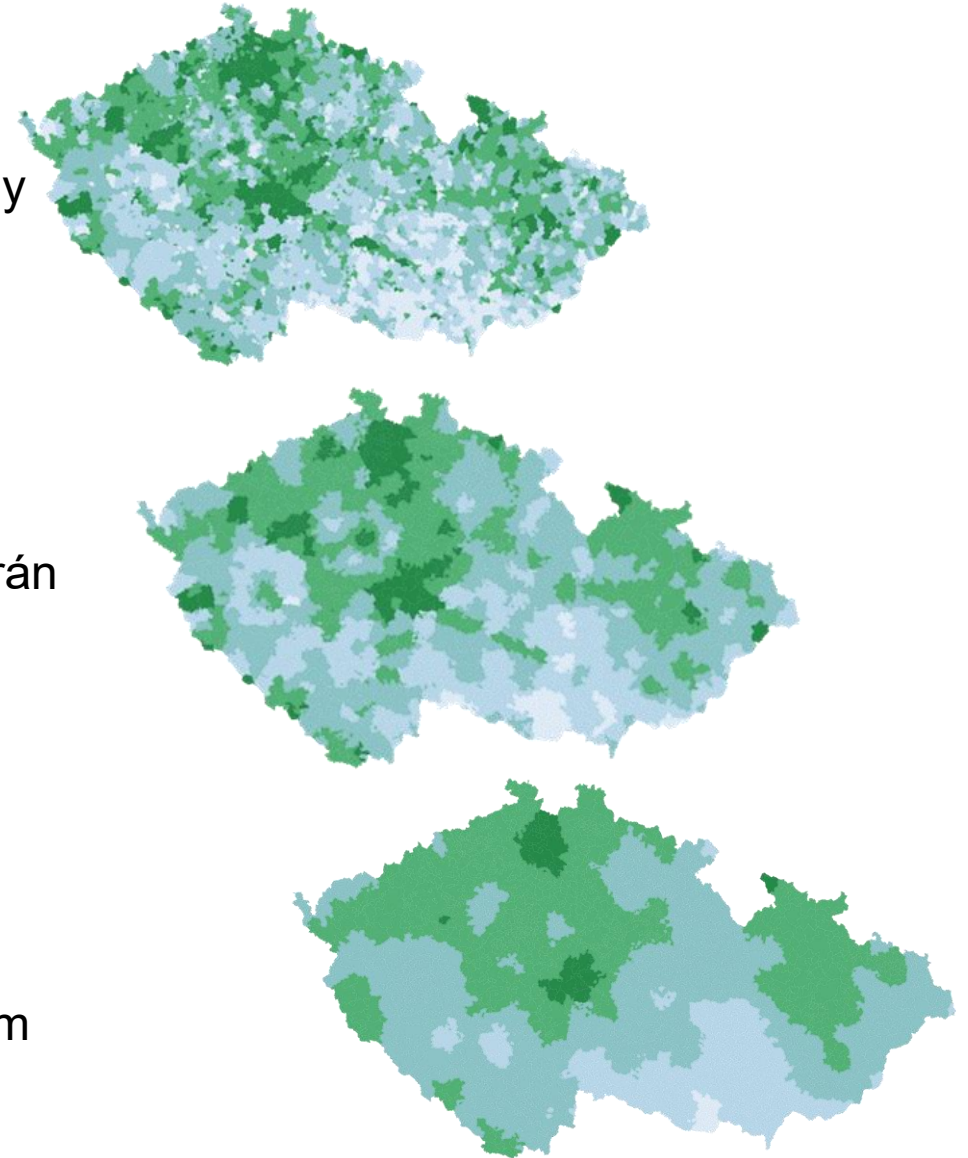
$$K(t) = 1_{\{|t| \leq 1\}}$$

- Výsledek je dominantně ovlivněn parametrem  $h$  (šířka okolí)

# Parametr $h$ (Bandwidth)

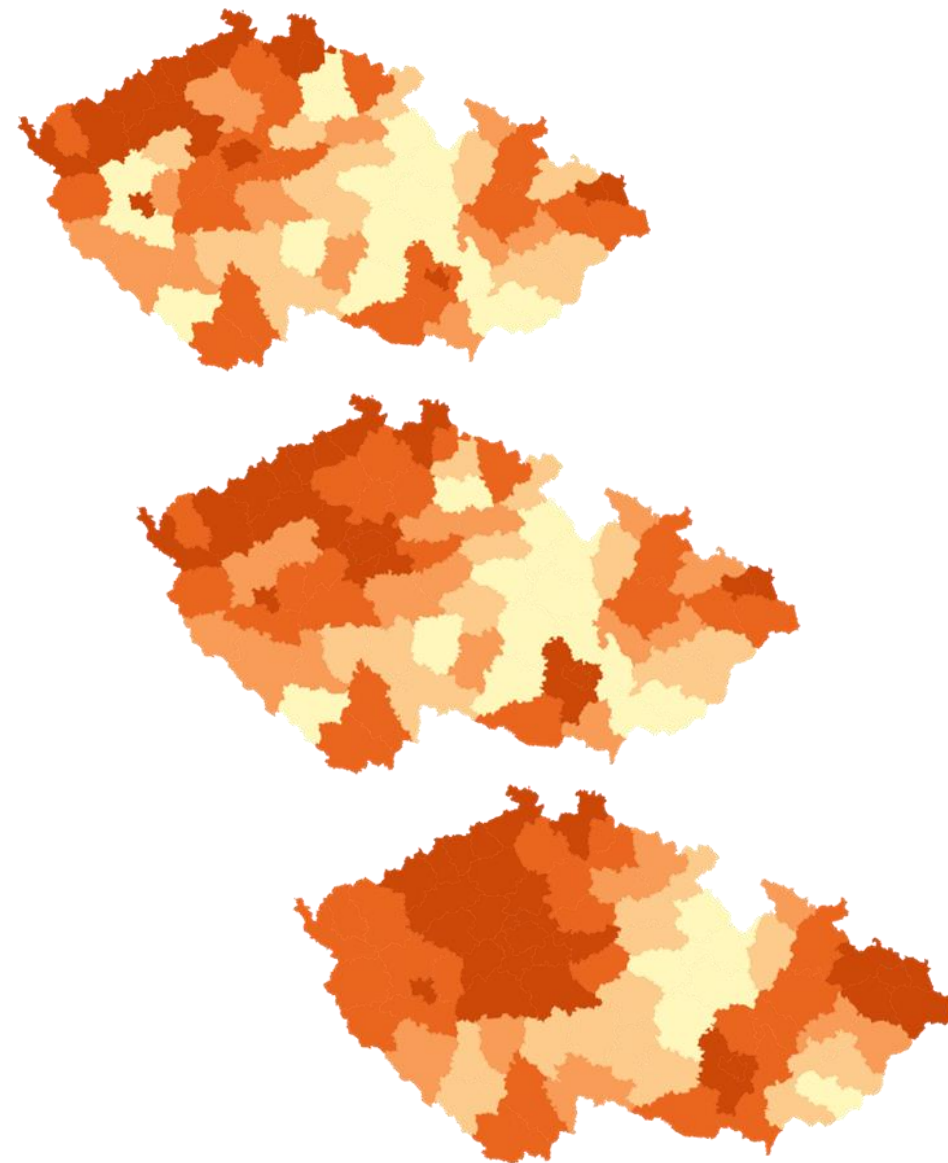
- Bandwidth  $h$  určuje, jak široké okolí je zahrnuto do výpočtu, a tedy  
→ jak hladký nebo detailní bude odhad.
  - Malé  $h$  → detailní povrch, vysoký rozptyl
  - Velké  $h$  → přehlazená mapa
- Volba  $h$  není triviální. Existuje několik možností
  - Cross-validation (nejčastější)
    - Výpočet chyby proběhne pro více hodnot  $h$ , následně je vybrán parametr s nejmenší
  - Heuristiky podle hustoty dat
    - např. menší  $h$  ve městech, větší na venkově
    - Adaptivní jádrové vyhlazování
  - Teoretické vzorce existují, ale v praxi se nepoužívají
- Pro jádrové vyhlazování je volba parametru  $h$  nejdůležitější
- Závisí na ní trade-off mezi systematickou odchylkou a rozptylem

$$\text{syst. odchylka}(\hat{\mu}) \propto h^2, \text{Var}(\hat{\mu}) \propto \frac{1}{nh^2}$$



# Limity jádrového vyhlazování

- Statistická omezení:
  - Ignoruje další ratingové faktory (pracuje jen s polohou)
  - Jeden globální parametr  $h$  je málo flexibilní při nerovnoměrné hustotě dat
  - Výsledek je extrémně citlivý na volbu  $h$
  - Okrajový efekt (Boundary bias) – zkreslení na hranicích
- Výpočetní a praktické limity:
  - Špatně škálovatelná metoda – vysoká výpočetní náročnost u velkých portfolií
  - Nevhodné jako finální tarifní model – spíše nástroj pro ilustraci a průzkum rizika



# Jádrové vyhlazování bez přesných souřadnic

- Není nutné pracovat s přesnými souřadnicemi škod
  - Lze využít aproximace polohy
  - Veškeré záznamy jsou umístěny do těžiště PSČ
- PSČ s vyšší expozicí mají větší vliv
- Ztráta vnitřního detailu
- Stále však jde o výrazně lepší odhad než původní mapa založená na hrubých průměrech PSČ

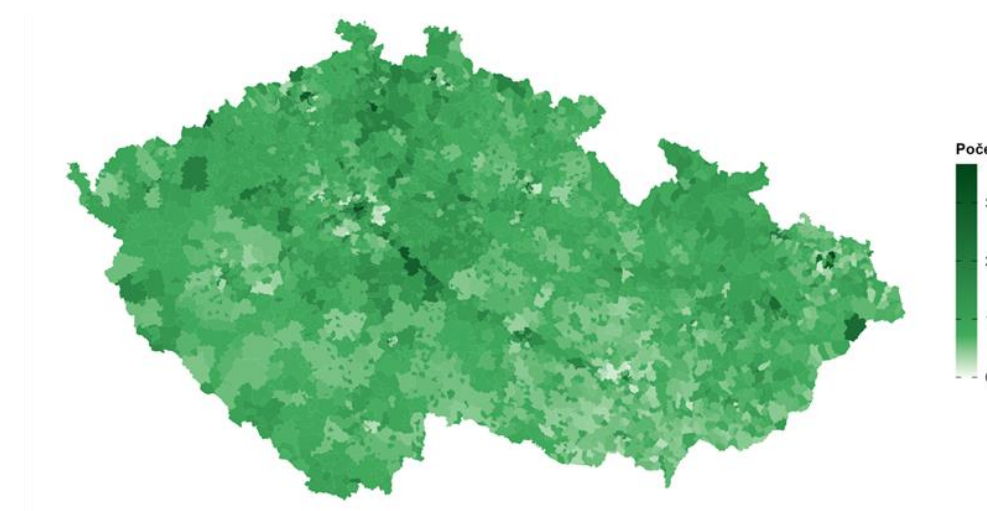
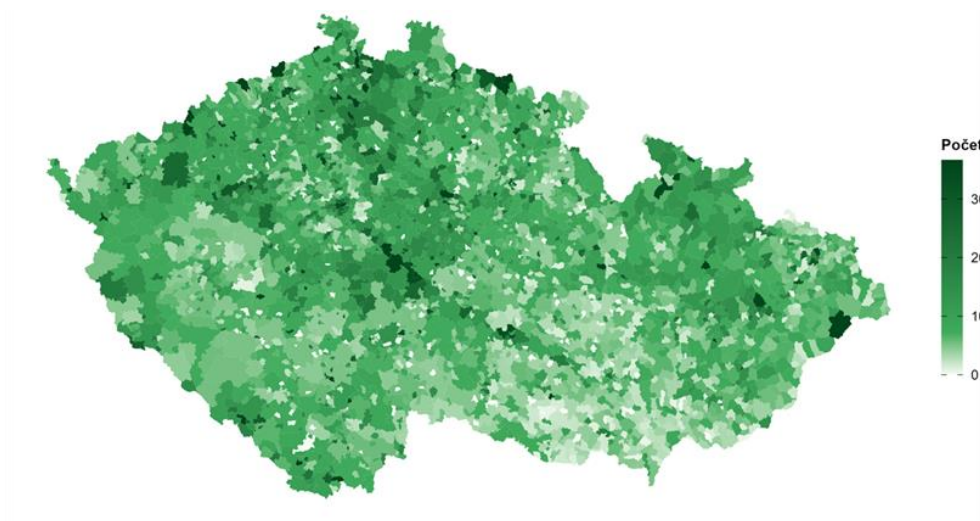
$$\hat{\mu}(x, y) = \frac{\sum_{j \in PSČ} K_h(d_j) \cdot \bar{X}_j \cdot n_j}{\sum_{j \in PSČ} K_h(d_j) \cdot n_j}$$

$\bar{X}_j$  - výběrový průměr

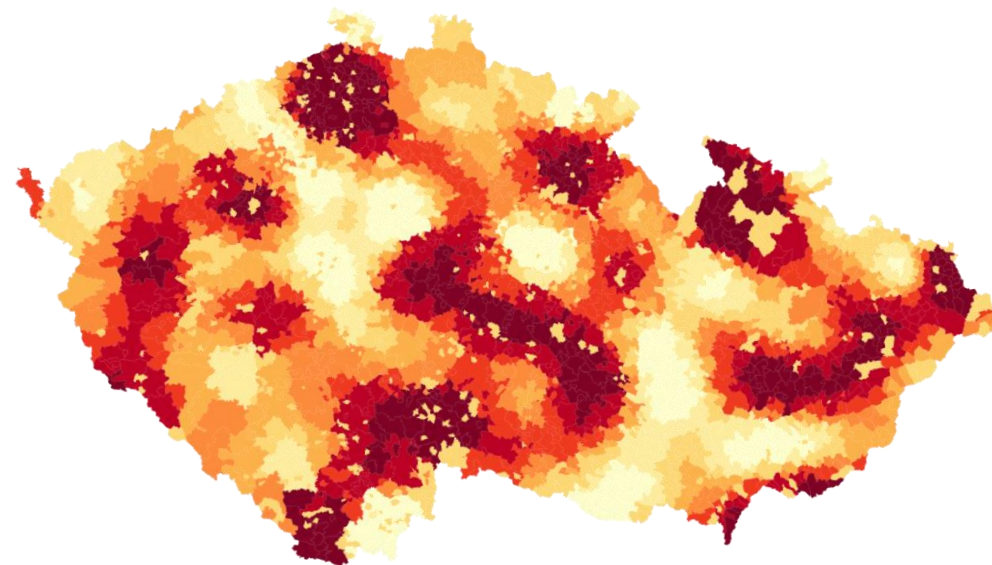
$n_j$  - expozice daného PSČ



# Jádrové vyhlazování – PSČ

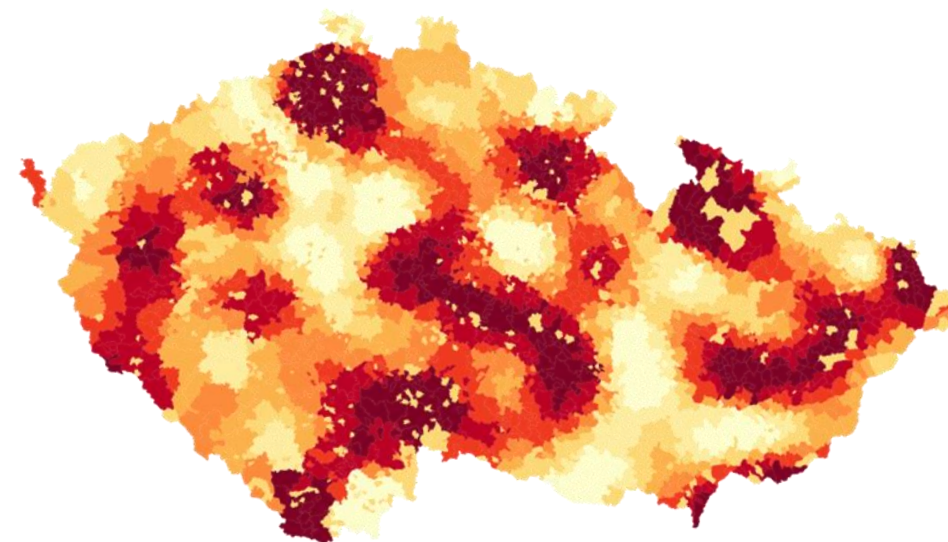


## 5. Zobecněné aditivní modely (GAM)



# Od jádrového vyhlazování ke GAM

- Kernel smoothing poskytuje správnou „výchozí úvahu“:
  - využívá informaci z okolí (geografických sousedů)
  - vytváří hladký prostorový povrch
- Jeho nevýhody však brání použití ve finálním tarifu:
  - nezahrnuje ostatní ratingové faktory
  - Vysoká citlivost modelu na volbu vyhlazovacího parametru
  - Vysoká výpočetní náročnost na velkých portfoliích
- Proto potřebujeme model, který:
  - Integruje prostor současně s ratingovými faktory
  - Automaticky volí hladkost
  - Přirozené rozšíření standardních GLM modelů
  - Vytváří hladký povrch
  - poskytuje stabilní odhady a transparentní statistická kritéria (AIC, GCV)



# GAM: Spojení prostoru a tradičních faktorů

- Tyto požadavky splňuje Generalized Additive Model (GAM)

- $GAM = GLM + \text{hladký prostorový efekt}$

$$g(\mu_i) = \beta_0 + s(x_i, y_i) + f_1(\text{věk}) + \beta_2(\text{typ paliva}) + \dots$$

- Cílem je modelovat plynulou geografickou mapu rizika
- Prostorový efekt je reprezentován hladkou funkcí  $s(x, y)$ :
  - vytváří dvourozměrný povrch
  - zachycuje prostorové trendy
  - potlačuje šum v oblastech s malou expozicí
  - respektuje blízkost pozorování (čím blíž, tím větší vliv na odhad)
- Umožňuje integrovat geografickou informaci přímo do tarifu bez složitých interakcí.

$g(\cdot)$  - link funkce

$f(\cdot)$  - hladká funkce

$\beta$  – regresní koeficient

# Prostorový efekt $s(x, y)$

- Funkce  $s(x, y)$  je složena pomocí bázových funkcí jejichž váhy jsou odhadnuty z dat

- Obecná reprezentace

$$s(x, y) = \sum_{m=1}^M \gamma_m b_m(x, y)$$

- $b_m(x, y)$  – bázové funkce (basis functions)
  - $\gamma_m$  – koeficienty (odhadnuty z dat)
- Vytváří plynulou geografickou mapu rizika
- Parametry jsou odhadovány současně v rámci modelu
- Eliminuje umělé skoky na hranicích regionů
- Jeho hladkost řídí penalizační parametr  $\lambda$

- Penalizace „zvlněné“ funkce
  - Penalizace omezuje, jak moc se tato kombinace bázových funkcí může ohýbat
  - Přidává k odhadu  $s(x, y)$  trest za příliš zakřivený povrch

- Význam bázové reprezentace
  - Umožňuje flexibilně modelovat hladký prostorový povrch
  - Výsledkem je hladká funkce  $s(x, y)$ , která zachycuje skutečné trendy a potlačuje šum.
- Nejčastější volba bází v geografii
  - Thin Plate Regression Splines (TPRS) – standardní volba pro 2D prostory
  - Duchon splines – modernější varianta, lépe škáluje ve větších datech

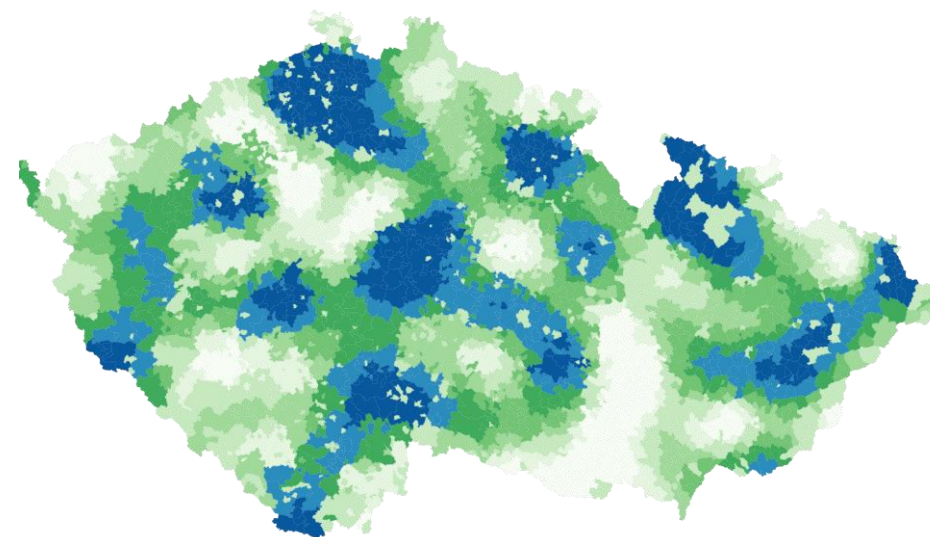
# Penalizovaná věrohodnost (Penalized Likelihood)

- GAM nehledá lokální průměr jako kernel smoothing
- Model optimalizuje celý tvar funkce najednou  $s(x, y)$  – tedy hledá globální optimum

$$\hat{s} = \underset{s}{\operatorname{argmin}} [-2\ell(\beta, s) + \lambda J(s)]$$

- $\ell(\beta, s)$  – *logaritmická věrohodnostní funkce*
- $J(s)$  – míra zakřivení funkce,  $\lambda$  – síla penalizace
- Optimalizace má dvě složky:
  - log-věrohodnost – aby model dobře popisoval data
  - penalizační člen – kontrolující hladkost funkce

$$J(s) = \int \int \left( \left( \frac{\partial^2 s}{\partial x^2} \right)^2 + \left( \frac{\partial^2 s}{\partial y^2} \right)^2 + 2 \left( \frac{\partial^2 s}{\partial x \partial y} \right)^2 \right) dx dy$$



# Měření chyby (Deviance) a volba hladkosti ( $\lambda$ )

- $\lambda$  řídí míru hladkosti prostorové funkce  $s(x, y)$
- Je to analogie k bandwidth parametru v jádrovém vyhlazování
- Parametr  $\lambda$  je určen při trénování modelu, aby vyvážil fit a penalizaci
- Model optimalizuje penalizovanou věrohodnost pomocí
  - GCV (Generalized Cross-Validation)

$$\lambda = \arg \min_{\lambda} \frac{D(\lambda)}{(n - edf(\lambda))^2}$$

- $D(\lambda)$  - míra chyby
- $edf(\lambda)$  - efektivní stupně volnosti (Effective Degrees of Freedom)
- REML (Restricted Maximum Likelihood)
  - Pohlíží na vyhlazování jako na smíšený model (fixed-random effects)
  - $\lambda$  je odhadováno jako poměr rozptylu šumu ku rozptylu signálu.

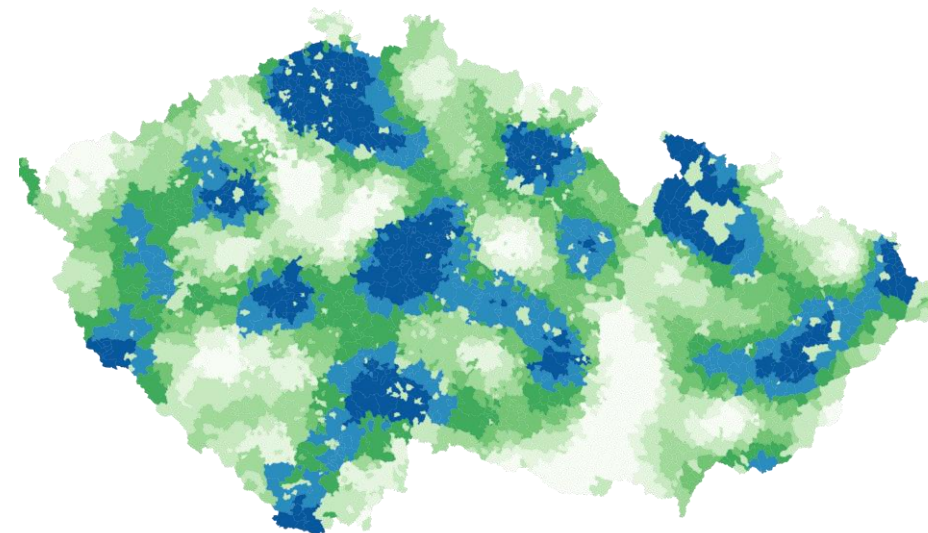
$$\lambda \propto \frac{\sigma_{\text{šumu}}^2}{\sigma_{\text{signal}}^2}, \lambda = \arg \max_{\lambda} REML(\lambda)$$

$$D_{Poisson} = 2 \sum_{i=1}^n \left( y_i \ln \left( \frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right)$$

$$D_{Gamma} = 2 \sum_{i=1}^n \left( -\ln \left( \frac{y_i}{\hat{\mu}_i} \right) + \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i} \right)$$

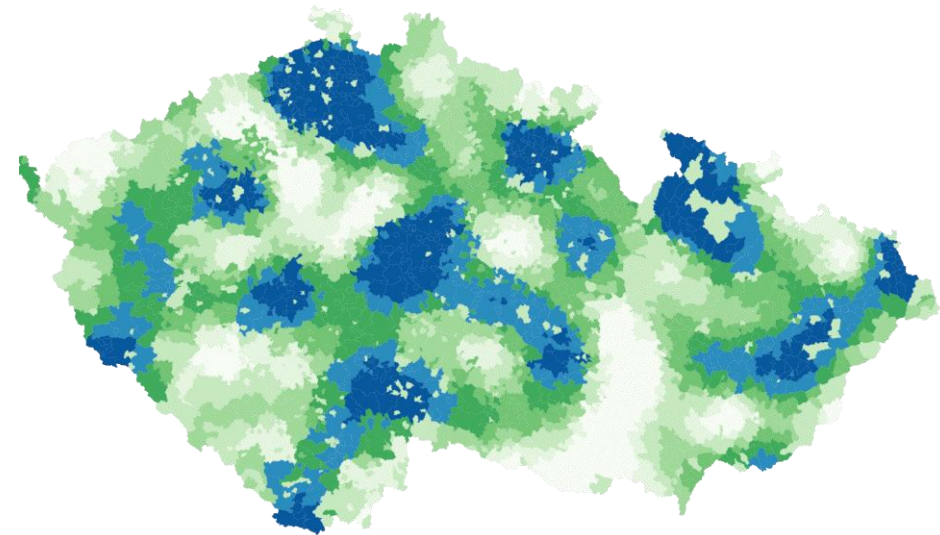
# Klíčové diagnostické metriky

- *edf* (effective degrees of freedom)
  - Míra složitosti prostorové funkce  $s(x, y)$
  - nízké *edf* → příliš hladká mapa (typicky 1–5), hrozí podfitování,
  - střední *edf* → rozumné vyvážení hladkosti a detailu (10–30),
  - vysoké *edf* → příliš flexibilní povrch, riziko přefitování (>40).
- $\lambda$  – síla penalizace zakřivení
  - Určuje, jak moc je penalizováno „ohnutí“ povrchu.
  - vysoké  $\lambda$  → hladký, stabilní povrch,
  - nízké  $\lambda$  → detailní, ale rizikový povrch.
- k-index – kontrola dostatečné velikosti báze
  - k-index < 1 → nedostatečný počet bází →  $s(x, y)$  naráží na limity, nutno zvýšit k.
- p-hodnota
  - Signifikance prostorového efektu
- Porovnání verzí
  - PSČ vs. GAM
  - Jádrové vyhlazování vs. GAM



# Přínos GAM

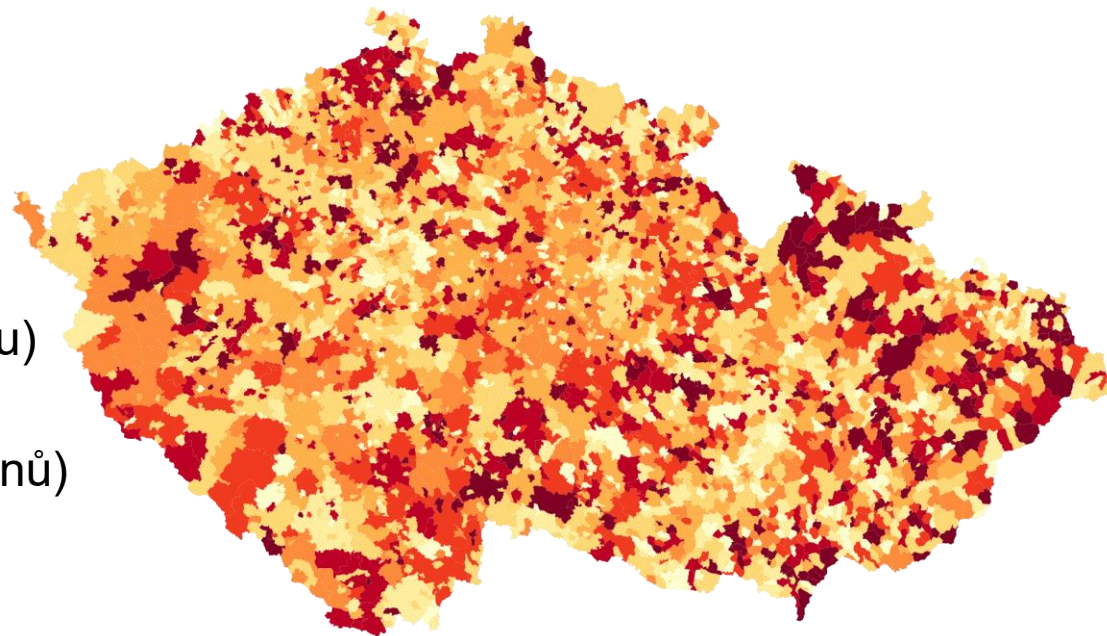
- Zvýšená interpretovatelnost
  - Aditivní struktura zajišťuje transparentnost
  - GAM umožňuje izolovat a vizualizovat efekt každého faktoru nezávisle
  - Výsledná sazba je prostým součtem vlivů. Díky tomu je model auditovatelný a obhajitelný před regulátorem
- Statistický přístup
- Etická a regulační obhajitelnost tarifu
  - Model produkuje spojitou rizikovou plochu – férové tarifování



## 5. Aplikace a dopady

# Validace, interpretace a Business Value GAM modelů

- Statistická validace
  - $edf$ ,  $\lambda$ ,  $k$  – index
  - Význam prostorového efektu
- Geografická validace
  - Stabilita v oblastech s nízkou kredibilitou
  - Absence umělých hran
  - Konzistence se sousedními oblastmi (kontinuita trendu)
- Obchodní validace
  - Dopad na portfolio (relativní změna sazeb podle regionů)
  - Srovnání se známými rizikovými oblastmi
- Porovnání verzí
  - PSČ vs. GAM
  - Jádrové vyhlazování vs. GAM



# Finální srovnání

- PSČ
  - Po částech konstantní mapa → ostré hranice
  - Vysoká nestabilita v malých objemech
  - Nevhodné jako finální geografická vrstva
- Teorie kredibility
  - Stabilizuje odhady při malých objemech
  - Neřeší prostorovou blízkost ani skutečný tvar rizika
  - Pouze zmírňuje problém

# Finální srovnání

- Jádrové vyhlazování
  - Plynulá mapa, intuitivní přístup
  - Citlivé na parametr  $h$ , špatně škálovatelné
  - Neumí kombinovat prostor + ratingové faktory
- GAM
  - Plynulá, stabilní a realistická mapa rizika
  - Automatická volba hladkosti  $\lambda$  (GCV/REML)
  - Kombinace prostoru + ratingových faktorů v rámci GLM
  - Nejvhodnější metoda pro tarifní modelování

Děkuji za pozornost.

# Apendix - zdroje

**Wood, S. N.** (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). CRC Press.

**Frees, E. W.** (2010). *Regression Modeling with Actuarial and Financial Applications*. Cambridge University Press.

**Bivand, R. S., Pebesma, E. J., & Gomez-Rubio, V.** (2013). *Applied Spatial Data Analysis with R* (2nd ed.). Springer.

- R balíček RCzechia - [github.com/jlacko/RCzechia](https://github.com/jlacko/RCzechia)
- Kriminalita - [kriminalita.policie.gov.cz](https://kriminalita.policie.gov.cz)
- Dopravní nehody - [nehody.policie.gov.cz/](https://nehody.policie.gov.cz/)
- Počet obyvatel - [csu.gov.cz](https://csu.gov.cz)