

Statistická syntéza indikátorů kvality života – návrh tvorby indexu v evropských regionech

Karel MACKŮ, Vít VOŽENÍLEK

Statistical synthesis of quality of life indicators – design of index construction in European regions

Abstract: *The quality of life is a complex theme closely linked with human existence. Both scientific and international political organizations have been focusing on this topic in recent decades. This research follows the European Union's challenges to deal with the quality of life as a multi-aspect phenomenon. Firstly, the authors' concept of quality of life, based on key aspects and their respective indicators is introduced. To assess the quality of life multidimensionally and comprehensively, the issue of data aggregation is addressed. Principal Component Analysis (PCA) is a statistical method used to reduce the dimensionality of the data and can also be used to construct a synthetic index. In its basic form, this method is designed for non-spatial data and does not take into account potential spatial relationships. However, geographically weighted PCA can capture them. Simple PCA, robust PCA, simple geographically weighted PCA, and robust geographically weighted PCA are tested on the quality of life data. The results of the paper describe differences in selected methods and the possibilities of their application in the regional evaluation of the quality of life in Europe. Using the PCA, a synthetic quality of life index is constructed based on the defined quality of life indicators on the pan-European scale at the regional level defined by NUTS 2 classification. Paper, compared to the existing studies, analyses the topic in a broader spatial scope and detail. The results do not bring innovation in terms of the used methods, but mean a challenge in the interpretative level.*

Keywords: *quality of life, Principal Component Analysis, Europe, synthesis, statistics*

Úvod do problematiky

Kvalita života je rozsáhlé, s lidskou existencí úzce propojené téma, kterému se věnuje řada vědních oborů, a to společenských, lékařských i přírodních. Po překonání průkopnického období formování konceptu kvality života (první ucelené publikace vydávají Smith (1973), Campbell, Converse a Rodgers (1976) a Andrews (1986)) se přibližně od 80. let 20. století vědecké výzkumy snaží kvalitu života lépe definovat a odvodit metody jejího měření. Důvodem je využití hodnocení kvality života jako nástroje státní správy a organizací pro plánování, rozhodování, identifikace problematických oblastí či přerozdělování financí pro místní rozvoj. Odborníci se shodují, že téma kvality života se skládá z velkého množství aspektů, které se snaží pokrýt dílčí části života. Bohužel, kvůli širokému chápání tématu zatím po dobu výzkumu nedošlo ke shodě na jednotné definici. V jedné ze starších prací toto konstatuje Andrews (1986), o tři desetiletí později se s Andrewsem shodují Morais a Camanho (2011).

Hodnocení kvality života se věnuje také Evropská unie v iniciativě Evropské komise *Beyond the GDP – Measuring progress in changing world* (European Commission 2009). Hlavní myšlenkou iniciativy je posunout kvalitu života za hranice jejího staršího konceptu, kdy byla chápána pouze ve smyslu ekonomické síly. V kontextu této iniciativy vznikla v roce 2008 za podpory francouzské vlády *Commission on the Measurement of Economic Performance and Social Progress* (CMEPSP).

DOI: <https://doi.org/10.33542/GC2019-2-06>

Jejím cílem bylo identifikovat nedostatky ve sledování rozvoje společnosti založeného pouze na hrubém domácím produktu (HDP) a navrhnout doporučení pro nové hodnocení kvality života v kontextu udržitelného rozvoje, pokrývajícího širší aspekty života. V roce 2009 představila komise dokument *GDP and beyond – measuring progress in a changing world* (European Commission 2009), ve kterém shrnula pět střednědobých opatření, která mají podpořit sledování rozvoje společnosti v současném světě. Jedním z nich je sledování kvality života a well-beingu obyvatel.

Na základě aktivity iniciativy *Beyond the GDP* vznikla řada doporučení pro nakládání s pojmem kvality života. Hlavními myšlenkami iniciativy formulovaných v reportu *Report of Commission on the Measurement of Economic Performance and Social Progress* (Stiglitz, Sen a Fitoussi 2009) jsou zejména:

- důraz na domácnosti a jejich příjmy
- stanovení nových datových zdrojů pro sledování kvality života
- hlubší a propracovanější analýzy
- multidimenzionální hodnocení

Multidimenzionální hodnocení kvality života (chápano jako zahrnující více aspektů) je klíčové, aby se ke kvalitě života přistupovalo jako ke komplexnímu pojmu, skládajícímu se z více dílčích aspektů, které hrají v životě svou roli. O agregaci dat hovoří *Report of the Task Force: Multidimensional measurement of the quality of life* (Eurostat and INSEE 2011), v zájmu snížení složitosti sledovaného tématu nabádá ke konstrukci syntetického indikátoru pro jednotlivé aspekty (Eurostat a INSEE 2011). Podobně zpráva *Quality of life: fact and views* (Eurostat 2015) pojímá multidimensionalitu pouze ve smyslu analytického zpracování několika dílčích témat. Stále zde chybí syntéza těchto aspektů do jedné komplexní veličiny.

Během výzkumu v posledních desetiletích se jako hlavní měřicí nástroj komplexních témat ustálilo využívání agregovaných číselných ukazatelů – indexů. Index je bezrozměrný ukazatel, snadno vnímatelný nebo porovnatelný, obsahující složitou informaci, jež je výsledkem syntézy. Proto je velmi důležitý způsob konstrukce indexu. Na druhou stranu index neumožňuje pochopení širších souvislostí a podstaty problému. Proto je nezbytné znát teoretický koncept konstrukce pro pochopení celého problému, vnitřní struktury sledovaného tématu a vztahů dílčích indikátorů. Tvorbou kompozitních indexů se ve svých pracích zabývají Mederly, Topercer a Nováček 2004, Bérenger a Verdier-Chouchane 2007, Martín a Mendoza 2013, Greyling a Tregenna 2016 a další. Řada autorů využívá pro agregaci dat pouze jednoduchých metod, například v podobě aritmetických operací prováděných nad standardizovanými vstupními daty, dávající všem vstupním indikátorům stejnou váhu. Tento nedostatek může být odstraněn odborným posouzením významnosti jednotlivých indikátorů, k čemuž je však vhodné využít znalostí tematiky, a v případě komplexního tématu, kterým je kvalita života, je hodnocení významnosti indikátorů náročný úkol. S cílem vyhnout se těmto složitým procesům hledají autoři alternativní řešení v oblasti metod vícerozměrné statistiky. V dostupné literatuře je jako metoda sestavení syntetického indexu často použita analýza hlavních komponent (PCA). V kontextu hodnocení kvality života ji používá Lo a Faber (1997), v různých modifikacích dále Li a Weng (2007) nebo Rao et al. (2012).

Popsané aktivity Evropské Unie definují nové metody a indikátory pro monitorování a evaluaci kvality života. Navržené indikátory pro sledování vybraných aspektů kvality života jsou zatím dostupné pouze na národních úrovních. Vystává zde výzva vyzkoušet možnosti komplexního hodnocení kvality života založené na objektivních, snadno dostupných statistických indikátorech. Tato myšlenka formuluje výzkumný cíl tohoto článku: vyzkoušet využitelnost analýzy hlavních komponent a jejich variací jako nástroje komplexního hodnocení kvality života, použitelného v celoevropském rozsahu na regionální úrovni.

Metody

Použitá data

Pro hodnocení kvality života byla již v předchozí fázi studie sestavena datová sada Quality of Life Dataset (QoLD). Na základě vybraných prací (např. Smith (1972), Bérenger a Verdier-Chouchane (2007), Hardeman a Dijkstra (2014), Eurostat (2015) a další) byly vymezeny nejčastěji se vyskytující aspekty kvality života, které byly označeny jako jádrové: ekonomická síla a materiální zabezpečení, zdraví, sociální prostředí, vzdělání a životní prostředí. Následně byly hledány takové indikátory, které tyto aspekty vhodně popisují. Podrobnost dat byla zvolena úrovní NUTS 2. Na této klasifikaci existuje problém s celistvostí dostupných dat – mnoho zajímavých indikátorů není na této podrobnosti dostupné, často nejsou dostupné aktuální hodnoty, anebo nejsou datové zdroje územně celistvé. Proto musí být na regionální úrovni indikátorů upravena v závislosti na dostupnosti regionálních statistických dat.

Výslednou datovou sadu indikátorů Quality of Life Dataset (QoLD) tvoří pět jádrových aspektů a jim odpovídajících 13 indikátorů. Použitá data shrnuje Tab. 1 (podrobnější popis vybraných indikátorů je představen v Tab. 2). Zdroji byly převážně databáze Eurostat a OECD Regional Database. Chybějící hodnoty byly doplněny dohledáním u národních statistických úřadů, v minimu případů byly chybějící hodnoty imputovány jako průměr z předchozích let. Dataset QoLD tvoří data v prostorovém rozsahu států EU 28 na úrovni NUTS 2, rozšířené o Norsko, Švýcarsko, Island, Severní Makedonii a Černou Horu. Balkánské země (Srbsko, Bosna a Hercegovina, Kosovo, Albánie, Turecko) a většina ostrovních a zámořských regionů byly pro velké množství chybějících údajů vypuštěny. Posledním rokem, ke kterému splňovala všechna data požadavky na úplnost, byl rok 2013.

Tab. 1. Přehled aspektů kvality života a jejich indikátorů (QoLD)

Aspekt	Indikátor	Zdroj
Ekonomická síla a materiální zabezpečení	HDP na obyvatele	Eurostat, OECD
	Průměrný příjem domácností	Eurostat, OECD
	Míra dlouhodobé zaměstnanosti	Eurostat
	Index ekonomického zatížení ¹	Eurostat
Zdraví	Střední délka života při narození	Eurostat
	Kojenecká životnost	Eurostat
Sociální prostředí	Počet lékařů (na 10 000 obyv.)	Eurostat, národní statistické úřady
	Kapacita nemocnic	Eurostat, národní statistické úřady
	Index stáří ²	Eurostat
Vzdělání	Podíl terciárně vzdělaných obyvatel	Eurostat
	NEET ³	Eurostat
Životní prostředí	Koncentrace ozónu (SOMO35)	European Environmental Agency
	Koncentrace částic PM2.5	European Environmental Agency

¹ Počet dětí ve věku 0 – 14 a osob ve věku 65 let a více připadající na 100 ekonomicky aktivních osob

² Počet osob ve věku 65 let a více připadající na 100 osob ve věku 0 – 14 let

³ NEET – mladí lidé věku 15 – 24, kteří jsou nezaměstnaní, nestudují ani se neúčastní jiného procesu připravující je na budoucí zaměstnaní

Tab. 2. Použité indikátory a jejich popis

Indikátor	Jednotka	Popis indikátoru
HDP na obyvatele	Euro PPS na obyvatele	Hrubý domácí produkt na obyvatele vyjádřený v PPS (standardy parity kupní síly)
Průměrný příjem domácnosti	Euro PPS na obyvatele	Čistý disponibilní příjem domácnosti: celkový příjem domácnosti po zdanění a dalších sociálních odpočtech, který je dostupný pro útraty a úspory. Pro zohlednění velikosti domácnosti je příjem podělený počtem členů domácnosti převedených na dospělé osoby. Hodnoty zohledňují paritu kupní síly jednotlivých zemí.
Míra dlouhodobé zaměstnanosti	%	Podíl dlouhodobě nezaměstnanosti je podíl nezaměstnaných osob od 12 měsíců a více v celkové ekonomicky aktivní populaci vyjádřený v procentech. Doba nezaměstnanosti je definována jako doba hledání zaměstnání nebo doba od posledního zaměstnání. Pro snadnější interpretaci vlivu indikátoru na komponenty byla nezaměstnanost nahrazena zaměstnaností, vypočtenou jako: 100 – míra nezaměstnanosti.
Index ekonomického zatížení	číselný poměr	Počet dětí ve věku 0 – 14 a osob ve věku 65 let a více připadající na 100 ekonomicky aktivních osob
Střední délka života při narození	rok	počet roků, kterých se pravděpodobně dožije narozená osoba
Kojenecká životnost	‰	Kojenecká úmrtnost je definována jako počet zemřelých do 1 roku života na 1 000 živě narozených v daném roce. Pro snadnější interpretaci vlivu indikátoru na komponenty byla použita modifikace nazvaná jako kojenecká životnost, hodnota vypočtena jako: 1 000 – hodnota kojenecké úmrtnosti
Počet lékařů	číselný poměr	Počet lékařů na 10 000 obyvatele
Kapacita nemocnic	číselný poměr	Počet lůžek v nemocničních zařízeních na 10 000 obyvatele
Index stáří	číselný poměr	Počet osob ve věku 65 let a více připadající na 100 osob ve věku 0 – 14 let
Podíl terciárně vzdělaných obyvatele	%	Ukazatel definován jako procento populace ve věku 25 – 64 let, která úspěšně ukončila terciární studium. Toto vzdělání se vztahuje na úroveň ISCED (Mezinárodní standardní klasifikace vzdělávání) 2011, úroveň 5 – 8 pro data od roku 2014 a na úroveň ISCED 1997 pro úroveň 5 – 6 pro data do roku 2013.
NEET	%	Neither in Employment nor in Education and Training – procento mladých lidí ve věku 15 – 24, kteří jsou nezaměstnaní, nestudují ani se neúčastní jiného procesu připravující je na budoucí zaměstnání
Koncentrace ozónu (SOMO35)	µg.m-3 . den	Roční průměr sum maximálních denních 8-hodinových koncentrací, větších než prahová hodnota 70 µg.m-3
Koncentrace částic PM2.5	µg.m-3	Roční průměrné hodnoty získané interpolací z měřících stanic

Detekce extrémních hodnot

Reálná data často obsahují extrémní hodnoty, které se výrazně liší od zbytku datového souboru. Ty mohou vznikat chybami při pořizování dat anebo popisují záznamy, které jsou svými hodnotami mimořádné a zasluhují větší pozornost. Jelikož extrémní hodnoty výrazně ovlivňují výsledky a kvalitu analýz, je nutné tyto pozorování identifikovat a zvážit možnosti zacházení s nimi. Je-li požadavkem využití kompletního datového souboru za účelem popisu skutečného stavu a jsou-li extrémní hodnoty považovány za věrohodné, pak je vhodné je do analýzy zahrnout (Dixon 1950). Proti odstraňování extrémních hodnot argumentují také Militký a Meloun (2003): data jsou odstraněním extrémních hodnot upravena tak, aby více vyhovovala předkládanému modelu, a nelze tedy dostatečně posoudit jeho vhodnost; variabilita redukovaného souboru bývá nízká, což není vhodné pro případné další hodnocení modelu.

Extrémní hodnoty mohou být identifikovány pomocí Mahalanobisovy vzdálenosti, která měří vzdálenost daného záznamu pozorování od výběrového centroidu. Vzdálenost popisuje, jak moc se jednotlivé záznamy liší od průměrného vektoru v definovaném d -rozměrném prostoru (d odpovídá počtu sledovaných atributů) (Mahalanobis 1936). Samotný výpočet vzdálenosti extrémní hodnoty neidentifikuje, musí být nadefinována prahová hodnota. Jak uvádí Hubert a Debruyne (2010), tato hodnota může být rovna odmocnině 97.5 kvantilu chí kvadrátu rozdělení s d stupni volnosti (kde d odpovídá počtu sledovaných atributů). Zároveň je možné počítat vzdálenost robustní metodou, která není při identifikaci extrémních hodnot ovlivněna extrémními hodnotami samotnými, její použití dává tedy logicky větší smysl.

Analýza hlavních komponent

Analýza hlavních komponent (PCA) je jedna z často využívaných metod zpracování více-rozměrných dat sloužící k extrakci proměnných (Jolliffe 2002). Dokáže snížit rozměr zpracovávaného souboru a hledat nové, skryté vlastnosti (tzv. komponenty), vznikající na základě závislosti vstupních dat. Hlavní myšlenkou PCA je vyhledávat nové komponenty ve směru největšího rozptylu v datech. V případě, že data zahrnují extrémní hodnoty, bude tento směr patrně zkreslený, komponenty jsou přitahovány ve směru nejvýznamnějších extrémních hodnot a nevystihují skutečný trend v datech. Proto je nutné použít robustní metody (popsáno např. v Hubert, Rousseeuw a Vanden Branden (2005)), které dokáží přítomnost extrémních hodnot zohlednit.

V mnoha geografických studiích je PCA aplikována bez ohledu na prostorovou informaci, čímž poskytuje neúplné pochopení daného procesu (Harris et al. 2015). Prostorově vážená PCA (GW PCA) je lokalizovaná verze PCA, která je nástrojem pro zkoumání prostorové heterogenity ve struktuře mnohorozměrných dat (Harris, Brunson a Charlton 2011). Při výpočtu je pro každý prvek vypočtena samostatná PCA, která nezahrnuje všechna data (jako globální model), ale pouze určitý počet prostorově vážených prvků vybraných na základě předem definovaného okolí. Tento přístup umožňuje monitorovat lokální variabilitu sledovaného jevu.

Pro zpracování datasetu QoLD byly postupně otestovány vybrané kombinace nastavení, jak doporučuje (Harris et al. 2015) – byla vypočtena jednoduchá a robustní GW PCA, pro jádrové funkce byly testovány možnosti bi-square, gaussian, boxcar, exponential a tricube. Pro každé toto nastavení byla ohodnocena míra vhodnosti modelu na základě cross-validace, podle které lze následně vybrat optimální nastavení modelu. Pro oba případy (jednoduchá i robustní PCA) se jevil jako nejvhodnější model požívající bi-square jádrovou funkci, která snižuje váhu okolí s rostoucí vzdáleností, což se projevuje přirozeným shlazením výsledků. Zatímco pro jednoduchou PCA byl optimální počet sousedů 273, což v případě 283 záznamů datového souboru zahrnuje téměř všechny územní jednotky, pro robustní PCA bylo doporučeno adaptivní sousedství o počtu 60 prvků.

Komponenty vypočítané PCA z indikátorů QoLD lze použít pro sestavení indexu kvality života. Každá nová komponenta reprezentuje nově odhalený aspekt kvality života odvozený matematickými postupy přímo z původních dat. Spojení komponent do jednoho indexu pak

poskytuje kompletní pohled na téma kvality života. Pro výběr vhodného počtu komponent bylo využito Kaiser kritérium, které popisuje, kolik procent variability z původních dat nové komponenty vysvětlují (Kaiser 1960). U každé komponenty bylo ohodnoceno, zdali má pozitivní či negativní přínos na kvalitu života. Agregace byla provedena váženým součtem, váha byla komponentě přiřazena podle množství vysvětleného rozptylu původních dat:

$$QoL = \sum_{i=1}^n W_i F_i \quad (1)$$

kde:

QoL – hodnota indexu kvality života

W_i – váha komponenty i

F_i – hodnota komponenty i

n – počet hlavních komponent

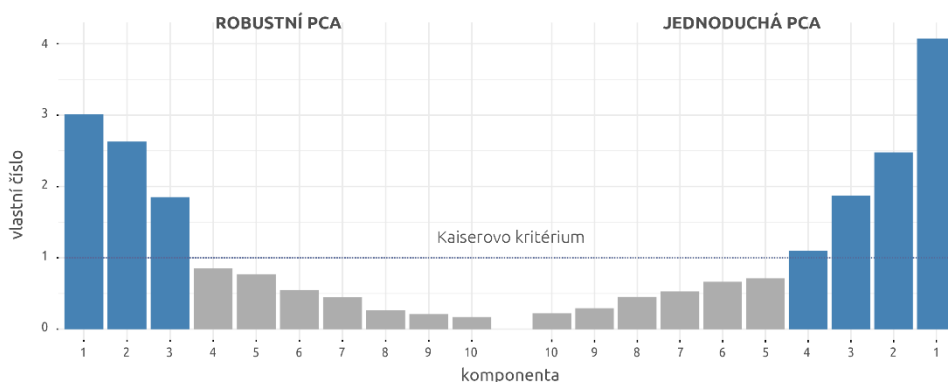
Použitý software

Statistické výpočty byly realizovány v prostředí software *R*, pro detekování extrémních hodnot byly využity balíčky *cellWise*, pro výpočet PCA balíčky *factoextra*, *robustbase* a *GWmodel*. Neprostorové vizualizace byly provedeny s využitím metod balíčku *ggplot2* a *corrplot*. Zpracování prostorových dat bylo řešeno v *ArcGIS Pro*.

Výsledky

Syntéza dat

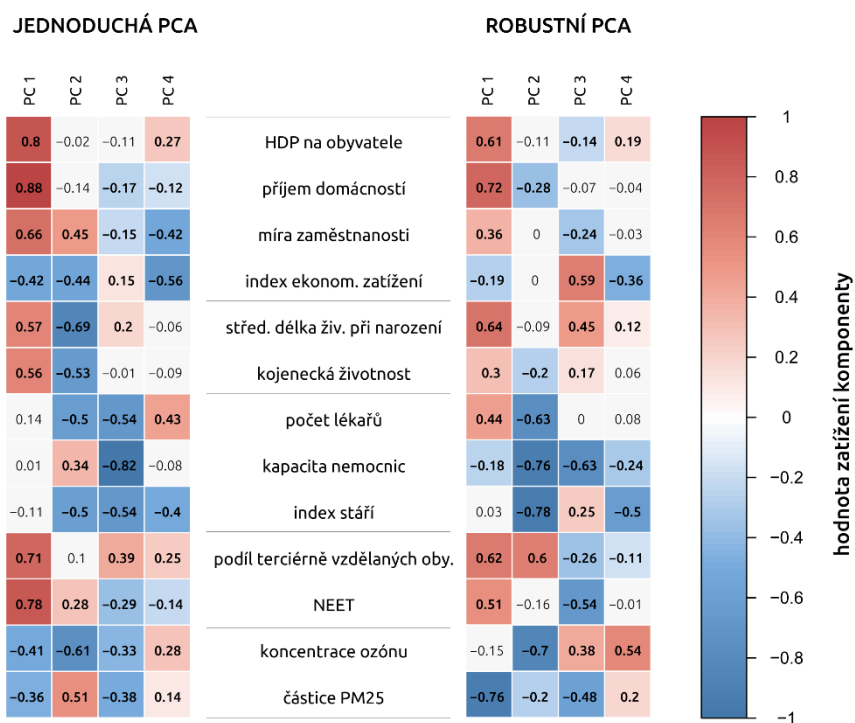
Analýza extrémních hodnot pomocí Mahalanobisovy vzdálenosti odhalila řadu pozorování, která lze považovat za extrémní. Kritická hranice Mahalanobisovy vzdálenosti byla stanovena hodnotou 21.3 (hodnota vypočtena přístupem odmocniny 97.5 kvantilu chí kvadrátu rozdělení s d stupni volnosti). Jako extrémní hodnota bylo takto identifikováno 21 regionů, což představuje přibližně 7 % z celého datového souboru. Detailnější rozbor identifikoval indikátory, které k vybočení nejčastěji přispívají – jsou jimi především index stáří, dlouhodobá zaměstnanost, střední délka života při narození a kojenecká životnost. Jako nejvíce vybočující byly analýzou označeny britský region *Inner London – West* (extrémně vysoké ekonomické indikátory) a francouzský *Corse* (vliv vysoké hodnoty indexu ekonomického zatížení a nízké hodnoty zaměstnanosti).



Obr. 1. Hodnoty vlastních čísel komponent pro stanovení jejich významnosti

Otestovány byly jak jednoduchá PCA, tak její robustní varianta (založena na odhadu Minimum Covariance Determinant). Vzhledem ke zjištěnému charakteru dat se věrohodnější výsledky očekávají právě u robustní PCA. Prvním krokem výpočtu je výběr počtu přijatých komponent. Při využití Kaiser kritéria byly u klasické PCA přijaté 4 komponenty s mírou vysvětleného rozptylu 74 %, v případě robustní PCA 3 komponenty s 68 %. Robustní přístup sice popisuje menší míru celkového rozptylu, nicméně 3 komponenty jsou vhodnější pro následnou interpretaci. Pokud by byl v obou přístupech zahrnut stejný počet komponent, daří se robustní metodou vysvětlit o cca 3 % více rozptylu ve srovnání s jednoduchou PCA. Rozdíly ve vypočtených zátěžích komponent popisuje Tab. 2. Je patrné, že na první komponentě není mezi klasickou a robustní PCA významný rozdíl. Druhé a třetí komponenty se již liší více. Nejdramatičtější rozdíl je na druhé komponentě, na třetí je více patrný trend podobnosti některých indikátorů. Lze tedy potvrdit, že extrémní hodnoty výrazně ovlivňují podobu modelu popsaneho pomocí PCA.

Tab. 3. Zatížení indikátorů kvality života na komponenty



Díky interpretaci byla každá komponenta vypočtena robustní PCA ohodnocena jako pozitivní nebo negativní, což je důležité pro sumární výpočet syntetického indexu kvality života QoL:

Komponenta 1 (27,5 % vysvětleného rozptylu v datech): do této komponenty pozitivně přispívají indikátory ekonomického aspektu (příjem, HDP na obyvatele, míra zaměstnanosti) a oba indikátory aspektu vzdělání s hodnotami zatížení komponent nad 0.5 (Tab. 3). Pozitivní vliv mají také oba indikátory aspektu zdraví (střední délka života při narození a kojenecká životnost). Komponentu lze charakterizovat jako míru ekonomické rozvinutosti s důrazem na vzdělání. Odráží se v ní pozitivní vliv bohatství, které bylo základem tradičního konceptu kvality života v minulém století, a vzdělání, které je v současnosti obecně přijímáno jako cesta k rozvoji udržitelné společnosti. Z těchto důvodů je komponenta pro závěrečnou sumarizaci ohodnocena pozitivně a nazvána *lidský kapitál*. Prostorovou vizualizaci komponentního skóre první komponenty zobrazuje Obr. 2a.

Komponenta 2 (23,9 % vysvětleného rozptylu v datech): druhá komponenta je nejvýrazněji negativně ovlivněna všemi indikátory sociálního aspektu s hodnotami zatížení komponenty pod -0.62 (Tab. 3) a oběma environmentálními indikátory. Významný pozitivní vliv zaznamenává podíl terciárně vzdělaného obyvatelstva. Interpretace druhé komponenty může být kontroverzní – vliv dostupnosti zdravotních služeb kontrastuje s indexem stárí, koncentrací škodlivého ozónu a podílem terciárně vzdělaného obyvatelstva. Jelikož nepřímá úměra negativních zatížení komponenty převažuje, nižší hodnoty komponentního skóre jsou považovány jako horší v kontextu kvality života, komponenta byla pro závěrečnou sumarizaci označena jako pozitivní a označena jako *sociální udržitelnost s příznivým vlivem životního prostředí*. Prostorovou vizualizaci komponentního skóre třetí komponenty zobrazuje Obr. 2b.

Komponenta 3 (16,8 % vysvětleného rozptylu v datech): ve třetí komponentě hraje nejvýznamnější roli s pozitivním vlivem index ekonomického zatížení, střední délka života při narození a koncentrace ozónu, negativně přispívá kapacita nemocnic, koncentrace částic PM2.5 a oba indikátory vzdělání. Opět lze sledovat míru nejednoznačnosti, nicméně pro závěrečnou sumarizaci byla komponenta nastavena jako negativní a nazvána *pasivní populace s ekonomickým zatížením*. Prostorovou vizualizaci komponentního skóre třetí komponenty zobrazuje Obr. 2c.

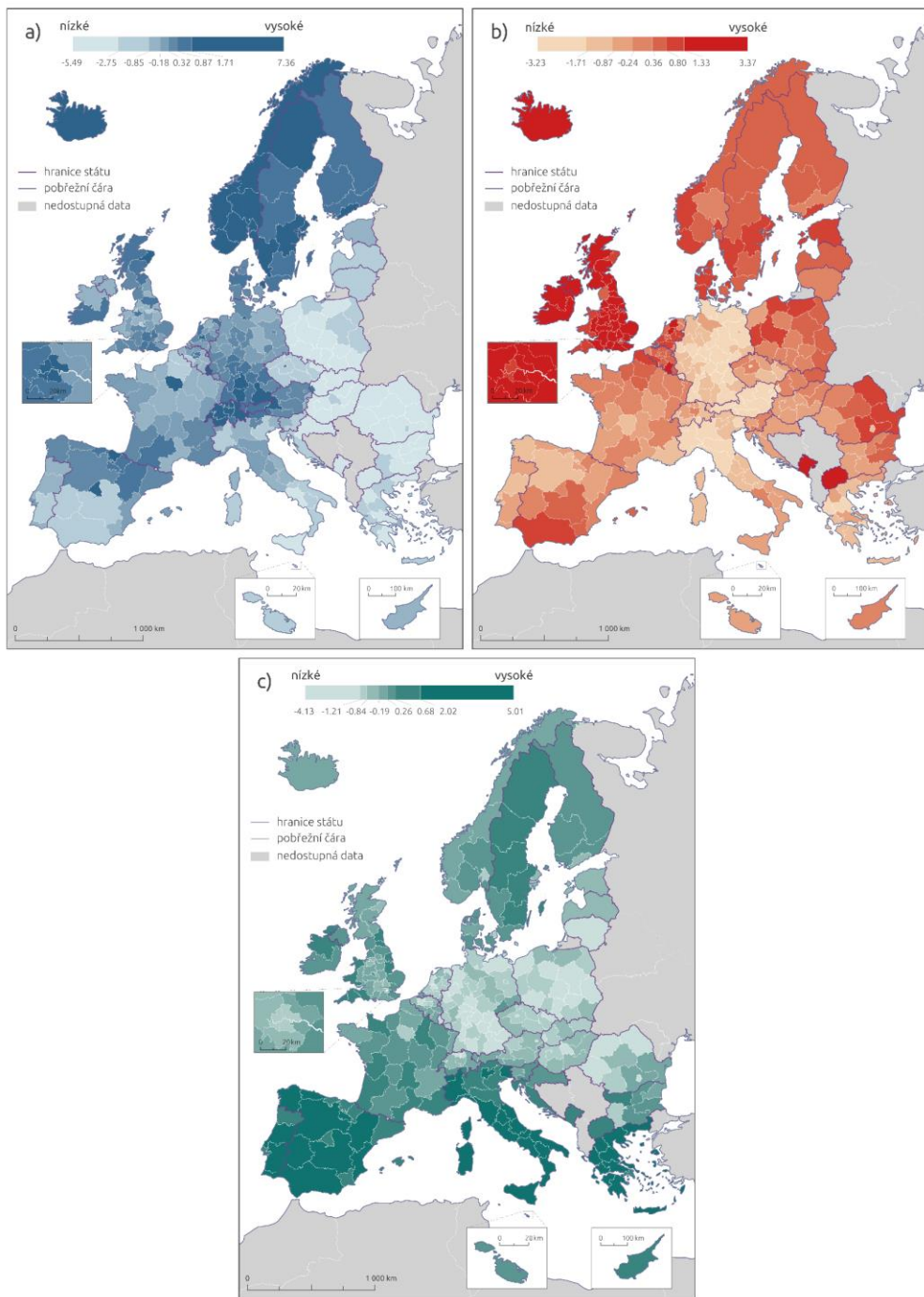
Výše vymezené tři komponenty mají silný potenciál být chápány jako nové aspekty kvality života evropských regionů opírající se o nejvýznamnější vliv vybraných indikátorů. Syntetický index kvality života QoL byl vypočten podle vzorce (2) z výše odvozených komponent reprezentujících nově odhalené aspekty kvality života. Znaménko každé komponenty bylo nastaveno podle hodnocení jejího přínosu k tématu kvality života. Váhy byly W_i stanoveny na základě míry vysvětleného rozptylu jednotlivých komponent, poměrově přepočteny tak, aby v součtu tvořily 100 % (míry vysvětleného rozptylu původních dat komponentami jsou uvedeny v popisu komponent v kapitole Syntéza dat).

$$QoL = \sum_1^n W_i F_i = 0.4 \times PC1 + 0.35 \times PC2 - 0.25 \times PC3 \quad (2)$$

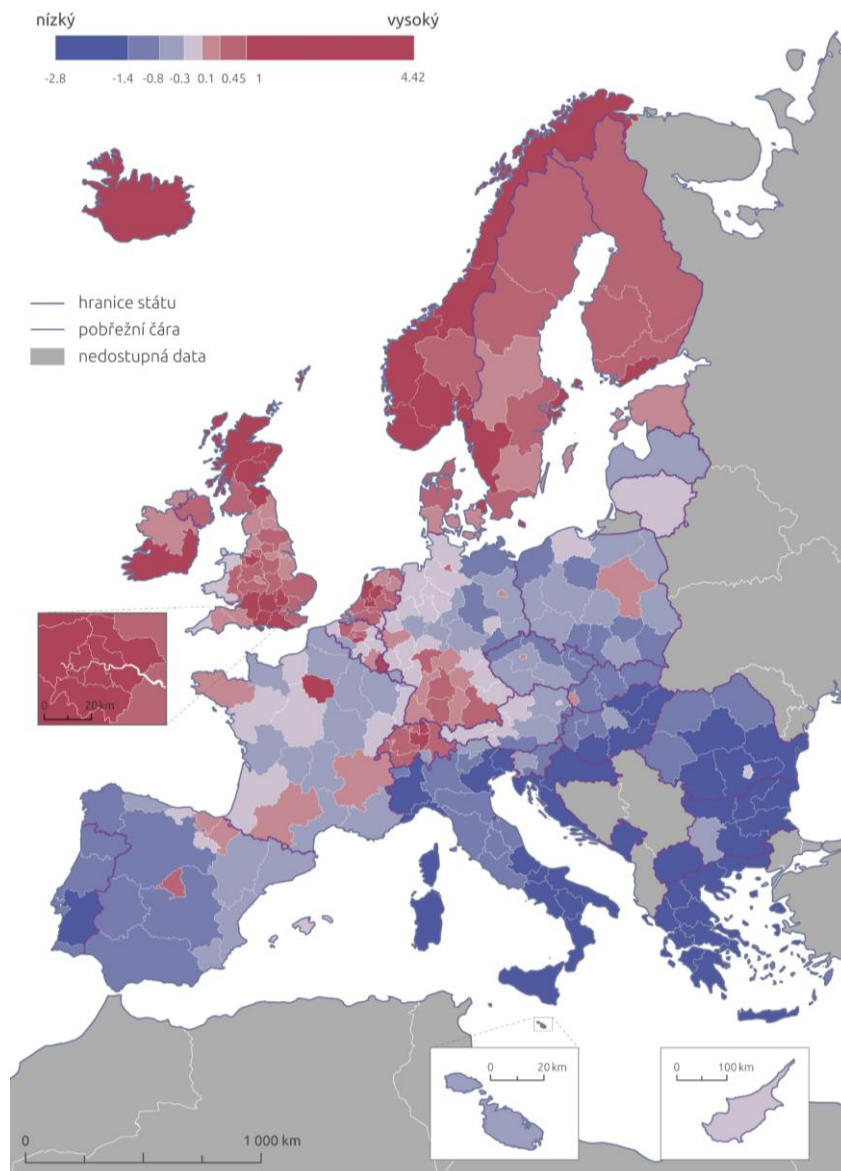
$QoL, W_i F_i$ – viz vzorec 1

$PC1 - PC3$ – viz kapitola Syntéza dat

Prostorovou vizualizací indexu QoL (Obr. 3) byly identifikovány oblasti s vysokou, respektive nízkou životní úrovní. Vizualizace identifikuje několik významných oblastí vysokých hodnot – Skandinávie, Britské ostrovy, oblast jižního Německa až Švýcarska a země Beneluxu. Nejnižší hodnoty výrazně dominují v jihovýchodní Evropě, pokračující v pásu přecházejícím pobřeží Středozemního moře až na Pyrenejský poloostrov. Absolutně nejvyšší hodnota indexu QoL je v regionu *Inner London – West* (hodnota indexu QoL 4,42), protože zde mají extrémně silný vliv ekonomické indikátory přispívající do první komponenty. Zajímavým postřehem je srovnání hodnot indexu QoL regionů hlavních měst vůči ostatním regionům. V řadě států si tyto administrativní jednotky vedou výrazně lépe než okolí. Národním příkladem jsou Bratislava, Bukurešť a Praha, v západních zemích Madrid a Paříž. Méně výrazný rozdíl mezi hlavním městem a jeho okolím je patrný v případě Vídně a Berlína, v Itálii region hlavního města Lazio nevystupuje nad svým okolím vůbec. Při pohledu na vstupní data se Lazio překvapivě v žádném indikátoru výrazně neliší od svého okolí. Mnohem výraznější rozdíly ve vstupních datech jsou patrné mezi severem a jihem Itálie (například v příjmu domácností nebo zaměstnanosti), v rámci souhrnného QoL indexu jsou však tyto rozdíly srovnány vlivem indikátorů na jednotlivé komponenty.



Obr. 2 Komponentní skóre tří nejvýznamnějších komponent (v intervalech podle septilů) vypočtených robustní PCA



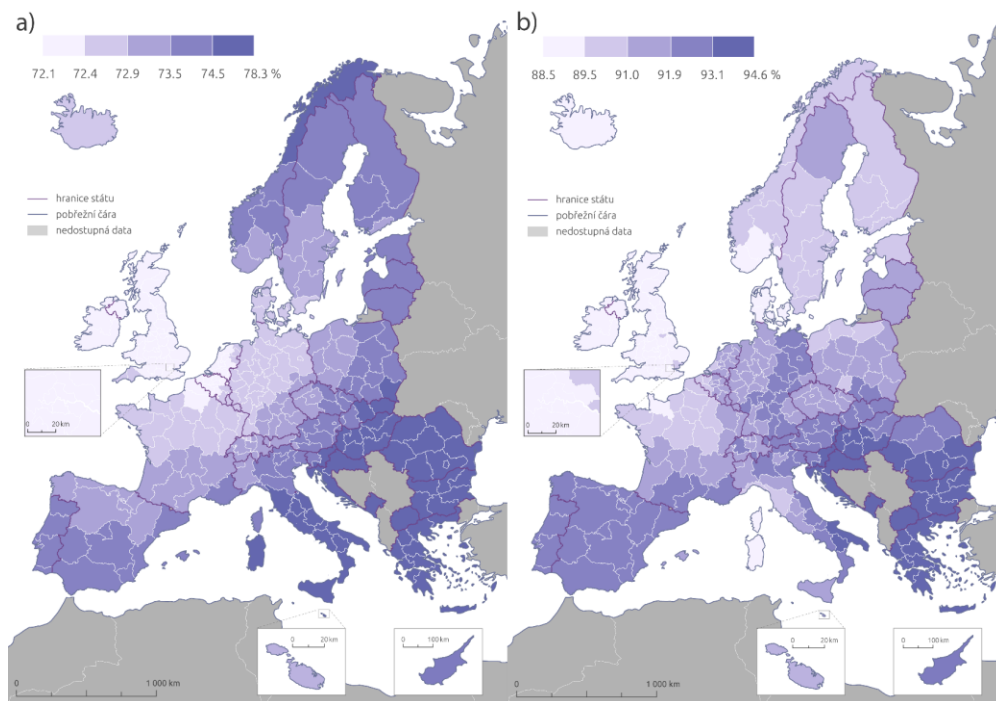
Obr. 3. Index kvality života regionů NUTS 2 v Evropě k roku 2013 (v intervalech podle septilů)

Prostorová variabilita modelu PCA

Jelikož byla zpracována prostorová data, je vhodné se ptát, zdali mají prostorové vazby mezi jednotlivými regiony vliv na výsledky PCA. K získání odpovědi na tuto otázku byla vypočtena GW PCA. Pro výpočet bylo nutné vyřešit otázku nastavení dvou parametrů – výběr jádrové funkce, která ovlivňuje ohodnocení vlivu okolní při lokálním odhadu a nastavení typu a velikosti prohledávané oblasti, tzv. *bandwidth* (adaptivní vzdálenost kontrolovaná počtem sousedů). Tento krok je klíčový, neboť jeho nastavení výrazně ovlivňuje výsledky. Určení optimálních hodnot parametrů bylo vyřešeno testováním různých nastavení, jak je popsáno v kapitole Metody.

Při využití PCA je obecně snaha maximalizovat vysvětlený rozptyl, GW PCA dokáže ve vysvětleném rozptylu identifikovat prostorové rozdílnosti. Po experimentování s výsledky různého nastavení *bandwidth* (optimální dle cross-validace, uživatelsky nastavené – např. 10 %

všech záznamů) byla použita hodnota počtu sousedů 283, což zahrnuje všechny záznamy datového souboru. Takto vytvořený model se dá považovat za téměř globální (liší se pouze ve váze regionů měnící se se vzdáleností) a je tedy nejvíce srovnatelný s globálním modelem použitým pro sestavení indexu kvality života. Zároveň umožňuje porovnat jednoduchou a robustní metodu, což v případě optimálních nastavení určených cross-validací nebylo srovnatelné. Prostorová variabilita vysvětleného rozptylu je vizualizována na Obr. 4.



Obr. 4. Množství vysvětleného rozptylu a) čtyřmi komponentami při jednoduché GW PCA, b) třemi komponentami při robustní GW PCA

Jednoduchá GW PCA je typická shladeným průběhem v pásu rostoucích hodnot (cca 72 – 78 % vysvětleného rozptylu) od západu k východu. Robustní GW PCA postrádá tak jasný prostorový trend, ale zato vysvětluje výrazně větší množství rozptylu (cca 88 – 94 % rozptylu původních dat). Obě metody dosahují nejvyšších hodnot v jihovýchodní části Evropy, lze tedy tvrdit, že zde nacházející se regiony jsou v nových komponentách nejlépe vysvětleny. Na druhou stranu, nejmenší míra vysvětleného rozptylu se v obou případech nachází na Britských ostrovech. Zde vybrané komponenty obsahují největší podíl informace, kterou se jimi nepodařilo vystihnout.

Druhou zajímavou výstupní informací je posouzení vlivu jednotlivých indikátorů lokálního modelu na komponenty, což umožňuje sledovat prostorovou variabilitu významnosti jednotlivých indikátorů. Jako nejvýraznější se na první komponentě robustní PCA projeví indikátory míry zaměstnanosti, kojenecké životnosti a NEET. Vliv těchto indikátorů je překvapivý, jelikož v celkovém globálním modelu mezi nejsilnější nepatří. V případě jednoduché PCA byly dominantní HDP na obyvatele, příjem domácností a NEET. Tento výsledek je již v souladu s globálním modelem PCA.

Diskuze

V analytické části byly představeny možnosti detekce extrémních hodnot a využití PCA v předkládané problematice kvality života. Úmyslně není diskutován výsledný index kvality života a jeho skutečná vypovídající hodnota, neboť záměrem autorů a cílem práce je prezentovat

možnosti hodnocení kvality života pomocí analýzy hlavních komponent a představit rozdílnost mezi jednoduchým, robustním a prostorově váženým přístupem. PCA se tedy osvědčila jako zajímavý nástroj pro agregaci dat do syntetického indexu a vymezení skrytých aspektů kvality života a jejich vazeb na původní indikátory. Výhodou postupu využívajícího multidimenzionální statistické metody je částečná eliminace autorových subjektivních pocitů a zpracování založené pouze na numerických výpočtech. Při tvorbě indexu QoL pomocí PCA je ze strany autora nezbytné pouze posouzení vlivu jednodušších komponent do závěrečné sumarizace. Ohodnocení komponenty jako pozitivní nebo negativní má zásadní dopad na hodnotu výsledného indexu. Při tomto nastavování se nejvíce projevuje autorův subjektivní pocit a může být často dvojsečný. PCA často nebývá interpretována, nové komponenty jsou pouze využity v dalších výpočtech. Přidaná hodnota v geografických aplikacích může spočívat právě v této interpretaci, kdy se autor snaží komponenty pojmenovat pro lepší porozumění. Tento úkon je z celého hodnocení nejnáročnější a nejvíce diskutabilní. Mohou totiž vznikat kontroverzní komponenty, do kterých stejnou mírou přispívají indikátory, které jsou chápány pozitivně, tak i ty které jsou chápány negativně. Příkladem je druhá komponenta, kde je přítomen negativní i pozitivní vliv pozitivně chápaných indikátorů (viz. Tab. 3), podobná situace nastala také u třetí komponenty. Zároveň je často problematické komponentu vhodně slovně popsat, jelikož nese syntetickou informaci, která může spojovat indikátory, které nemají v teoretické rovině žádnou smysluplnou spojitost. Proto je pojmenování nových komponent nutno brát s velkou rezervou. Z těchto důvodů může být vhodnější hodnotit jednotlivé komponenty separovaně, a do syntézy indexem se pouštět až v případech, kdy je interpretace jasná a dobře uchopitelná.

O vlivu extrémních hodnot na libovolné statistické metody není pochyb. Představený postup jejich detekce umožňuje identifikovat problematické záznamy, kterým je při práci vhodné věnovat větší pozornost. Pro zajištění větší správnosti je nutné s extrémními hodnotami vědomě pracovat, v tomto případě byly ošetřeny pomocí robustní varianty PCA. Ta se ukázala jako efektivnější z hlediska poměru vysvětleného rozptylu k počtu komponent a usnadnila tak interpretaci celého modelu.

Nejkritičtější nahlíží autoři na možnosti využití GW PCA. Prostorová složka je při využití statistických metod často opomíjena, proto bylo vhodné do analýzy kvality života tento fakt zahrnout. Dostupné metody vyžadují velkou pozornost z hlediska nastavení, zejména v nastavení parametrů pracujících s prostorovou informací. Jak bylo představeno výše, vliv definování bandwidth se zásadně projevuje na poskytovaných výsledcích. Autoři se domnívají, že důraz na lokální složku může být nejužitečnější v momentu, kdy je cílem sledovat změny způsobené jen vybraným okolím (např. variabilita v rámci regionu a jeho blízkého okolí, kde může docházet k migraci obyvatel). V celkovém hodnocení se autoři přiklání k použití modelu, který se svou prostorovou vlastností blíží k modelu globálnímu: takový model je vhodný pro interpretaci a zároveň dokáže do jisté míry zohlednit proměnlivost způsobenou prostorovými vazbami. I v takovém případě je však interpretovatelnost diskutabilní (viz. neshoda významnosti indikátorů v globálním a GW modelu). Nejpoužitelnější se autorům jeví využití prostorové variability vysvětleného rozptylu, která může sloužit jako míra kvality a nejistoty vypočteného modelu. Pokud v analýze nejsou hledány vlivy způsobené konkrétním prostorovým okolím, jsou průhlednější výsledky zajištěny použitím globálního modelu PCA. GW PCA také bohužel nelze využít pro konstrukci indexu, jak bylo představeno v kapitole Syntéza dat. To vyplývá z podstaty metody, kdy by musel být index konstruován samostatně v každém regionu na základě lokálního modelu. Takový přístup je prakticky nemožné interpretovat.

Závěr

V práci bylo představeno zpracování navrženého datasetu indikátorů kvality života (QoLD) metodami vícerozměrné statistiky. Jelikož je kvalita života velice komplexní jev, nelze u žádného navrženého hodnocení tvrdit, že se jedná o to jediné správné. Při každém hodnocení se jedná pouze o aproximaci sledovaného tématu s ohledem na vymezený koncept

a použitá data, což je nutné brát při celém hodnocení v potaz. Cílem této práce nebylo konkrétní hodnocení kvality života s hlubokou diskuzí a interpretací výsledků, ale především představení možnosti využití pokročilých statistických metod a postup jejich zpracování při aplikaci na geografickou úlohu. Byl kladen důraz na využití správných metod, vymezených na základě charakteru vstupních dat. Využití vícerozměrné statistiky není inovací metodologickou, nýbrž aplikační výzvou. Použití analýzy hlavních komponent jako nástroje konstrukce indexu QoL je pro rozsáhlé a proměnlivé území (kterým Evropa je) tou novou výzvou zejména v interpretační rovině. V přehledu dostupné literatury je jen málo prací zabývajících se tématem kvality života v tak velkém územním rozsahu, jako je celá Evropa v klasifikaci NUTS 2. Aby byla zahrnuta korektnost pro zpracování prostorových dat, byly vyzkoušeny možnosti použití GW PCA. Tato metoda dle očekávání identifikovala prostorovou variabilitu sledovaného jevu, nicméně se ukázala být velice obtížná pro parametrizaci a rozumnou interpretaci. V konfrontaci s globálním modelem nabízí výrazně odlišné výsledky, které jsou pro potřeby hodnocení kvality života matoucí. Nicméně využití GW PCA bylo zajímavým výchozím bodem pro další budoucí hodnocení tématu kvality života s ohledem na prostorový kontext.

Literatura

- ANDREWS, F. M. ed. 1986: *Research on the Quality of Life*. Michigan (Survey Research Center, Institute of social research, University of Michigan).
- BÉRENGER, V., VERDIER-CHOUCHE, A. 2007: Multidimensional Measures of Well-Being: Standard of Living and Quality of Life Across Countries. *World Development*, 35(7), 1259-1276. DOI: <https://doi.org/10.1016/j.worlddev.2006.10.011>.
- CAMPBELL, A., CONVERSE, P. E., RODGERS, W. L. 1976: *The Quality of American Life: Perceptions, Evaluations and Satisfaction*. New York (Russell Sage Foundation).
- DIXON, W. J. 1950: Analysis of Extreme Values. *The Annals of Mathematical Statistics*, 21(4), 488-506. DOI: <https://doi.org/10.1214/aoms/1177729747>.
- EUROPEAN COMMISSION 2009: *GDP and beyond: Measuring progress in a changing world*. Brussels (European Commission).
- EUROSTAT 2015: Quality of Life – facts and views. Luxembourg (Publications Office of the European Union). DOI: <https://doi.org/10.2785/59737>.
- EUROSTAT, INSEE 2011: *Report of the Task Force: Multidimensional measurement of the quality of life*. Paříž (Eurostat, INSEE – Institute national de la statistique et des études économiques). 27 p.
- GREYLING, T., TREGENNA, F. 2016: Construction and Analysis of a Composite Quality of Life Index for a Region of South Africa. *Social Indicators Research*, 131(3), 887-930. DOI: <https://doi.org/10.1007/s11205-016-1294-5>.
- HARDEMAN, S., DIJKSTRA, L. 2014: *The EU Regional Human Development Index – JRC Science and Policy Reports*. Luxembourg (Publications Office of the European Union). DOI: <https://doi.org/10.2760/26355>.
- HARRIS, P. et al. 2015: Enhancements to a geographically weighted principal component analysis in the context of an application to an environmental data set. *Geographical Analysis*, 47(2), 146-172. DOI: <https://doi.org/10.1111/gean.12048>.
- HARRIS, P., BRUNSDON, C., CHARLTON, M. 2011: Geographically weighted principal components analysis. *International Journal of Geographical Information Science*, 25(10), 1717-1736. DOI: <https://doi.org/10.1080/13658816.2011.554838>.
- HUBERT, M., DEBRUYNE, M. 2010: Minimum covariance determinant. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1), 36-43. DOI: <https://doi.org/10.1002/wics.61>.
- HUBERT, M., ROUSSEEUW, P. J., VANDEN BRANDEN, K. 2005: ROBPCA: A New Approach to Robust Principal Component Analysis. *Technometrics*, 47(1), 64-79. DOI: <https://doi.org/10.1198/004017004000000563>.
- JOLLIFFE, I. T. 2002: *Principal Component Analysis*. New York: Springer-Verlag (Springer Series in Statistics). DOI: <https://doi.org/10.1007/b98835>.

- KAISER, H. F. 1960: The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20(1), 141-151. DOI: <https://doi.org/10.1177/001316446002000116>.
- LI, G., WENG, Q. 2007: Measuring the quality of life in city of Indianapolis by integration of remote sensing and census data. *International Journal of Remote Sensing*, 28(2), 249-267. DOI: <https://doi.org/10.1080/01431160600735624>.
- LO, C. P., FABER, B. J. 1997: Integration of landsat thematic mapper and census data for quality of life assessment. *Remote Sensing of Environment*, 62(2), 143-157. DOI: [https://doi.org/10.1016/S0034-4257\(97\)00088-6](https://doi.org/10.1016/S0034-4257(97)00088-6).
- MAHALANOBIS, P. C. 1936: On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1), 49-55.
- MARTÍN, J. C., MENDOZA, C. 2013: A DEA Approach to Measure the Quality-of-Life in the Municipalities of the Canary Islands. *Social Indicators Research*, 113(1), 335-353. DOI: <https://doi.org/10.1007/s11205-012-0096-7>.
- MEDERLY, P., TOPERCER, J., NOVÁČEK, P. 2004: *Indikátory kvality života a udržitelného rozvoje: kvantitativní, vícerozměrný a variantní přístup*. Praha (Univerzita Karlova, Fakulta sociálních věd CESES).
- MILITKÝ, J., MELOUN, M. 2003: Metoda hlavních komponent a exploratorní analýza vícerozměrných dat. *Sborník přednášek z konference: Zajištění kvality analytických výsledků – Komorní Lhotka*, 38-62.
- MORAIS, P., CAMANHO, A. S. 2011: Evaluation of performance of European cities with the aim to promote quality of life improvements. *Omega*, 39(4), 398-409. DOI: <https://doi.org/10.1016/j.omega.2010.09.003>.
- RAO, K. R. M. et al. 2012: Assessment of Quality of Life in Uttarakhand, India using geospatial techniques. *Geocarto International*, 27(4), 315-328. DOI: <https://doi.org/10.1080/10106049.2011.627470>.
- SMITH, D. M. 1972: Geography and social indicators. *South African Geographical Journal*, 54(1), 43-57. DOI: <https://doi.org/10.1080/03736245.1972.10559497>.
- SMITH, D. M. 1973: The Geography of Social Well-Being in the United States: An Introduction to Territorial Social Indicators. *Social Indicators Research*, 1(2), 257-259. DOI: <https://doi.org/10.1007/BF00302892>.
- STIGLITZ, J. E., SEN, A., FITOUSSI, J. P. 2009: *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris. <https://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report>

Poděkování: Tento článek byl podpořen projektem č. 2016-1-CZ01-KA203-024040 (Spatialonomy) financovaným Evropskou Unií v rámci programu Erasmus+; projektem IGA_PrF_2019_014 Interní Grantové Agentury University Palackého v Olomouci a projektem Grantové Agentury České Republiky č. 18-05432S - Prostorová syntéza založená na pokročilých metodách geocomputation.

Adresy autorů

Mgr. Karel Macků
Katedra geoinformatiky
Univerzita Palackého v Olomouci
17. listopadu 50, 771 46 Olomouc
Česká republika
karel.macku@upol.cz

prof. RNDr. Vít Voženílek, CSc.
Katedra geoinformatiky
Univerzita Palackého v Olomouci
17. listopadu 50, 771 46 Olomouc
Česká republika
vit.vozenilek@upol.cz