

Disy Informationssysteme GmbH, Karlsruhe; Hochschule Karlsruhe

# FRAMEWORK ZUR ERSTELLUNG KOMPLEXER, RÄUMLICHER VERGLEICHSINDIZES AUF BASIS OFFENER DATEN – ANWENDUNG AM BEISPIEL DER FAHRRADFREUNDLICHKEIT IN STÄDTEN BADEN-WÜRTTEMBERGS

Nina Kirschner, Julian Bruns, Andreas Abecker, Johannes Kutterer, Detlef Günther-Diringer

**Zusammenfassung:** Um Objekte, Orte oder Regionen bezüglich einer komplexen, qualitativen, abstrakten Eigenschaft möglichst objektiv zu bewerten und damit auch vergleichen zu können, kann man sie solange rekursiv auf einfachere, konkretere Eigenschaften zurückführen, bis man bei Merkmalen (Indikatoren) angekommen ist, die sich quantitativ, und idealerweise auch objektiv, beschreiben lassen. Kann man die entsprechenden Merkmalswerte auf Basis von (offenen) Datenquellen bestimmen, ist die Berechnung eines solchen Vergleichswertes u.U. sogar automatisierbar. Eine solche aufwendige Aggregation von Merkmalswerten wird dann als komplexer Vergleichsindex (KVI) bezeichnet.

In diesem Beitrag wird ein methodisches Framework entwickelt, mit dem ein KVI für räumliche Objekte mithilfe offener Daten berechnet werden kann. Dieses erfüllt die folgenden Anforderungen: Die Werte des Index müssen quantitativ sein, seine Berechnung muss transparent sein, der Index muss anpassbar und verfügbar sein, er muss sich allein aus offenen Daten berechnen lassen und er muss für jeden nutzbar sein.

Der Ansatz wird auf realen Daten und Anwendungsszenarien getestet. Dazu wird exemplarisch ein Vergleichsindex für Fahrradfreundlichkeit von Städten in Baden-Württemberg berechnet. Die Ergebnisse sind anschaulich und transparent nachvollziehbar. Im Beispiel sind sie inhaltlich jedoch eher ernüchternd, was auf die heute noch unzureichende Verfügbarkeit und Qualität offener Daten zurückzuführen ist.

Dagegen ist der Ansatz methodisch-technisch überzeugend. Die Analysepipeline bietet die Möglichkeit, den Index durch eine Änderung der Gewichte einzelner Indikatoren anzupassen. Damit kann jeder Nutzer individualisierte Indizes für spezifische Zwecke und Anforderungen erstellen, berechnen und visualisieren lassen. Vielversprechende zukünftige Arbeiten – vorausgesetzt, die Datenlage verbessert sich – liegen z. B. im Bereich der Benutzerfreundlichkeit und im Austausch und kollaborativen Weiterentwickeln von KVI.

**Schlüsselwörter:** Open Data, Vergleichsindex, komplexer Vergleichsindex, zusammengesetzter Index, Geodatenanalyse

## FRAMEWORK FOR THE CREATION OF COMPLEX, SPATIAL COMPOSITE INDICES BASED ON OPEN DATA – APPLICATION USING THE EXAMPLE OF BICYCLE-FRIENDLINESS IN CITIES OF BADEN-WÜRTTEMBERG

**Abstract:** In order to evaluate objects, places or regions with respect to a complex, qualitative, abstract property as objectively as possible and thus to be able to compare them, one can recursively trace them back to simpler, more concrete properties until one has arrived at features (indicators) that can be described quantitatively, and ideally also objectively. If the corresponding values can be determined based on (open) data sources, the calculation of such a value may even be automated. Such an elaborate aggregation of characteristic values is then called a composite index.

In this paper a methodical framework is presented which allows the calculation of a composite indicator for spatial objects based on open data. This framework fulfills the following requirements: the values of the indicator must be quantitative, its calculation must be transparent, the indicator must be adaptable and available, it must be calculated based on open data alone, and it must be usable for everyone.

The approach is tested on real data and application scenarios. For this purpose, a composite index for bicycle-friendliness of cities in Baden-Württemberg is calculated as an example. The results are clear and transparent. In the example, however, they are not too convincing from the content point of view which is due to the currently still insufficient availability and quality of open data. In contrast, from the methodological and technical point of view, the approach is indeed convincing. The analysis pipeline offers the possibility of adjusting the index by changing the weights of individual indicators. This allows each user to create, calculate and visualize individu-

alized indices for specific purposes and requirements. Promising future work – provided that the data situation improves – lies, for example, in the area of user-friendliness and in the exchange and collaborative further development of composite indices.

**Keywords:** Composite indicators, geodata analysis, open data

### Autoren

B. Sc. Nina Kirschner

Dr. Julian Bruns

Dr. Andreas Abecker

Dr. Johannes Kutterer

Disy Informationssysteme GmbH

Ludwig-Erhard-Allee 6

D-76131 Karlsruhe

E: [nina.kirschner@disy.net](mailto:nina.kirschner@disy.net)

[Julian.bruns@disy.net](mailto:Julian.bruns@disy.net)

[Andreas.abecker@disy.net](mailto:Andreas.abecker@disy.net)

[Johannes.kutterer@disy.net](mailto:Johannes.kutterer@disy.net)

Prof. Dr. Detlef Günther-Diringer

Hochschule Karlsruhe

Hoffstraße 3

D-76133 Karlsruhe

[detlef.guenther-diringer@hs-karlsruhe.de](mailto:detlef.guenther-diringer@hs-karlsruhe.de)

## 1 EINLEITUNG

Um Objekte, Orte oder Regionen bezüglich einer komplexen, qualitativen Eigenschaft möglichst objektiv zu bewerten und damit auch vergleichen zu können, versucht man häufig, abstrakte Eigenschaften solange rekursiv auf einfachere, konkretere Eigenschaften zurückzuführen, bis man bei Merkmalen angekommen ist, die sich quantitativ, und idealerweise auch objektiv, beschreiben lassen. Wenn sich die entsprechenden Merkmalswerte auf Basis von (offenen) Datenquellen bestimmen lassen, kann die Berechnung eines solchen Vergleichswerts u.U. sogar automatisiert werden. Bekannte Beispiele für zahlenbasierte Metriken oder Indizes reichen von sehr einfachen Varianten, wie dem Big-Mac-Index (Economist 2019) für die Inflation, bis zu hochkomplexen Varianten, wie dem Human Development Index (HDI) (UNDP 2019), der eine Vielzahl von Datenquellen und Informationen verknüpft. Angelehnt an Barclay et. al (2018) werden solche aufwendigen Aggregationen als *komplexer Vergleichsindex (KVI)* bezeichnet. Ein mög-

licher Ablauf zur Erstellung eines KVI kann, basierend auf OECD (2008), der Abbildung 1 entnommen werden.

Heutzutage wird in Infografiken, wissenschaftlichen oder journalistischen Auswertungen etc. i. d. R. jedoch nur das quantitative Endergebnis einer solchen KVI-Berechnung dargestellt. Datengrundlage, quantitative Werteableitung, Gewichtung und Aggregation von Einzelfaktoren sind selten komplett veröffentlicht. Derartige Intransparenz kann zu grundlegendem Misstrauen führen und dadurch den Nutzen eines KVI reduzieren. Außerdem muss man zur Berechnung eines KVI häufig generalisierende Festlegungen treffen, insbesondere die Gewichtungen der einzelnen Faktoren, was sowohl den Nutzen für Einzelpersonen als auch den Nutzen hinsichtlich spezieller Bedürfnisse und Interessen reduziert.

Ausgehend von den sechs Hauptproblemen bei der Indexerstellung in Barclay et. al (2018) werden die folgenden Anforderungen aufgestellt, die für die Demokratisierung der KVI erreicht werden sollten:

1. Die Index-Werte sollen quantitativ und direkt vergleichbar sein.
2. Es soll transparent sein, wie sich der Index-Wert berechnet.
3. Der Index soll auf offenen Daten basieren.
4. Der Index soll nicht nur für Experten nutzbar und erstellbar sein.
5. Der Index soll möglichst einfach anpassbar sein, z.B. für verschiedene Rahmenbedingungen oder Lebenssituationen der Nutzer.
6. Der Index soll für jede beliebige geographische Fläche berechnet werden können, wie Stadtteile, Städte, Bundesländer etc.

Insbesondere die Anforderungen 2-5 stehen dabei im Fokus dieses Beitrags. Diese Eigenschaften können das Vertrauen in einen solchen KVI und damit auch dessen Wirksamkeit stärken. Durch die leichte Erstellbarkeit und Anpassbarkeit lassen sich verschiedenste Indikatorkombinationen testen und damit auch individuelle Bedarfe und Anforderungen abbilden. Softwaretechnische Zielsetzung dieses Beitrags und

des zugrunde liegenden Projekts ist die Entwicklung eines Frameworks, das die obigen Anforderungen auf Basis von geographischen Analysen mit einer Analytics-Plattform komplett auf offenen Daten erfüllen kann und damit die Erstellung von räumlichen KVI erheblich erleichtert. Dieses Framework wird auf realen Daten und Anwendungsszenarien getestet.

Im Rahmen dieses Beitrags wird diese Idee am Beispiel eines Index für die Fahrradfreundlichkeit einer Stadt untersucht. Hierfür werden vier Varianten eines KVI erstellt: (1) ein möglichst allgemeiner Fahrradindex sowie drei spezifische Varianten für typische Nutzergruppen des Fahrrads, nämlich: (2) Familie mit Lastenrad, (3) Sportfahrer mit Rennrad und (4) Freizeitfahrer mit Elektrofahrrad. Im Rahmen der Praxisevaluation zeigt sich dabei einerseits die starke Abhängigkeit des Verfahrens von den zur Verfügung stehenden offenen (räumlichen) Daten, andererseits aber auch die Stärken des Ansatzes in punkto Transparenz und Nachvollziehbarkeit.

## 2 STAND DER FORSCHUNG

Die Verwendung von Indizes zur schnellen Vermittlung von Informationen ist ein Standardverfahren in der Geoinformatik. In diesem Abschnitt werden ausgewählte Beispiele für die Methodik und Anwendung von räumlichen KVI vorgestellt – mit Fokus auf dem betrachteten Anwendungsbeispiel der Fahrradfreundlichkeit.

Grundlegende Vorgehensweisen und Methodiken für die Erstellung von KVI werden z.B. in Barclay et. al (2018), OECD (2008), Saltelli (2007), Munda & Nardo (2005), Li et al. (2015) und Mazziotta & Pareto (2013) betrachtet. Diese werden dabei im englischen Sprachgebrauch als *Composite Index* beschrieben und diskutiert. In den Arbeiten werden die allgemeinen Anforderungen und Anwendungen diskutiert und Hinweise für die Erstellung gegeben.

Bassolas et al. (2019) entwickeln eine Metrik und darauf basierende Klassifikation von Städten zur Bestimmung der Mobilität und deren Auswirkungen auf die Stadt. Diese werden mit einer Vielzahl an bestehenden Indizes von urbaner Qualität verglichen und gezeigt, dass ein hierarchischer Ansatz basierend auf geostatistischen Hotspotanalysen bei geringeren Kosten und verbesserter Skalierbarkeit zu ähnlichen Er-

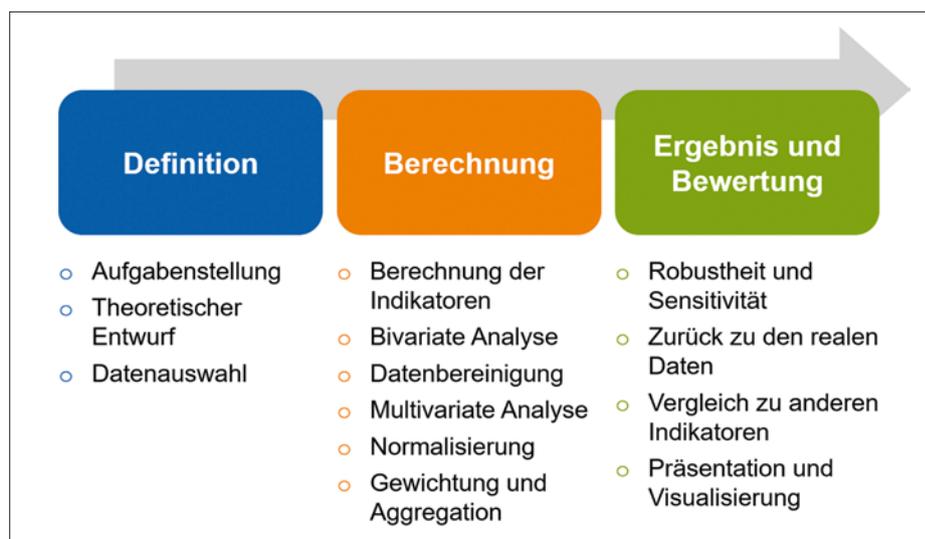


Abbildung 1: Schritte zur Erstellung eines komplexen Vergleichsindex

gebnissen kommt. Hierbei wird insbesondere auf die Vielzahl an oft heterogenen Prädiktoren in der Literatur eingegangen. Hierfür werden Beispiele aus Ewing et al. (2015) verwendet, welche in ihrer Arbeit einen Überblick der Indikatoren für die Ausdehnung der Stadtgebiete präsentieren. Weitere Details zur Implementierung finden sich hierzu in Google (2019).

In der Literatur wird auch eine Vielzahl an (räumlichen) Indikatoren für die Fahrradfreundlichkeit vorgeschlagen, die unterschiedliche Herangehensweisen und Prädiktoren verwenden. Alleine der Begriff „bikeability“ bei Google Scholar ergibt mehr als 2.400 Treffer in 2019. Daher kann hier nur eine kurze Auswahl an relevanten Arbeiten vorgestellt werden.

Im Bereich der Fahrradfreundlichkeit sind zwei der bekanntesten Indizes der ADFC Fahrradklima-Test (ADFC 2018b) und der Copenhagenize Index (Copenhagenize 2018). Der Copenhagenize Index (Copenhagenize 2018) vergleicht im Jahr 2019 insgesamt 115 Städte weltweit in 14 verschiedenen Parametern. Diese Parameter stammen aus drei Kategorien: (1) *Straßenlandschaft* – z.B. Fahrradstellplätze, (2) *Kultur* – z.B. Frauenanteil und (3) *Ambition* – z.B. Stadtplanung. Dabei werden nur die besten 20 Plätze im Städte-Ranking veröffentlicht. Die genaue Methodik wird dabei jedoch nicht veröffentlicht, sondern, laut den Autoren, nur von diesen angewandt, um einheitliche Ergebnisse zu garantieren. Der ADFC Fahrradklima-Test (ADFC 2018b) wurde über qualitative Umfragen durchgeführt und hat im Jahr 2018

insgesamt 683 Städte in Deutschland bewertet. Dabei wurde 2018 auch eine Unterkategorie zum Thema Familien und Kinder eingeführt, um deren spezielle Bedürfnisse abzudecken.

Loidl & Zagel (2014) stellen in ihrer Arbeit ein Bewertungsmodell vor, welches auf einzelnen Indikatoren aufbaut. Hierbei werden in einem iterativen Arbeitsablauf zuerst relevante Indikatoren identifiziert und anschließend modelliert, evaluiert und kalibriert. Dieser Ansatz basiert auf einer Linearkombination der einzelnen Indikatoren und wird für die Region Salzburg und den Use Case der Sicherheit von Fahrradfahrern evaluiert. Experten können die Modelle leicht modifizieren und wiederverwenden. Die Ergebnisse werden dabei direkt in einem GIS weiter genutzt. Der Unterschied zum vorliegenden Beitrag besteht darin, dass jeder einzelne Straßenabschnitt bewertet wird, wohingegen hier ein Indikator für eine beliebige Fläche, wie eine Stadt, erstellt wird.

Ein weiteres Beispiel für einen lokalen und konkreten Anwendungsfall von Indikatoren ist in Forster (2019) zu finden. Dabei wurde für einzelne Straßenabschnitte und Kreuzungen das Unfallrisiko für Fahrräder in Form eines Bewertungsindex berechnet und auf einer Karte visualisiert. Dies wurde exemplarisch für die Stadt Augsburg durchgeführt und evaluiert.

Weitere Vorgehensweisen für die Indikatorerstellung finden sich zum Beispiel bei Gehring (2016) und Krenn (2015), die 100 m × 100 m Rasterkarten erstellen, in denen jede Zelle einen eigenen Fahrrad-

freundlichkeitswert erhält. Motta (2017) und Winters (2013) nutzen Umfragen für die Erstellung der Indizes; in Motta (2017) werden diese direkt in die Karten eingebaut und bei Winters (2013) werden die Prädiktoren für die Berechnung des Index auf Basis der Umfragen erstellt.

Weitere typische Anwendungen von räumlichen Vergleichsindizes lassen sich in der Immobilienbranche zum Vergleich von Standorten (z. B. Adam 2018), beim Zukunftsranking der Wirtschaftswoche (Wirtschaftswoche 2019), bei Stauindizes von Firmen wie INRIX, TomTom und ADAC oder bei Fußgängerindizes, wie z. B. von Dörrzapf (2019), finden, um nur einige weitere mögliche Anwendungen zu nennen.

Obwohl die Nutzung von (komplexen) Vergleichsindizes in der Geoinformatik weit verbreitet ist, ist ihre Erstellung heutzutage noch mit hohem (manuellem) Arbeitsaufwand verbunden. Es ist sowohl ein hohes Domänen- als auch IT-Wissen erforderlich, um diese Indizes zu definieren und zu modifizieren. Dies erschwert den Praxiseinsatz und die freie Gestaltung solcher Indizes für den individuellen Gebrauch. Weiterhin muss die Datengrundlage oft aufwendig erhoben werden und ist nicht frei verfügbar. Eine Replizierbarkeit und direkter Vergleich der Ergebnisse und Methoden ist daher meist kaum oder gar nicht gegeben. Durch das Erfüllen der in Abschnitt 1 beschriebenen sechs Anforderungen sollen diese Probleme durch diese Arbeit deutlich reduziert werden.

### 3 METHODIK

In diesem Abschnitt wird der Lösungsansatz beschrieben: die Methodenkombination zur Berechnung der KVI. Der Inhalt ist dabei im zweiten Schritt in Abbildung 1 enthalten. Dies stellt den statistisch-technischen Kern und Hauptbeitrag der Arbeit dar: Ein allgemeines Framework, welches ermöglicht, mit offenen Daten einen KVI mit räumlichem Bezug zu entwickeln, der einfach zu erstellen und gleichzeitig transparent ist. Die Methodik lässt sich dabei in drei Phasen unterteilen: (1) Vorverarbeitung, (2) Berechnung des KVI und (3) Vermittlung und Nachvollziehbarkeit der Ergebnisse.

Die Vorverarbeitung wird durch eine Datenbank (hier PostgreSQL mit PostGIS-Erweiterung) durchgeführt. Sämtliche raumbezogenen Operationen werden dabei in

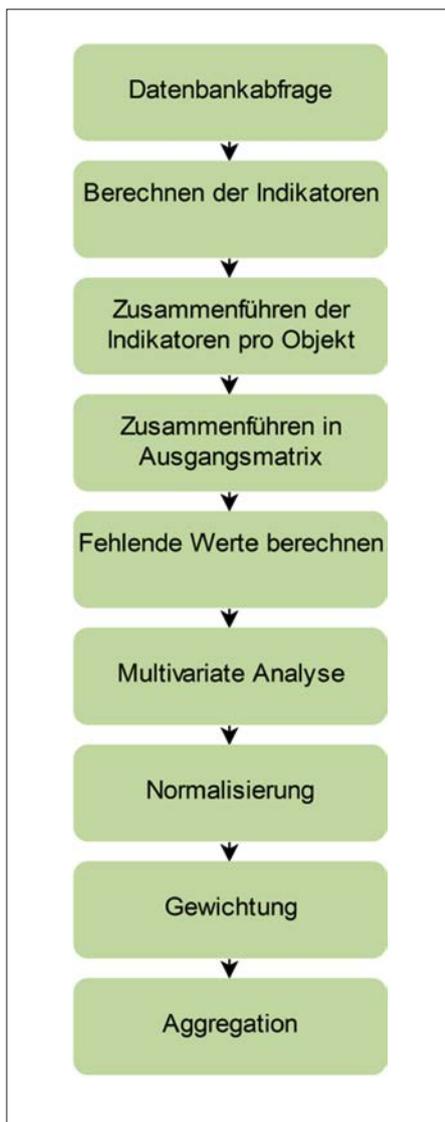


Abbildung 2: Verarbeitungsschritte zur Berechnung eines KVI

der Datenbank ausgeführt, sodass nach der Vorverarbeitung kein Raumbezug im Sinne der Regeln von Anselin (1989) vorliegt. Dies erlaubt die Verwendung beliebiger Analyseplattformen zur Berechnung von KVI, die dennoch einen räumlichen Bezug haben. Die folgenden Vorverarbeitungen und damit Abstraktionen werden in der Datenbank mithilfe von SQL-Befehlen durchgeführt: Zuschneiden auf ein beliebiges Untersuchungsgebiet, Berechnung von Längen und Flächen, die Höhe an einer Strecke in regelmäßigen Abständen, Rasterwerte in Umkreis von Features.

Die Berechnung der KVI wird im Rahmen der hier beschriebenen Arbeit mit der Analytics-Plattform KNIME (KNIME 2019) durchgeführt. Die Schritte werden in Abbildung 2 aufgezeigt. Jeder dieser Schritte wird im Folgenden einzeln beschrieben

und in die Gesamtberechnung des KVI eingeordnet.

Im ersten Schritt werden vom Anwender die benötigten vorverarbeiteten Daten aus der Datenbank gemäß den gewählten Indikatoren geladen. Jeder dieser einfachen Indikatoren wird berechnet und dient als Grundlage für den übergreifenden KVI. In diesem Beitrag werden fünf unterschiedliche Verfahren angewandt, um aus den Rohdaten in der Datenbank Indikatoren zu berechnen: (1) Anzahl an Objekten je Fläche, z. B. Einwohner pro Fläche; (2) Anzahl an Objekten innerhalb eines Buffers, z. B. Anzahl von Fahrradläden im Umkreis von Fahrradstraßen; (3) Verhältnis zwischen zwei Attributen, z. B. Anzahl der Streckenkilometer Fahrradstraßen im Verhältnis zu allen Straßenkilometern; (4) durchschnittliche Steigung, z. B. durchschnittliche Steigung einer Fahrradstraße und (5) der Mittelwert innerhalb einer Fläche, z. B. durchschnittlicher Niederschlag innerhalb einer Stadt über fünf Jahre. In Abbildung 3 wird der Ablauf beispielhaft für die Berechnung von Operation 2 in Kombination mit Operation 3 aufgezeigt. Hierbei werden parallel zuerst alle relevanten Werte aus der Datenbank gesammelt und anschließend zusammengeführt und final der einfache Indikator berechnet. Hierbei kann jeder der grundlegenden fünf Operatoren kombiniert werden.

Im nächsten Schritt der KVI-Berechnung werden die einzelnen Indikatoren zuerst für jede einzelne Untersuchungseinheit zusammengeführt (z. B. eine Stadt) und anschließend in eine übergreifende Ausgangsmatrix zur vereinfachten Berechnung und einheitlichen Durchführung der folgenden Schritte für den übergreifenden geographischen Untersuchungsgegenstand (z. B. ein Land) eingetragen.

Durch die implizite Anforderung, dass jede Analyse für jede geographische Einheit gleichartig ist, besteht die Anforderung an die Daten, dass diese nicht lückenhaft für einzelne Untersuchungsgegenstände sein dürfen. In diesem Schritt werden, wenn möglich, über Imputation Datensätze ergänzt. Falls dies nicht möglich ist, wird der betroffene Datensatz aus der Berechnung herausgenommen. Diese Entscheidung muss vom Anwender je nach Datenumlage getroffen und im Sinne der Nachvollziehbarkeit dokumentiert werden. Für die Imputation werden wohlbekanntere Verfah-

ren wie Kriging oder lineare Regressionen verwendet.

Die multivariate Analyse hilft bei der Entscheidung, welche Gewichtungsmethode und welche Aggregationsmethode gewählt wird. Diese dient als Entscheidungsunterstützung für den Anwender, indem die Zusammenhänge zwischen den Daten aufgezeigt werden. Die Korrelationsanalyse und Hauptkomponentenanalyse haben sich in der Praxis als leicht verständlich gezeigt und werden hierbei verwendet. Da sich Daten stark in den Skalen verändern können, werden für die Korrelationsanalyse diese in ihren Z-Wert transformiert. Anschließend werden die Indikatoren in einer Korrelationsmatrix verglichen. Indikatoren mit absoluten Korrelationen von nahe 1 können dabei zusammengeführt werden, da diese sehr ähnliche Informationen beinhalten. Bei einer Hauptkomponentenanalyse wird der bestehende Variablenraum über eine Linearkombination in einen neuen Raum projiziert, sodass die Varianz der einzelnen Dimensionen maximiert wird. Werden manche Variablen gar nicht oder ungenügend auf dieser Hauptachse abgebildet, wird eine zweite Hauptachse bestimmt, die senkrecht zur ersten ist. Es werden so lange Hauptachsen bestimmt, bis alle Variablen zufriedenstellend abgebildet werden. Das Ergebnis der Hauptkomponentenanalyse ist eine Ladungsmatrix, die die Korrelation der Merkmale zu den Hauptachsen angibt und eine Wertemat-

rix, die die Werte der Objekte hinsichtlich der Hauptachsen angibt (siehe z. B. Bill 2016, Backhaus 2018). Dies ermöglicht es, den Einfluss einzelner Parametergruppen auf die Berechnung des KVI darzustellen und zu interpretieren. So können hierdurch thematische Schwerpunkte oder Übergruppen identifiziert und wieder als einzelne KVI berechnet werden, was den späteren Kombinationsaufwand reduzieren kann. Als Hilfestellung wird hierzu das Kaiser-Kriterium dem Anwender gegeben. Das Kaiser-Kriterium besagt, dass alle Hauptkomponenten mit einem Eigenwert  $> 1$  für eine weitere Nutzung zu betrachten sind (Backhaus 2018).

Basierend auf den Ergebnissen der multivariaten Analysen und dem Umgang mit den fehlenden Daten liegt dem Nutzer nun eine Menge an einzelnen Indikatoren vor, die in den folgenden Schritten derart kombiniert werden müssen, dass sie in einer einzelnen Kennzahl angezeigt werden können. Da die Skalen zwischen einzelnen Werten oft unterschiedlich sein können – z. B. ist die Einwohnerzahl einer Großstadt über 100.000, während die Anzahl an Fahrradläden mehrere Potenzen kleiner ist – müssen diese vor der Kombination normiert werden. Die Normalisierungsmethode wird passend zu der Datengrundlage und der Definition des komplexen Vergleichsindex gewählt. Die Auswahl sollte sorgfältig getroffen werden, um einen aussagekräftigen KVI zu erhalten (OECD 2008).

Um die Präferenzen der Anwender abzubilden und die Anpassbarkeit des KVI zu garantieren, müssen die (normierten) Indikatoren gegeneinander gewichtet werden. Die Gewichtung muss dabei für den Anwender intuitiv sein. Aus diesem Grund wurde sich hier standardmäßig für eine Verhältnisschätzung entschieden. Dabei wird jedem einzelnen Indikator eine vom Nutzer gewählte Wertigkeit (z. B. prozentual) zugeordnet und damit multipliziert. Dies erlaubt eine Anpassung der Wichtigkeit jedes Indikators und eine sehr einfache Nachverfolgung der Gewichtung im Nachhinein. Ein weiterer Vorteil dieser Methode ist, dass keine technische Manipulation der Daten vorgenommen wird. Nachteil ist, dass diese Methode bei sehr vielen Indikatoren aufgrund des Zeitaufwands ungeeignet ist (z. B. OECD 2008, Flitter et al. 2016).

Zuletzt müssen die Gewichte mit den Indikatoren zusammengeführt werden. Dies geschieht im Rahmen einer Aggregation. Durch die Wahl der Verhältnisschätzung ergibt sich das arithmetische Mittel als einfachste Aggregationsform und wird als Standardvariante gewählt. Hierbei wird ein Produkt aus den normalisierten Werten der Indikatoren der Merkmalsträger und dem Gewicht gebildet und anschließend summiert.

Die Aggregation, Gewichtung und Normalisierung werden getrennt betrachtet, da es dadurch leicht möglich ist, diese

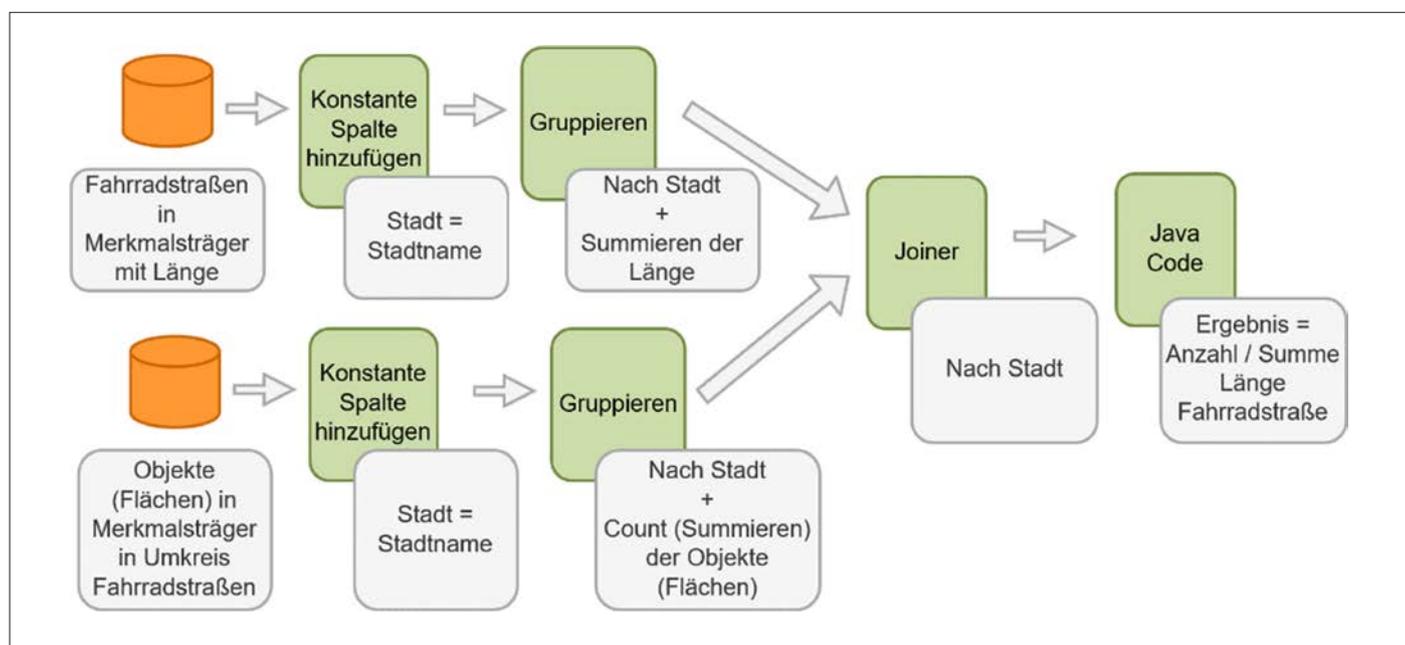


Abbildung 3: Berechnung von Anzahl Objekten innerhalb eines Buffers und Verhältnis zu zweitem Wert

jeweils einzeln dem Bedarf oder Anwendungsfall anzupassen. Dies erlaubt eine größere Flexibilität in der praktischen Anwendung. Gleichzeitig kann damit auch die Auswirkung der einzelnen Schritte leicht nachvollzogen werden – sowohl in der Betrachtung der Schritte als auch im Ergebnis.

**4 EVALUATIONSAUFBAU**

Um die Ergebnisse einordnen zu können, werden zwei intrinsische, quantitative Evaluationen und zwei qualitative Evaluationen konzipiert und realisiert. Für die quantitative Evaluation wird hierfür eine Clusteranalyse und eine Unsicherheitsanalyse durchgeführt. Für die qualitativen Evaluationen werden bestehende Indizes und verschiedene mit der Methodik erstellte KVI verglichen und der Einfluss der Schritte und Daten auf die Ergebnisse diskutiert.

Bei der Clusteranalyse werden nicht die Variablen gebündelt, sondern die Objekte. Die Objekte werden in Gruppen, Cluster oder Klassen zusammengefasst. Die Objekte, die der gleichen Klasse angehören, sollten möglichst ähnlich sein. Objekte, die nicht derselben Klasse angehören, sollten hingegen möglichst unähnlich sein. Wie möglichst ähnlich oder möglichst unähnlich definiert wird, hängt vom verwendeten Ansatz ab. Die Eingabe einer Clusteranalyse ist eine standardisierte Datenmatrix mit den Variablen der Objekte. Vor der Berechnung der Cluster muss ein Algorithmus gewählt werden. In diesem Beitrag wird mit dem k-Medoids-Algorithmus gearbeitet, bei dem zunächst eine Distanzmatrix berechnet wird. K-medoids wird hierbei statt k-Means gewählt, da dieses Verfahren robuster gegenüber Ausreißern ist und der Medoid auf einem der beste-

$$t = \frac{\bar{X}(q,CI) - \bar{X}(q)}{S(q)}$$

Mit

- $t$  = t-Wert
- $\bar{X}(q,CI)$  = Mittelwert des Indikators q über die Objekte im Cluster CI
- $\bar{X}(q)$  = Gesamtmittelwert des Indikators q
- $S(q)$  = Standardabweichung des Indikators q über alle Objekte

Gleichung 1

henden Punkte liegt. Anhand dieser Matrix werden die Gruppen gebildet (z. B. Backhaus 2018, Fahrmeir 2015). Als Entscheidungshilfe, wie viele Cluster berechnet werden, wird das Elbow-Kriterium verwendet, woraus sich in diesem Beispiel eine Anzahl von vier Clustern ergibt. Der „Elbow“ zeigt einen Sprung in der Veränderung in der Fehlerquadratsumme der Cluster (z. B. Backhaus 2018).

Zur tiefergehenden Interpretation der Cluster ist der t-Wert hilfreich. Der t-Wert beschreibt den Einfluss der einzelnen Parameter auf die Zuordnung zum Cluster. Der t-Wert berechnet sich als gewichteter Z-Wert der Variable im Cluster nach Gleichung (1).

Ein negativer t-Wert zeigt, dass der Merkmalsträger in diesem Cluster im Vergleich zu den anderen Clustern unterrepräsentiert ist. Ein positiver t-Wert zeigt hingegen, dass ein Indikator in diesem Cluster im Vergleich zu den anderen Clustern überrepräsentiert ist (Backhaus 2018).

Durch eine Unsicherheitsanalyse soll der KVI hinsichtlich seiner Verlässlichkeit des Ergebnisses bewertet werden. Dabei werden innerhalb der letzten drei Schritte

der KVI-Berechnung – der Normalisierung, Gewichtung und Aggregation – alternative Methoden verwendet und die Ergebnisse miteinander verglichen und aggregiert. Dies soll die Robustheit des Index gegenüber Veränderungen aufzeigen. Es werden aus drei verschiedenen Normalisierungsmethoden, zwei Gewichtungsmethoden und zwei Aggregationsmethoden zehn mögliche Ergebnisse berechnet. Diese werden in einem Boxplot dargestellt. Für jedes Objekt werden darin der minimale und maximale Wert, das untere und obere Quantil, die Ausreißer sowie der Median angezeigt.

Im Sinne der sechs Anforderungen, die an die Erstellung eines KVI in der Einleitung definiert wurden, wird sich hier für die Rangfolge als Normalisierungsmethode entschieden. Nachteil dieser Methode ist, dass Informationen zu den Abständen verloren gehen. Dafür fallen Ausreißer nicht ins Gewicht, die Interpretation ist eindeutig und die Ergebnisse eignen sich für die Weiterverarbeitung und Präsentation (Nardo et al. 2005). Alternativen, wie die Z-Wert-Transformation, erleichtern zwar die Automatisierung, jedoch wurde die Nach-

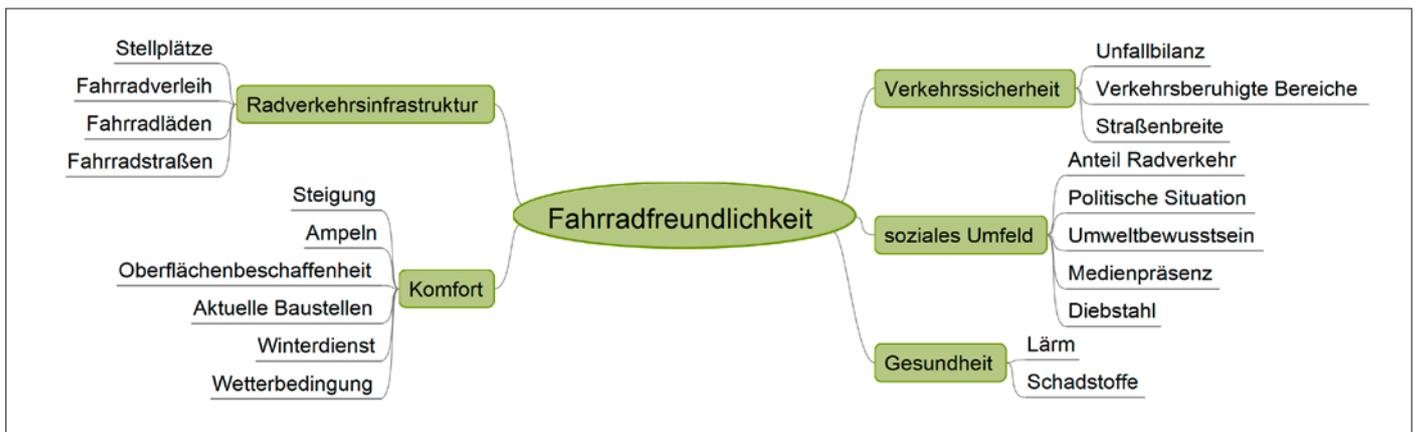


Abbildung 4: Themenbereiche und Indikatoren des Vergleichsindex zur Fahrradfreundlichkeit

vollziehbarkeit für Menschen als Priorität gesehen.

## 5 EVALUATIONSSZENARIO UND ERGEBNISSE

Um das entwickelte Framework und die eingesetzte Methodik für die Berechnung von räumlichen KVI zu evaluieren, wurde die Schaffung eines Index für Fahrradfreundlichkeit als Beispielszenario gewählt. Fahrradfreundlichkeit ist ein sowohl wissenschaftlich wie auch gesellschaftlich stark diskutiertes Thema, zu dem es eine Vielzahl an Ansichten, Bedarfen und persönlichen Präferenzen gibt. Gleichzeitig sind die bestehenden Indizes oft intransparent und nicht nachvollziehbar. Es besteht also ein Bedarf, hier frei und flexibel räumliche KVI zu erstellen, auszutauschen und zu vergleichen.

Um einen aussagekräftigen Index zu erhalten, werden die Indikatoren noch vollkommen unabhängig von den vorhandenen Daten ausgewählt (OECD 2008). Daher wird zunächst der theoretische Entwurf des Index basierend auf existierenden Indizes definiert. Die Überlegungen bauen auf den Bewertungskriterien des Copenhagenindex auf. Die Kriterien werden hinterfragt und daraufhin angepasst, verworfen oder

erweitert. Das Ergebnis ist der Abbildung 4 zu entnehmen. Es ergeben sich fünf Themenklassen, die sich wiederum aus Unterklassen zusammensetzen. Jede dieser Themenklassen kann dabei auch als eigener, einzelner KVI gesehen werden.

Untersuchungsgegenstand sind die Großstädte (Einwohnerzahl von mindestens 100.000) in Baden-Württemberg. Damit werden in dieser Arbeit acht geographische Orte verglichen, die sogenannten Merkmalsträger bzw. Objekte. Baden-Württemberg wurde gewählt, weil es eine große geographische Heterogenität in den relevanten Parametern für die Fahrradfreundlichkeit, von dem sehr flachen Karlsruhe bis hin zur „Kesselstadt“ Stuttgart – bei gleichzeitig hoher Anzahl an Großstädten, besitzt. Ein weiterer Grund für die Wahl dieses Bundeslands ist die Zielsetzung der Generalisierbarkeit des Anwendungsbeispiels – Baden-Württemberg startet jetzt erst seine Open-Data-Initiativen und hat daher noch wenige zusätzliche Quellen gegenüber den bundes- bzw. weltweit verfügbaren offenen Datenquellen. Dies erleichtert die breite Anwendung und Beurteilung des Verfahrens, da keine Datenquellen verwendet werden, die sonst nicht verfügbar sind. Nach Wahl des Untersu-

chungsgegenstands und der gewählten Indikatoren werden die Daten gesucht, welche für die Umsetzung dieses Index benötigt werden. Die relevanten, offenen Daten sind der Tabelle 1 zu entnehmen. Somit ist dieser Index auch für andere Bereiche Deutschlands einsetzbar. Des Weiteren ist die sowohl räumliche als auch die zeitliche Auflösung der Daten unterschiedlich. Dieses Problem ist wohlbekannt in der Geoinformatik, z. B. bei Vergleichen zwischen Daten aus der Fernerkundung und bestehenden Informationen von staatlichen Informationsquellen. Im Rahmen der Evaluation werden immer die jeweils aktuellsten Daten verwendet. Da der KVI sich jederzeit neu berechnen lässt und die einfließenden Daten transparent sind, lässt sich das Ergebnis dadurch klar einordnen. Räumlich werden Daten auf einzelne Kennzahlen heruntergebrochen, sodass die Unterschiede in der räumlichen Auflösung durch die Transformation die Problematik verlieren.

Zu einigen wichtigen Themen, wie den Baustellen, Winterdienst, Anteil Radverkehr und Diebstählen, liegen zum Zeitpunkt der Arbeit keine offenen Daten vor. Betroffen sind davon 31,82% der zuvor definierten Indikatoren (siehe Abbildung 5). Zusätzlich sind 27,27% der Daten ungeeignet. Da

Datensatz	Datenquelle	Räuml. Auflösung	Aktualität	Ausdehnung	Indikator
Unfallorte	Unfallatlas (Destatis)	–	2016/2017	9 Bundesländer	Unfallbilanz
Niederschlag	DWD	1 km	1881 – 2017	Deutschland	Wetterbedingung
Wind	DWD	200 m	1881 – 2000	Deutschland	Wetterbedingung
DGM	SRTM	~ 30 m	2000	Weltweit	Steigung
Lärm	UBA	–	2017	Deutschland	Lärm
Umweltsensoren	OpenSenseMap	–	Aktuell	Weltweit	Schadstoffe
OSM Lines	OSM	–	Aktuell	Weltweit	Steigung, Oberflächenbeschaffenheit, Fahrradstraßen, verkehrsberuhigte Bereiche, Straßenbreite
OSM Points	OSM	–	Aktuell	Weltweit	Ampeln, Stellplätze Fahrradläden
Fahrradverleih	NextBike	–	Aktuell	Weltweit	Fahrradverleih
Fahrradverleih	Call a Bike, RegioRad	–	Aktuell	Deutschland	Fahrradverleih
Social Media	Twitter	–	Aktuell	Weltweit	Medienpräsenz
Social Media	Google Trends	–	Aktuell	Weltweit	Medienpräsenz
Verwaltungsgrenzen	BKG	–	2018	Deutschland	

Table 1: Datengrundlage zur Berechnung des komplexen Vergleichsindex

von betroffen sind beispielsweise die Indikatoren Stellplätze, Oberflächenbeschaffenheit, Schadstoffe etc. aufgrund unvollständiger Datensätze. Die geringe Datengrundlage führt zu einem unsicheren und irreführenden Ergebnis. Das Problem könnte behoben werden, indem mehr Daten mit einberechnet werden. Dies ist aufgrund fehlender offener Daten derzeit nur begrenzt möglich; in der Zukunft sowie in anderen Untersuchungsräumen kann jedoch von ausreichenden Datengrundlagen ausgegangen werden.

Nachdem die einzelnen Indikatoren für alle Merkmalsträger berechnet und zusammengeführt wurden, wird untersucht, ob fehlende Daten berechnet werden können. Im Beispiel der Fahrradfreundlichkeit betreffen die fehlenden Datensätze die Oberflächenbeschaffenheit, die Straßenbreite und den Lärm. Für die Indikatoren Oberflächenbeschaffenheit und Straßenbreite lassen sich die fehlenden Daten nicht berechnen, da es keinen Anhaltspunkt gibt, in was sich die unbekanntesten Straßenabschnitte aufteilen lassen; zudem sind zu viele Abschnitte unbekannt, um diese zu löschen. Beim Datensatz Lärm haben zwei der Merkmalsträger, Heidelberg und Pforzheim, keine Daten. Anhand der Korrelationsmatrix wird überprüft, ob es eine starke Beziehung zu einer anderen Variablen gibt. Es ist zu erkennen, dass eine Korrelation mit einem Koeffizienten von 0,7 mit den Straßen auftritt. Die Korrelation zwischen Lärm und Straße wird als plausibel erachtet. Da die Regressionsanalyse hier nur mit sechs Stichproben berechnet werden konnte, kann ein Einfluss von Ausreißern nicht ausgeschlossen werden.

Nach der Berechnung der Indikatoren des Themenbereichs *Radverkehrsinfrastruktur* weisen die Indikatoren – mit Ausnahme der Stellplätze – eine hohe Aussagekraft auf. Beim Indikator Stellplätze wird die Aussagekraft angezweifelt, da die Stellplätze mit der Einwohnerzahl korrelieren.

Die Steigung und die Wetterbedingung stellen die aussagekräftigen Indikatoren des Themenbereichs *Komfort* dar. Die fehlenden Werte der Oberflächenbeschaffenheit lassen sich, wie oben bereits beschrieben, nicht berechnen. Damit der Indikator Ampeln aussagekräftig wird, werden Informationen zur Ampelschaltung benötigt.

Der Themenbereich *Gesundheit* setzt sich zusammen aus dem Indikator Lärm und

dem Indikator Schadstoffe. Der Indikator Lärm ist aussagekräftig, wenn nicht berücksichtigt wird, dass die Daten nur für sechs der acht beobachteten Städte vorliegen. Der Indikator Schadstoffe wird als nicht aussagekräftig eingestuft, da aus den Daten, aufgrund starker Ausreißer und zu wenig Messpunkten, kein sinnvolles Raster interpoliert werden kann.

Das Themenfeld *soziales Umfeld* besteht aus einem Indikator, der Medienpräsenz, die als aussagekräftig eingestuft wird.

Der Themenbereich *Sicherheit* setzt sich zusammen aus der Straßenbreite, den verkehrsberuhigten Bereichen und der Unfallbilanz. Die verkehrsberuhigten Bereiche stellen den einzigen aussagekräftigen Indikator dieses Themenbereichs dar. Die Straßenbreite ist, wie die Oberflächenbeschaffenheit, an zu wenigen Straßenabschnitten bekannt. Um einen aussagekräftigen Indikator zu erhalten, sollten die Unfälle pro Radfahrer berechnet werden. Da es keine Datensätze zum Radverkehrsanteil gibt, wird dieser Indikator als nicht aussagekräftig betrachtet.

Als Zwischenzusammenfassung für die multivariate Analyse lässt sich sagen, dass die Aussagekraft der Indikatoren stark von der Qualität und Verfügbarkeit der Datengrundlage abhängt. Es zeigt sich, dass sich der Themenbereich Radverkehrsinfrastruktur gut berechnen lässt. Die Themenbereiche Komfort, Gesundheit, soziales Um-

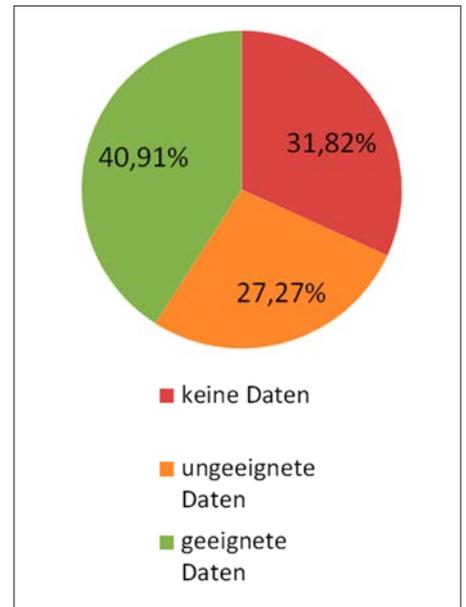


Abbildung 5: Kreisdiagramm mit Anteil der Indikatoren ohne, mit ungeeigneten und mit geeigneten Daten

feld und Sicherheit hingegen haben jeweils nur einen bis zwei aussagekräftige Indikatoren.

Das Ergebnis der Hauptkomponentenanalyse lässt sich Abbildung 6 entnehmen. Basierend auf dem Kaiser-Kriterium (siehe Abschnitt 3) (Backhaus 2018) lassen sich vier Hauptkomponenten identifizieren. Die gebildeten Gruppen sind dabei stark heterogen, was für eine Interpretation als gemeinsame Hauptkomponente nicht zielfüh-

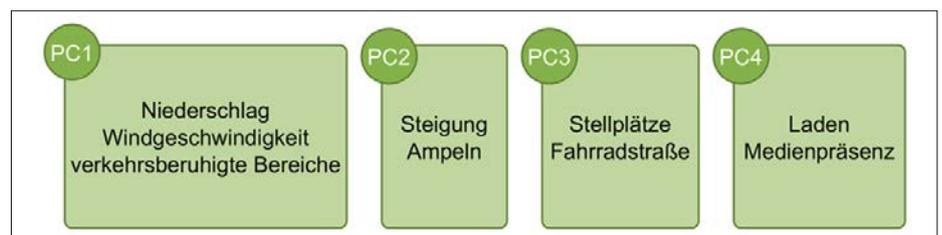


Abbildung 6: Indikatoren in Hauptkomponenten

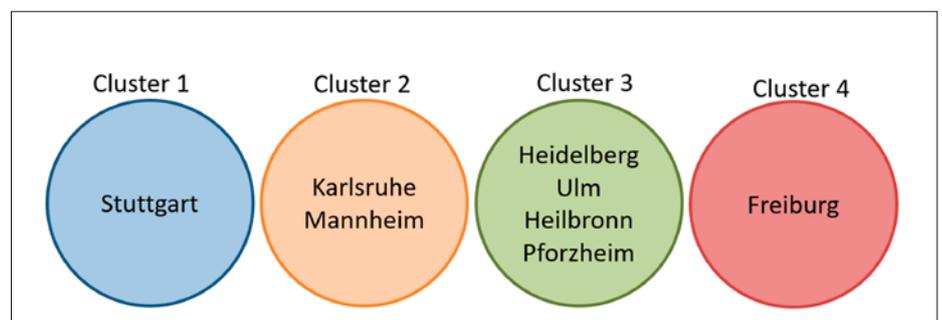


Abbildung 7: Clusteranalyse der Merkmalsträger nach k-Medoids mit vier Klassen

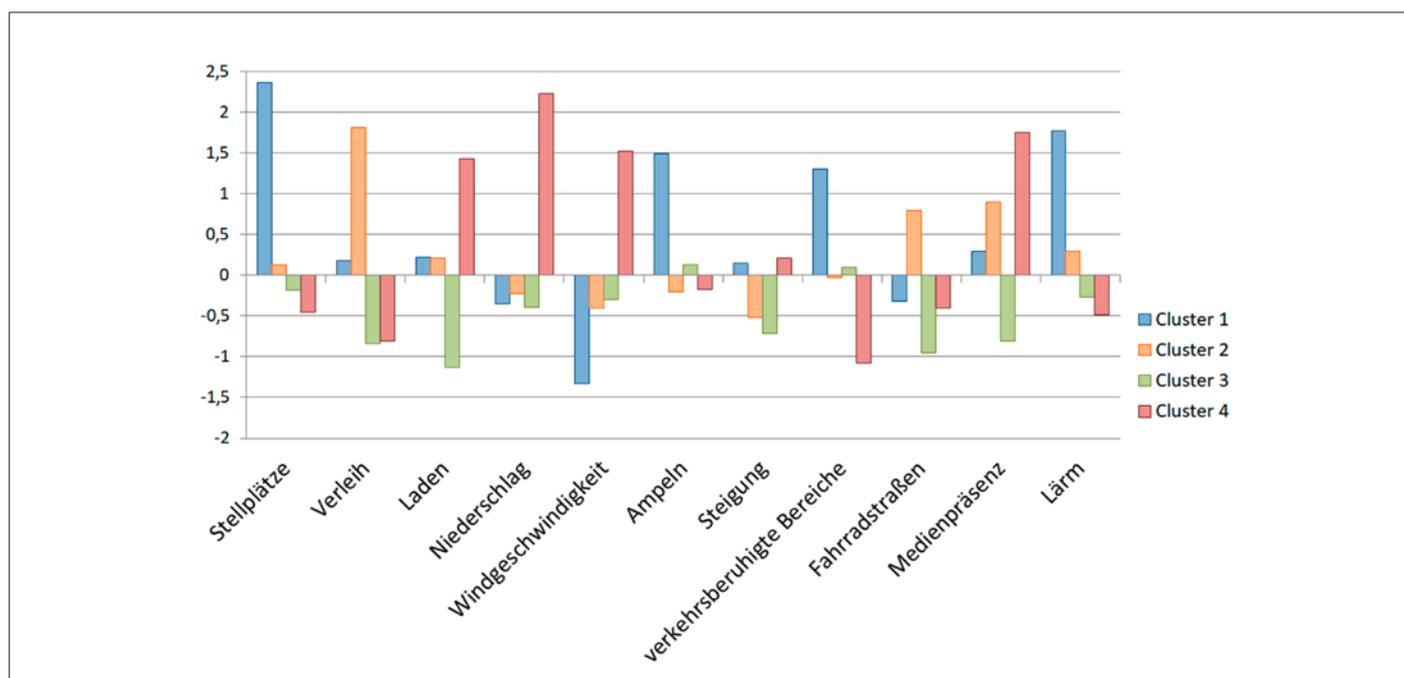


Abbildung 8:  $t$ -Werte der Cluster für die einzelnen Variablen

rend ist. Mithilfe der Wertematrix lässt sich bestimmen, aus welchen Hauptkomponenten sich die Merkmalsträger zusammensetzen (Backhaus 2018). Da in dieser Arbeit keine thematische Verbindung in den Hauptkomponenten auftritt, ist das Berechnen einer Wertematrix nicht notwendig. In dieser Arbeit können aufgrund der Datenbeschaffenheit nur acht Stichproben für die multivariate Analyse verglichen werden. Dies kann für das gewählte Anwendungsbeispiel zu Überanpassung (Overfitting) führen. Die zugrunde liegende Methodik und die weiteren Ergebnisse sind davon jedoch nicht betroffen.

Welche Merkmalsträger sich im selben Cluster befinden und somit über ähnliche Eigenschaften verfügen, ist der Abbildung 7 zu entnehmen. Hierbei fällt auf, dass sich die Merkmalsträger Stuttgart und Freiburg von den anderen Merkmalsträgern unterscheiden und jeweils allein ein Cluster bilden.

Aus dem Diagramm Abbildung 8 mit normierten  $t$ -Werten geht hervor, dass Cluster 1 stark von den Indikatoren Stellplätze, Ampeln, verkehrsberuhigten Bereichen und Lärm repräsentiert wird. Die Indikatoren Verleih und Fahrradstraßen repräsentieren Cluster 2. Cluster 3 hingegen wird von keinem Indikator stark repräsentiert, aber von den Indikatoren Verleih, Laden, Fahrradstraßen und Medienpräsenz unterrepräsentiert. In Cluster 4 sind die Indikatoren La-

den, Niederschlag, Windgeschwindigkeit und Medienpräsenz repräsentativ für das Cluster und die verkehrsberuhigten Bereiche unterrepräsentativ für das Cluster. Da die Cluster keine sinnvolle Interpretation zulassen, ist anzunehmen, dass die Cluster zufällig sind.

Die Unsicherheitsanalyse zeigt, dass der KVI hinsichtlich der Normalisierungsmethode, der Gewichtungsmethode und der Aggregationsmethode in den großen Städten Stuttgart, Karlsruhe und Mannheim robust ist. Das ist daran zu erkennen, dass die Städte konstant das gleiche Ranking erreichen. In Abbildung 9 ist daher für diese

Städte nur der Medianstrich zu sehen. In den restlichen Städten lässt die Robustheit nach. Dies ist auf die teilweise unzureichende Datenlage für offene Daten zurückzuführen. Für Pforzheim mussten hier beispielsweise einige Daten abgeleitet werden. Dies, in Zusammenhang mit einigen überstarken Gewichten wie den Fahrradstellplätzen oder der Medienpräsenz, führt vermutlich zur Instabilität der Ergebnisse für die Städte mit weniger vorhandenen offenen Daten.

Abbildung 10 zeigt eine Visualisierung des in diesem Beitrag berechneten allgemeinen KVI für Fahrradfreundlichkeit. Da

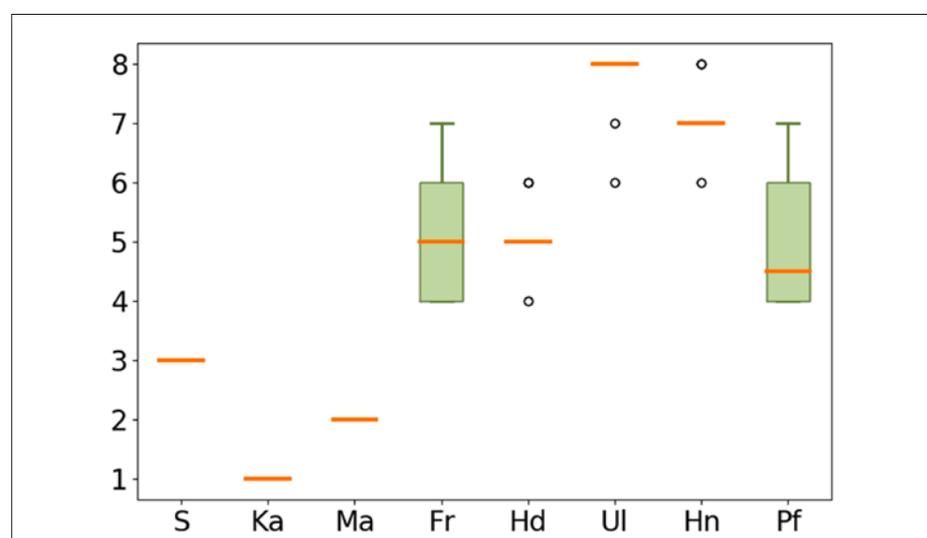


Abbildung 9: Ergebnis der Unsicherheitsanalyse in einem Boxplot-Diagramm

hier zunächst eine methodisch-technische Machbarkeitsanalyse durchgeführt werden soll, bei der die inhaltlichen Aspekte noch zweitrangig sind, wird Indikatoren mit schlechter Datengrundlage ein geringeres Gewicht zugeteilt oder sie werden komplett ausgelassen. Das Ergebnis ist ein individuelles Ranking der Städte. Im weiteren Verlauf des Projekts soll jeder Nutzer die Gewichte nach seinen eigenen Bedürfnissen wählen können. In diesem Beitrag wurde dies für die drei beschriebenen speziellen Nutzergruppen einmal durchgeführt. In Abbildung 11 wird dies für die Familienfreundlichkeit aufgezeigt. Die Indikatoren, denen in diesem Fall mehr Gewicht zugeteilt wurde, sind in Fettschrift gekennzeichnet.

In Abbildung 10 und 11 wird in der Legende die Zusammensetzung der Indikatoren und der gewählten Themenbereiche beschrieben. Die Gewichtung kann (siehe Abbildung 11) besonders hervorgehoben werden. Das Ranking der Städte wird als Tabelle und als aggregiertes Ergebnis der einzelnen Indikatoren angezeigt. Die Balkenlänge zeigt auch den relativen Abstand in der Aggregation an. Für jede Stadt wurde zusätzlich ein Spinnendiagramm erstellt und auf der Karte angezeigt. Dies beschreibt, in welchen der Kategorien aus der Legende die einzelnen Städte wie abgeschnitten haben. Dies erleichtert die weitere Beurteilung des Gesamrankings und verdeutlicht, wie sich die Städte unterscheiden.

Tabelle 2 zeigt kurz die Ergebnisse für die Profile „Sportfahrer mit Rennrad“ und „Freizeifahrer mit Elektrofahrrad“. Es zeigt

sich dabei, dass sich bei allen Profilen Karlsruhe als fahrradfreundlichste Stadt platziert, die einzelnen Gewichtungen aber stark das einzelne Ranking beeinflussen. Gleichzeitig ist durch die Spinnendiagramme und die Legende mit Zusammensetzung des Ergebnisses der Grund für das Ranking klar und offen ersichtlich.

Beim Vergleich der Ergebnisse der hier erstellten KVI mit bestehenden Indizes lassen sich Unterschiede erklären. Insbesondere beim Vergleich mit dem Fahrradklimatest des ADFC zeigen sich unerwartete Resultate – vor allem das relativ gute Ranking von Stuttgart. Diese Ergebnisse lassen sich an der Tatsache erklären, dass nicht alle relevanten Indikatoren berücksichtigt werden konnten. Die dafür benötigten Daten mit räumlichem Bezug zu den gewählten Städten sind oft nicht oder nicht durchgehend verfügbar. Ein weiterer Grund ist, dass keine weichen Faktoren mit einberechnet werden, wie dies z. B. durch Umfragen geschehen kann. Die dadurch erfasste subjektive Wahrnehmung von Fahrradfreundlichkeit kann fehlende Daten ersetzen, führt jedoch zu einer geringeren Nachvollziehbarkeit. Dies kann jedoch auch klar in den Ergebnissen nachvollzogen werden und insbesondere durch die Legende (Zusammensetzung der Ergebnisse) und die Spinnendiagramme vermittelt werden.

## 6 DISKUSSION UND AUSBLICK

In diesem Beitrag wurde ein Framework für die Erstellung räumlicher KVI auf Basis offener Geodaten vorgestellt und am Beispiel eines Fahrradfreundlichkeitsindex in der

Praxis evaluiert. Das Ziel dieses Framework war es, die Erstellung und Verwendung von KVIs zu demokratisieren. Um dies zu erfüllen, wurden sechs Anforderungen als Ziel definiert, die in der Praxis erfüllt werden sollen. Diese Eigenschaften sind aus dem aktuellen Stand der Forschung zu KVI abgeleitet. Es konnte gezeigt werden, dass mit der vorliegenden Methodik vier von sechs dieser Eigenschaften uneingeschränkt erreicht worden sind: Die Transparenz des Index, die Anpassbarkeit des Index auf individuelle Lebenssituationen, die geographisch unabhängige Anwendbarkeit und die Erstellung eines KVI rein auf quantitativen Werten. Bei der Berechnung des KVI kann jeder einzelne Schritt nachvollzogen und individuell beurteilt werden. Gerade bei den teils unerwarteten Ergebnissen kann gesehen werden, wodurch diese im Ranking entstehen, wie diese angepasst werden könnten und wie die Aussage daher zu bewerten ist. Die eingeflossenen Parameter und die Rankings in den gewählten Themenbereichen sind in Abbildung 10 direkt sichtbar. Die Anpassbarkeit lässt sich leicht und individuell über die Wahl der Einzelgewichte erreichen und kann so die Bedürfnisse verschiedenster Gruppen und Lebensabschnitte abbilden. Dies wurde durch drei Profile von Nutzergruppen abgebildet und die Ergebnisse haben sich dabei ausreichend unterschieden.

Nur eingeschränkt konnten die Ziele der einfachen Verwendbarkeit für Nichtexperten und der reinen Nutzung von offenen Daten erreicht werden. Sowohl die Regres-

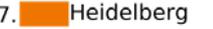
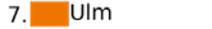
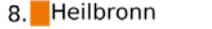
	Sportfahrer mit Rennrad	Freizeifahrer mit Elektrofahrrad
Indikatoren, die stärker gewichtet werden	Stellplätze, Fahrradläden, Fahrradstraßen, Ampeln, Niederschlag, Wind, Lärm	Fahrradläden, Fahrradstraße, Ampeln, Niederschlag, Medienpräsenz
Ranking der Städte	<ol style="list-style-type: none"> <li>1.  Karlsruhe</li> <li>2.  Mannheim</li> <li>3.  Freiburg</li> <li>4.  Stuttgart</li> <li>5.  Pforzheim</li> <li>6.  Ulm</li> <li>7.  Heidelberg</li> <li>8.  Heilbronn</li> </ol>	<ol style="list-style-type: none"> <li>1.  Karlsruhe</li> <li>2.  Pforzheim</li> <li>3.  Stuttgart</li> <li>4.  Mannheim</li> <li>5.  Freiburg</li> <li>6.  Heidelberg</li> <li>7.  Ulm</li> <li>8.  Heilbronn</li> </ol>

Tabelle 2: Ergebnis des komplexen Vergleichsindex für die Nutzergruppen Sportfahrer und Freizeifahrer

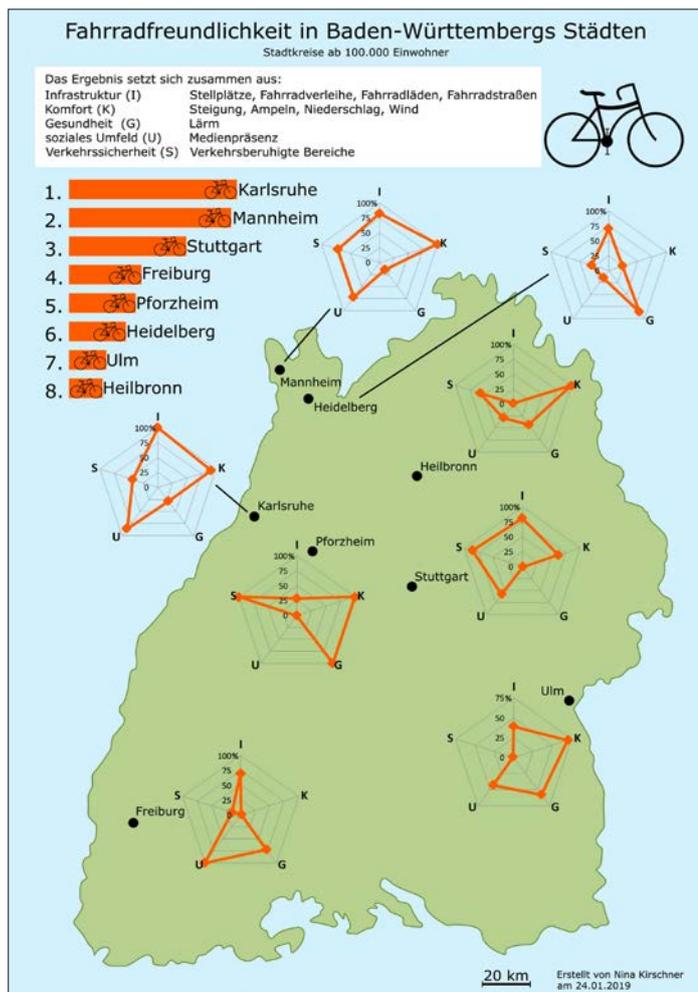


Abbildung 10: Visualisierung des komplexen Vergleichsindex als Infografik

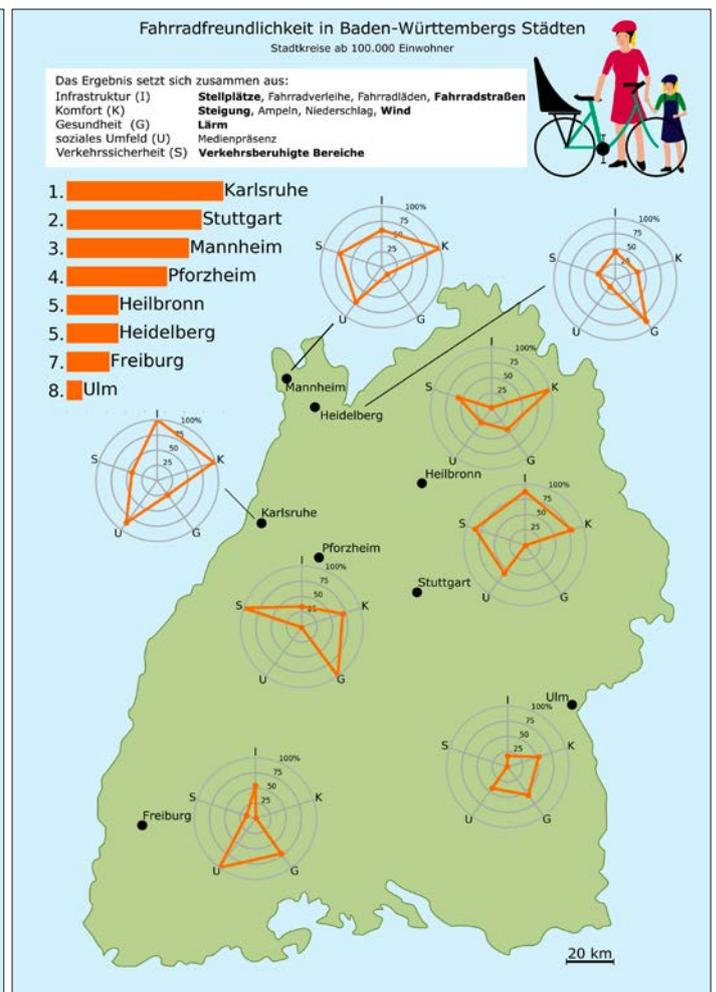


Abbildung 11: Visualisierung des komplexen Vergleichsindex mit Schwerpunkt auf Familienfreundlichkeit als Infografik

sionsanalyse, die multivariate Analyse und die Bewertung der Robustheit sind aufwendige Verfahren, die viel Zeit in Anspruch nehmen. Zusätzlich wird zur korrekten Kombination und Interpretation der Zwischenergebnisse und damit Auswahl der Parameter ein Grundwissen in der Beurteilung und Methodik benötigt. Des Weiteren sind diese Methoden nur geeignet, wenn viele Objekte betrachtet werden. Da in diesem Beitrag acht Merkmalsträger miteinander verglichen werden, folgt daraus, dass Probleme bei der Interpretation der Ergebnisse dieser Methoden auftreten. Das Problem bei Indizes, dass das Ergebnis häufig fehlinterpretiert wird, wird bei einem Weglassen der Analysen verstärkt. Andererseits ist die Möglichkeit, dies überhaupt zu betrachten, als neuartig zu betrachten. Die ausschließliche Nutzung von offenen Daten konnte erreicht werden, jedoch hat sich bei der Evaluation gezeigt, dass heutzutage noch viele benötigte Daten nur unzureichend vorhanden sind. Dies zeigt sich vor

allem in dem relativen hohen Ranking der Stadt Stuttgart, da hier einige Faktoren, wie das Sicherheitsgefühl und die Verkehrslage/Radinfrastruktur, nicht zufriedenstellend abgebildet werden konnten. Gleichzeitig hat sich hierbei gezeigt, dass genau diese Lücken und die daraus resultierenden Probleme transparent aufgezeigt werden konnten. Ein Nutzer kann dies erkennen und die Ergebnisse entsprechend einordnen.

Insgesamt lässt sich hieraus der Schluss ziehen, dass das vorgestellte Framework zur Erstellung von KVI die gesteckten Erwartungen aus methodischer und aus softwaretechnischer Sicht erfüllen kann. Aber die erstellten KVI sind in ihrer Qualität enorm von den verfügbaren Daten abhängig. Bei kritischer Reflektion der berechneten Rankings und Vergleich mit Erfahrungen aus eigener Anschauung der Fahrradfahrsituation in den jeweiligen Städten sind die KVI-Werte an vielen Stellen unerwartet, kontraintuitiv oder offensichtlich unsinnig. Dies hängt ein-

zig an der Verfügbarkeit und Qualität offener Daten. Hier ist die Politik gefordert, die seit Jahren formulierten Ziele der Bereitstellung sinnvoller, qualitativ hochwertiger Datengrundlagen auch umzusetzen. Zivilgesellschaft (wie die OSM-Community) und Wissenschaft können hier helfen, aber sie können ordentliche Prozesse und Infrastrukturen zur Bereitstellung von Open Government Data nicht ersetzen. Immerhin: auch wenn das Problem der Abhängigkeit von der Qualität der Eingangsdaten bei jedem Index auftaucht, fällt es in unserem Framework durch die ausschließliche Verwendung von quantitativen Daten und Methoden stärker auf, da die Anforderung der Transparenz erfüllt wird.

In Zukunft wird davon ausgegangen, dass die Problematik der Datengrundlage stark reduziert wird. In Deutschland haben in den letzten Jahren immer mehr Verwaltungseinheiten begonnen, ihre Daten offen zur Verfügung zu stellen (Kloiber 2014). Mithilfe weiterer öffentlicher Daten

kann der erstellte Index erweitert werden, wodurch die Aussagekraft erhöht wird. Seit dem 12.12.2019 hat z.B. das Land Baden-Württemberg ein neues Geoportal für offene Geodaten (Baden-Württemberg 2019a), ebenso ist ein allgemeines Datenportal für offene Daten im Aufbau (Baden-Württemberg 2019b). In Bundesländern, wie Nordrhein-Westfalen und Rheinland-Pfalz, ist dies bereits stärker fortgeschritten. Dies bekräftigt die Annahme, dass die Ergebnisse und Verfahren dieser Arbeit in Zukunft noch stärker in der Praxis Anwendung finden werden und gleichzeitig die Qualität stetig steigen wird.

Als zukünftige Erweiterung der hier vorgestellten Methodik ist die Anwendung auf weitere Sachverhalte und räumliche Gebiete geplant, um damit die Generalisierbarkeit des Ansatzes zu zeigen. Da im vorliegenden Beitrag zunächst methodische und technische Grundlagen untersucht wurden, sind sicherlich auch zum Thema Benutzerfreundlichkeit von KVI-Werkzeugen und zur weitergehenden Automatisierung der Abläufe noch vielfältige zukünftige Arbeiten erforderlich.

Eine weitere Erweiterung besteht in der Betrachtung der sich ändernden Ergebnisse über die Zeit. Bei jeder Berechnung des komplexen Vergleichsindex entsteht ein anderes Ergebnis, da die aktuellsten Daten für

die Berechnung verwendet werden. Je nach Jahreszeit und Uhrzeit können sich die einzelnen Indikatoren ändern. Wird dieser Index über einen längeren Zeitraum verwendet, kann die Veränderung des Index verglichen und ausgewertet werden. Dies ermöglicht es Planern, Entscheidern und Gutachtern, auch den Einfluss von planerischen und steuernden Entscheidungen (wie z.B. Ausbau des Fahrradwegenetzes, Ändern der Verkehrssteuerung u.Ä.) über die Zeit hinweg zu beobachten.

Diese Arbeit ist im Rahmen des mFUND-Projekts WEKOVI durchgeführt worden und die Ergebnisse werden darin weiterentwickelt. Um die Anwendbarkeit für Nichtexperten zu erleichtern, wird der hier erstellte komplexe Vergleichsindex zunächst in der Software StreamPipes nachgebaut. StreamPipes ist ein Verarbeitungstool für Datenströme und wurde vom Forschungszentrum für Informatik (FZI) entwickelt (FZI 2019). Das übergeordnete Ziel des Projekts WEKOVI ist es, eine benutzerfreundliche und erweiterbare Software-Plattform, welche Anwender bei der Erstellung, Berechnung, Veröffentlichung und kollaborativen Weiterentwicklung komplexer Vergleichsindizes (KVI) auf der Basis offener Verwaltungsdaten unterstützt (BMVI 2018). Da auch StreamPipes das Ziel verfolgt, komplexe Datenverarbeitungsprozesse durch Domänenexperten

ohne vertiefte Programmierkenntnisse modellieren zu lassen, liegt eine Erweiterung auf KVI-Berechnungsprozesse durch Realisierung von Konnektoren zu Geodatenquellen und Open-Data-Portalen nahe, ebenso wie die Realisierung KVI-orientierter Operatoren für die Verarbeitung von Geodaten.

Zuletzt ist die Entwicklung einer kollaborativen Komponente für WEKOVI geplant. Hier geht es darum, dass jeder Interessierte eigenständig KVIs entwickeln und auch veröffentlichen kann. So können erstellte Analyse-Pipelines in StreamPipes ausgetauscht werden. Verschiedene Anwender können gewählte Gewichtungen, genutzte Parameter und zugrunde liegende Datenquellen dokumentieren, veröffentlichen und in der Gruppe diskutieren bzw. individuell oder in der Gruppe weiterentwickeln. Dies ermöglicht eine kritische und eigenständige Auseinandersetzung mit den Themen im Sinne von mündigen Bürgern bzw. dem Gedanken der Citizen Science (Resch 2013).

#### DANKSAGUNG

Die Verfasser danken dem Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) für die Förderung im Rahmen des mFund Projekts WEKOVI (FKZ: 19F202 0A).

## Literatur

- Adam, B. (2018): Digitalisierung in der Immobilienbewertung. In: Gündling, H.; Schulz-Wulkow, C. (Eds.): Next Generation Real Estate. Frankfurt School, Frankfurt/M.
- ADFC (2018a): Fahrrad oder Auto? ADFC Baden-Württemberg e.V.
- ADFC (2018b): ADFC-FAHRRADKLIMA-TEST 2018. <http://www.fahrradklima-test.de/>, Zugriff 10/2018.
- Anselin, L. (1989): What is Special about Spatial Data? Alternative Perspectives on Spatial Data Analysis. National Center for Geographic Information and Analysis, Technical Paper 89-4. Santa Barbara, CA.
- Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R. (2018): Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. Springer Gabler, Wiesbaden.
- Baden-Württemberg (2019a): Geoportal Baden-Württemberg. <https://www.geoportal-bw.de/>, Zugriff 12/2019.
- Baden-Württemberg (2019b): Datenportal Baden-Württemberg. <https://datentest.service-bw.de/>, Zugriff 12/2019.
- Barclay, M.; Dixon-Woods, M.; Lyratzopoulos, G. (2018): The problem with composite indicators. In: *BMJ Quality & Safety*, 4, S. 338-344.
- Bassolas, A.; Barbosa-Filho, H.; Dickinson, B. et al. (2019): Hierarchical organization of urban mobility and its connection with city livability. In: *Nat Commun.*, 10, Art. Nr. 4817. doi:10.1038/s41467-019-12809-y.
- Bill, R. (2016): Grundlagen der Geo-Informationssysteme. Wichmann, Berlin/Offenbach.
- BMVI (2018): BMVI – Werkzeuge für die einfache Erstellung komplexer Vergleichsindizes – WEKOVI. <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/mfund-projekte/wekovi.html>, Zugriff 04/2020.
- Copenhagenize (2018): The Criteria for the Copenhagenize Index. <http://copenhagenizeindex.eu/criteria.html> Handbuch, Zugriff 10/2018.
- Dörrzapf, L.; Kovács-Gyori, A.; Resch, B.; Zeile, P. (2019): Defining and assessing walkability: a concept for an integrated approach using surveys, biosensors and geospatial analysis. In: *Urban Development Issues*, 62 (1), S. 5-15.
- Economist (2019): The Big Mac index. <https://www.economist.com/news/2019/07/10/the-big-mac-index>, Zugriff 08/2019.
- Ewing, R.; Hamidi, S. (2015): Compactness versus Sprawl: A Review of Recent Evidence from the United States. In: *Journal of Planning Literature*, 30 (4), S. 413-432.
- Fahrmeir, L. (2015): *Multivariate statistische Verfahren*. W. de Gruyter, Berlin.
- Flitter, H.; Laube, P.; Hägi, S. (2016): Eignungsanalysen. <http://gitta.info/Suitability/de/text/Suitability.pdf>, Zugriff 02/2019.
- Forster, N. (2019): Analyse von Gefahrenstellen zwischen Pkw- und Fahrradfahrern Mithilfe von Geodaten. In: *gis.Science*, 2/2019.
- FZI (2019): StreamPipes Homepage. <https://www.streampipes.org/>, Zugriff 02/2019.
- Gehring, D. (2016): Bikeability – Index für Dresden – Wie fahrradfreundlich ist Dresden? <https://nbn-resolving.org/urn:nbn:de:bsz:14-qcoca-224390>, Zugriff 12/2019.
- Google (2019): <https://ai.googleblog.com/2019/11/new-insights-into-human-mobility-with.html>, Zugriff 12/2019.
- Kloiber, J. (2014): Von offenen Daten zu nützlichen Anwendungen – Ein kollaborativer Prozess. In: Hill, H.; Martini, M.; Wagner E. (Hrsg.): *Transparenz, Partizipation, Kollaboration: Die digitale Verwaltung neu denken*. Nomos, Baden-Baden, S. 203-210.
- KNIME (2019): KNIME Software. <https://www.knime.com/knime-software>, Zugriff 01/2019.
- Krenn, P.; Oja, P.; Titze, S. (2015): Development of a Bikeability Index to Assess the Bicycle-Friendliness of Urban Environments. In: *Open Journal of Civil Engineering*, 5/2015.
- Li, S.; Dragicevic, S.; Anton, F.; Sester, M.; Winter, S.; Coltekin, A.; Pettit, C.; Jiang, B.; Harworth, J.; Stein, A.; Cheng, T. (2015): Geospatial Big Data Handling Theory and Methods: A Review and Research Challenges. In: *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, S. 119-133.
- Loidl, M.; Zagel, B. (2014): Assessing bicycle safety in multiple networks with different data models. In: Vogler, R.; Car, A.; Strobl, J.; Griesebner, G. (Eds.): *GI-Forum 2-2016*. Wichmann, Berlin/Offenbach.
- Mazziotta, M.; Pareto, A. (2013): Methods for constructing composite indices: One for all or all for one. In: *Rivista Italiana di Economia Demografia e Statistica*, 67 (2), S. 67-80.
- Motta, B. G. (2017): A bikeability index for Curitiba (Brazil). <https://essay.utwente.nl/71736/>, Zugriff 12/2019.
- Munda, G.; Nardo, M. (2005): Constructing consistent composite indicators: the issue of weights. In: *EUR 21834 EN*, 2005.
- Nardo, M.; Saisana, M.; Saltelli, A.; Tarantola, S. (2005): Tools for composite indicators building. <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC31473/EUR%2021682%20EN.pdf>, Zugriff 1/2019.
- OECD (2008): *Handbook on constructing composite indicators: Methodology and user guide*. <http://www.oecd.org/els/soc/handbook-on-constructing-composite-indicators-methodology-and-user-guide.htm>, Zugriff 10/2018.
- Resch, B. (2013): *People as Sensors and Collective Sensing – Contextual Observations Complementing Geo-Sensor Network Measurements*. In: *Progress in Location-Based Services*. Springer, Berlin/Heidelberg, S. 391-407.
- Saltelli, A. (2007): Composite indicators between analysis and advocacy. In: *Social Indicators Research*, 81 (1), S. 65-77.
- UNDP (2019): Human Development Index (HDI). <http://hdr.undp.org/en/content/human-development-index-hdi>, Zugriff 08/2019.
- Winters, M.; Brauer, M.; Setton, E. M.; Teschke, K. (2013): Mapping Bikeability: A Spatial Tool to Support Sustainable Travel. In: *Environment and Planning B: Planning and Design*, 40.
- Wirtschaftswoche (2019): Städteranking 2019. <https://www.wiwo.de/politik/deutschland/staedteranking-2019-deutschlands-zukunftsfaehigste-staedte/25231576.html>, Zugriff 12/2019.