



INSTITUT FÜR ENERGIE-  
UND UMWELTFORSCHUNG  
HEIDELBERG

# Roadmap O-Lkw - AP5: Berücksichtigung realer Lkw Einsatzprofile

Status der Arbeiten und erzielte Ergebnisse

Jan Kräck, Hinrich Helms & Maybritt Schillinger

Heidelberg, September 2020

---



Das Vorhaben wurde im BMUB-Förderprogramm „Erneuerbar Mobil“  
unter der FKZ 16EM3153-1 gefördert

#### IMPRESSUM

Autoren: Jan Kräck, Hinrich Helms & Maybritt Schil-  
linger

Herausgeber: ifeu – Institut für Energie- und Umweltfor-  
schung Heidelberg GmbH  
Wilckensstraße 3, D-69120 Heidelberg

Erscheinungsjahr: 2020

# Inhalt

---

<b>Abbildungsverzeichnis</b>	<b>4</b>
<b>Tabellenverzeichnis</b>	<b>5</b>
<b>1 Verlauf und Status von AP 5</b>	<b>6</b>
<b>2 Geplante Arbeiten in AP 5</b>	<b>8</b>
AP 5.1: Analyse realer Lkw-Einsatzdaten	8
AP 5.2: Abgleich des Netzausbaumodells mit realen Lkw-Einsatzdaten	10
AP 5.3: Schlussfolgerungen hinsichtlich des Roadmappings	11
<b>3 Praktische Durchführung und Herausforderungen im AP 5</b>	<b>12</b>
<b>4 AP 5.1: Analyse realer Lkw-Einsatzdaten</b>	<b>14</b>
4.1 Aufbereitung des Datensatzes	14
4.2 Analyse der internationalen Verkehre und deren Einfluss auf das Elektrifizierungspotenzial	17
4.3 Entwicklung einer Methodik zu Identifizierung von Pendelverkehrsanteilen	21
4.3.1 Definition und Beschreibung der untersuchten Features	22
4.3.2 Gruppierung mittels Clustering-Verfahren	24
4.3.3 Bewertung der Clusteringverfahren	25
4.4 Ergebnisse	31
<b>5 AP 5.2: Abgleich des Netzausbaumodells mit realen Lkw-Einsatzdaten</b>	<b>33</b>
5.1 Methodik	33
5.2 Ergebnisse	34
<b>6 AP 5.3: Schlussfolgerungen hinsichtlich des Road Mappings</b>	<b>36</b>
<b>Literaturverzeichnis</b>	<b>37</b>



# Abbildungsverzeichnis

---

Abbildung 1: Marktanteile Scantias an den Nutzfahrzeugen der Klasse über 3,5t in den Jahren 2010 bis 2017 ohne Linien- und Reisebusse in der EU und den Ländern der Europäische Freihandelszone (EFTA). Eigene Darstellung basierend auf Daten der (“ACEA - European Automobile Manufacturers’ Association” 2018)	14
Abbildung 2: Anteil der Lkw-Alterklassen an den zurückgelegten Fahrzeugkilometern in ausgewählten europäischen Ländern im Jahr 2017. Eigene Darstellung basierend auf (EUROSTAT 2019a)	15
Abbildung 3: Fahrleistungen in Mio. Fahrzeugkilometern von Lkw (beladen) nach zulässigem Gesamtgewicht unterteilt in ausgewählten Ländern der EU im Jahr 2017. Eigene Darstellung basierend auf (EUROSTAT 2019a)	16
Abbildung 4: Fahrleistungen nationaler und internationaler Straßentransporte durch beladene und unbeladene Lkw in ausgewählten Ländern Europas im Jahr 2017. Eigene Darstellung basierend auf (EUROSTAT 2019a)	17
Abbildung 5: Transitfahrten in Tausend Fahrten von Lkw der Meldeländer durch Deutschland im Jahr 2017. Eigene Darstellung basierend auf (EUROSTAT 2019a)	18
Abbildung 6: Mittlere Preise für Diesel und Strom (Gewerbekunden) inklusive Steuern und Angaben in ausgewählten europäischen Ländern im Jahr 2018. Eigene Darstellung basierend auf (European Environment Agency 2019) und (EUROSTAT 2019b)	19
Abbildung 7: Zusätzliches Potenzial an elektrifizierter Fahrleistung aufgrund des Infrastrukturausbaus in den entsprechenden Nachbarländer und der Umstellung zusätzlicher Lkw der internationalen und Transitverkehre auf den Oberleitungsbetrieb. Exemplarische Darstellung mit Ersatzwerten, da die bisherigen Zwischenergebnisse nicht zugänglich sind.	20
Abbildung 8: Auszug einer Tabelle mit Nutzern und einem zugeordneten Featureset.	24
Abbildung 9: Exemplarische Darstellung der Durchfahungshäufigkeit einzelner Abschnitte durch einen Nutzer	26
Abbildung 10: Exemplarischer Graphenplot eines Nutzers, der die Anzahl der Durchfahrungen von Routen (Kanten) darstellt.	26
Abbildung 11: Ausschnitt der Rangfolge hinsichtlich der Unterscheidungs- bzw. -Vorhersageperformance von Pendelfahrten verschiedener Featuresets.	27
Abbildung 12: Clustering mit KMeans und den Features num_visits_med, var_counts und mfe_un	29

Abbildung 13: Clustering mit DBSCAN und den Features num_visits_med, long_cycles_per_edge, sim_score, min_samples = 7, eps = 0.8	29
Abbildung 14: Ausbaunetz 2030 des Infrastrukturausbaumodells (ifeu) und Straßennetz (E-Roads) des Datenmodells bei Scania.	34
Abbildung 15: Automatische Zählstellen entlang der E-Roads sowie auf weiteren Autobahnen und Bundesstraßen innerhalb Deutschlands.	35

## Tabellenverzeichnis

---

Tabelle 1: Geeignetste Features im Falle des K-Means-Algorithmus zur Klassifizierung von Pendelverkehren.	28
Tabelle 2: Geeignetste Features im Falle des DB-Scan-Algorithmus zur Klassifizierung von Pendelverkehren.	28
Tabelle 3: Gegenüberstellung der Vor- und Nachteile des K-MEANS- und DB-SCAN- Algorithmus.	30
Tabelle 4: Vergleich der performantesten Features zur Selektion von Pendelverkehren	32

# 1 Verlauf und Status von AP 5

---

Eingangs soll kurz der Verlauf und abschließende Status der Arbeiten im AP 5 nachgezeichnet werden:

- Ein erster Austausch und Kontakt bezüglich der Floating-Car-Data von Scania kam im Juli 2017 im Nachgang an die damalige ERS-Konferenz zustande, auf der Analysen auf Basis dieser Daten dargestellt wurden und Scania eine prinzipielle Verwendung dieser Daten offerierte.
- Im April 2018 erfolgte eine konzeptionelle Abstimmung mit Scania zu Einbindungsmöglichkeiten dieser Daten in den Kontext des Gesamtvorhabens.
- Im September 2018 kam dann die Einwilligung seitens Scania, dass diese Daten zur Identifikation von geeigneten Profilen für den Oberleitungsausbau gemeinsam genutzt werden könnten. Im September, November und Dezember 2018 fanden weitere Abstimmungsgespräche hierzu statt.
- Im April 2019 fanden während eines Forschungsaufenthaltes im Rahmen des COLLERS-Projektes erste persönliche Treffen bei Scania statt, in denen weiterführende technische Diskussionen zur Verwendung der Daten geführt sowie Vertraulichkeits- und Geheimhaltungsbedingungen geklärt wurden. Zusätzlich erfolgte eine erste Einführung und Sichtung des Datensatzes.
- Im Rahmen dieser Gespräche wurde klar, dass die Analyse der Daten zur Nutzung im Projekt mit großem Aufwand verbunden ist, woraufhin ein Antrag zur Aufstockung des Projektvorhabens „Roadmap O-Lkw“ als zusätzliche Arbeitspaket 5 gestellt wurde, der im August 2019 bewilligt wurde. In der Zwischenzeit fand weiterhin ein kontinuierlicher fachlicher Austausch zur Datennutzung sowie zu Modellannahmen und Methoden statt.
- Vom Zeitpunkt der Bewilligung an, wurde die Zusammenarbeit mit Scania und die Datenanalysetätigkeit intensiviert: Anfang September 2019 erfolgte ein einwöchiger Arbeitsbesuch bei Scania der zunächst zur Diskussion vieler methodischer Aspekte im Zusammenhang der Modellierung genutzt wurde. Ferner wurden sowohl die Arbeiten zur Analyse der internationalen Verkehre weitergeführt als auch erste Analysen zur Identifizierung von Pendelverkehren begonnen. Änderungen am Datensatz wurden laufend mit den jeweiligen zuständigen Ansprechpartnern bei Scania erörtert.
- Ende Oktober 2019 wurde die Zusammenarbeit seitens Scania offiziell pausiert. Ein Zugriff auf die Datenbank wurde zunächst noch bis Ende 2019 eingeräumt. De Facto war die Datenbank jedoch phasenweise nicht mehr für ifeu erreichbar. Kontinuierliches Arbeiten an den Daten war mit der nun eingeschränkten Hilfestellung seitens des dortigen Teams, z.B. bei technischen und inhaltlichen Problemen, nur noch phasenweise möglich.
- Durch den zeitgleichen Verantwortlichkeitswechsel auf verschiedenen Ebenen, herrschte Unsicherheit über den Fortgang des Projektes, das daher vorläufig beendet wurde.
- Mit dem Projektträger wurde im Februar 2020 vereinbart sich im weiteren Projektverlauf auf die Dokumentation der bereits durchgeführten methodischen Arbeiten zu fokussieren.

Die bereits durchgeführten methodischen Arbeiten und erste Ergebnisse werden in diesem Bericht dokumentiert um zu zeigen, welcher Nutzen aus den Datenanalysen für zentrale Fragestellungen bzgl. Einführung eines O-Lkw-Systems zu erwarten ist. Dafür werden

- die Motivation des Vorhabens und die ursprünglich geplanten Arbeiten dargestellt,
- die praktische Durchführung und die zentralen Herausforderungen beschrieben und
- die bisherigen Ergebnisse der einzelnen Unterarbeitspakete umfangreich dokumentiert.

Der methodische Arbeitsstand ermöglicht bei einem Zugang zu weiteren Daten den direkten Einstieg in eine Datenanalyse



## 2 Motivation und geplante Arbeiten in AP 5

---

Verkehrsmodelle wie PTV Validate, das im Projekt Roadmap O-Lkw eingesetzt wird, bestimmen die Verkehre des Güterverkehrs auf Basis von Warenströmen und den damit verbundenen Transportfahrten. Damit können jedoch keine Rückschlüsse auf die realen Einsatzprofile einzelner Nutzfahrzeuge gezogen werden. Streckenverläufe, Fahrtstrecken und Wegeketten entsprechen somit eher den Warenströmen zwischen Orten, als dem tatsächlichen Fahrtenverlauf von Fahrzeugen. Um diese zu untersuchen, sollten zur Unterstützung der bisherigen Methodik Analysen auf Basis der Floating-Vehicle-Daten von Scania erfolgen. Damit können potenziell vorteilhafte Streckenabschnitte für einen vorrangigen Ausbau auf Basis realer Fahrzeugdaten, inklusive internationaler Verkehre, identifiziert werden.

Die Arbeiten sollten das Gesamtvorhaben ergänzen und wurden im Rahmen einer Aufstockung beauftragt (siehe Abschnitt 1). Ziel war es, reale Fahrzeugnutzungsdaten der Firma Scania für das Gesamtvorhaben nutzbar zu machen und die Aussagekraft und Robustheit der Analysen mit Blick auf die anstehenden Investitionsentscheidungen für den Oberleitungsausbau sowie die Entwicklung von Fördersystemen zu stärken. Als Aufstockung wurden die Arbeiten im Gesamtvorhaben Roadmap O-Lkw in einem zusätzlichen Arbeitspaket 5 zusammengefasst. Die Bearbeitung war in drei Schritten geplant, die als Unterarbeitspakete ausgewiesen wurden.

### AP 5.1: Analyse realer Lkw-Einsatzdaten

Dieses Arbeitspaket umfasst verschiedene Auswertungen auf den Nutzungsdaten von Fahrzeugen der Firma Scania. Europaweit sind etwa 400.000 Scania-Nutzfahrzeuge im Einsatz, wovon ca. 100.000 Fahrzeuge über ein Telematiksystem hoch aufgelöste Betriebsdaten an ein von Scania betriebenes Portal senden. Es liegen also Fahrzeugnutzungsdaten für ganz Europa vor, wobei der Marktanteil von Scania sich zwischen den Ländern teils deutlich unterscheidet (Dies macht eine Normalisierung und Skalierung der Daten notwendig, s.u.). Die Daten sind auf einem HADOOP-Rechencluster<sup>1</sup> gespeichert. Dieses ermöglicht die performante Ausführung von Auswertungsroutinen auf dem gesamten Datensatz oder Teilen davon. In den Datensätzen sind u.a. folgende Parameter der Fahrzeuge enthalten:

- Fahrzeugtyp
- Konfiguration / Ausstattung

---

<sup>1</sup> Computer-Cluster, das für die Speicherung und Analyse von großen Mengen strukturierter und unstrukturierter Daten in einer verteilten Rechenumgebung eingesetzt wird, da mittels eines herkömmlichen Computersystems die vorliegenden Datenmengen nicht mehr in vernünftiger Rechenzeit verarbeitet werden können.

- Herkunftsland
- Während des Fahrzeugbetriebs gesendete Daten, u.a.
  - GPS-Position
  - Momentane Geschwindigkeit
  - Kumulierter Kraftstoffverbrauch

Im Einzelnen waren in diesem Arbeitspaket folgende Analysen geplant, die teilweise auch wichtige Grundlagen für das nachfolgende AP 5.2 darstellen sollten:

- Vorbereitende statistische Analysen
  - Extraktion statistischer Daten zu den Fahrzeugen (Zulassungsort, Aktionsradien, Wegelängenverteilungen, Aufenthaltsorte (Länder), Beladungsfaktoren, ...).
  - Entsprechende Zusammenfassung der Fahrzeuge zu sinnvollen Gruppen für die weiteren Auswertungen
  - Extraktion statistischer Daten zu den einzelnen Relationen in Deutschland (Befahrungsstärke, Kraftstoffverbrauch, Anteil von Pendelfahrten auf der Relation)
- Implementierung des in AP 1.4 des Gesamtvorhabens entwickelten TCO-Rechners in einen Algorithmus, mit dem TCO automatisiert für beliebige reale Fahrzeugeinsatzprofile in der Datenbank berechnet werden können.
- Entwicklung einer Metrik zur Selektion für O-Lkw geeigneter Fahrprofile aus den Fahrzeugnutzungsdaten anhand der bisher im Projekt erarbeiteten Eignungskriterien.
- Untersuchung der Konzentration für O-Lkw geeigneter Verkehre auf einzelne Nutzer bzw. Nutzergruppen. Ziel war es dabei, mögliche Ankernutzer in einer Einführungsphase näher einzugrenzen<sup>1</sup>.
- Identifikation von kurzen bis mittleren Strecken (bis etwa 100 km), auf denen ein besonders hoher Anteil von Pendelverkehren zu verzeichnen ist. Solche Strecken eignen sich nach Erkenntnissen aus dem Basisprojekt besonders gut als Einführungsstrecken zu Beginn einer Systemeinführung von O-Lkw.
- Abschätzung des Einflusses internationaler Verflechtung auf die Marktpotentiale von O-Lkw. Konkret sollten dabei folgende Fragen untersucht werden:
  - Welche Unterschiede gibt es bei den Fahrzeugnutzungsmustern sowie beim Anteil gebietsfremder Fahrzeuge zwischen verschiedenen Nachbarländern Deutschlands? Welchen Einfluss hat dies auf das O-Lkw-Potential in diesen Ländern?
  - Für welche internationalen Verkehre wäre selbst bei einem rein nationalen Oberleitungsausbau eine Umstellung auf O-Lkw wirtschaftlich sinnvoll?
  - Wie stark könnte Deutschland vom Ausbau eines Oberleitungsnetzes in den Nachbarländern profitieren (oder auch umgekehrt)?
  - Welche Auswirkungen hätte eine selektive Elektrifizierung von ausländischen Hauptkorridoren? Welche Verkehre könnten sinnvoll elektrifiziert werden, wenn europaweit auf sämtlichen TEN-Korridoren Oberleitungen errichtet würden (z.B. im Rahmen einer Erweiterung der Infrastrukturrichtlinie AFID)?

---

<sup>1</sup> Eine solche Untersuchung wurde im Vorhaben sonst nur beispielhaft für einige exemplarische Strecken auf Grundlage von Daten zu Logistikimmobilien durch die Fraunhofer-Arbeitsgruppe SCS durchgeführt.

Durch die Analysen sollten aus den zugrundeliegenden Lkw-Einsatzprofilen für das europäische Gebiet (mit Schwerpunkt auf Deutschland und seinen Nachbarländern) praxisnahe Erkenntnisse für die Potentiale eines O-Lkw-Systems gewonnen werden.

Damit die Fahrzeugnutzungsdaten eines einzelnen Fahrzeugherstellers für die hier vorgesehenen Zwecke genutzt werden können, war eine Normalisierung der Daten und Skalierung an den Gesamtverkehrsmengen vorgesehen, um regions- oder anwendungsspezifische Unterschiede im Marktanteil des Herstellers auszugleichen.

## **AP 5.2: Abgleich des Netzausbaumodells mit realen Lkw-Einsatzdaten**

Im Gesamtvorhaben wurde ein mathematisches Modell zur Berechnung des Markthochlaufs von O-Lkw und geobasierter Optimierung des Infrastrukturaufbaus entwickelt (AP 1.4 sowie 2.4). Das Optimierungsmodell beruht auf relationsspezifischen Verkehrsmengen schwerer Nutzfahrzeuge aus dem Verkehrsmodell PTV Validate. Aus diesen Daten werden im Optimierungsmodell generalisierte Lkw-Einsatzprofile erstellt, jedoch unter Annahme von Pendelbeziehungen auf Basis von Gütermengenbewegungen auf den einzelnen Relationen zwischen den verkehrsverursachenden Gebieten.

Dieser Schritt der Erstellung von Fahrzeugeinsatzprofilen aus den Relationsdaten, der erhebliche Auswirkungen auf die Ergebnisse des Modells hat, sollte mithilfe der realen Fahrzeugnutzungsdaten verfeinert werden. Dabei sollten relationsspezifisch wesentliche Kennwerte aus den Nutzungsdaten in das Modell integriert werden. Im Einzelnen waren folgende Arbeitsschritte geplant:

- Ermittlung eines realitätsnahen Anteils von Pendelverkehren pro Relation und Einspeisung ins Optimierungsmodell
- Grobe Einordnung der Verkehre, die keine Pendelverkehre sind, hinsichtlich ihrer Eignung für O-Lkw. Daraufhin sollte entschieden werden, ob und ggf. wie diese Verkehre im Optimierungsmodell berücksichtigt werden können.
- Ermittlung von typischen realen Kraftstoffverbräuchen aus den Nutzungsdaten pro Streckenabschnitt auf dem deutschen Autobahnnetz und Einspeisung ins Optimierungsmodell
- Rechenläufe mit dem Optimierungsmodell auf der neuen Datengrundlage und Vergleich mit bisherigen Ergebnissen

Als Ergebnis sollte die Belastbarkeit der mithilfe des Optimierungsmodells getroffenen Aussagen gesteigert werden.

## AP 5.3: Schlussfolgerungen hinsichtlich des Roadmappings

Dieses Arbeitspaket sollte die Verbindung zu den Ergebnissen des Basisvorhabens sicherstellen. Im Kern ging es darum, die Implikationen der Analysen für staatliches Handeln in Deutschland und damit für das nationale Roadmapping herauszuarbeiten. Das sollte insbesondere die folgenden Punkte betreffen:

- Abschätzung der Marktpotentiale für O-Lkw
- Spezifischer Zuschnitt von Förderinstrumenten auf Einsatzprofile (z.B. gezielte Förderung von Pendelverkehren mit O-Lkw in der Einführungsphase)
- Auswahl vordringlicher Ausbaukorridore
- Notwendigkeit internationaler Koordination beim Oberleitungsausbau

## 3 Praktische Durchführung und Herausforderungen im AP 5

---

Um die Datenmengen des Flottendatensatzes bei Scania zu verarbeiten, kam ein dort bereits eingeführtes Framework zum Einsatz, das die verteilte Verarbeitung großer Datensätze über Computer-Cluster hinweg mit einfachen Programmiermodellen ermöglicht. Dieses Cluster basiert auf Apache-Hadoop und ist so konzipiert, dass von einzelnen Servern bis zu Tausenden von Rechnern skaliert werden können, wobei jeder Rechner lokale Berechnung und Speicherung ermöglicht.

Der Zugriff auf die Daten und die Analyse erfolgt über Apache Spark™, eine vereinheitlichte Analyse-Engine für die Verarbeitung großer Datenmengen, wahlweise über die Programmiersprachen Java, Scala, Python, R, sowie SQL oder wird über HIVE mit HiveQL bewerkstelligt. Als Programmierumgebung diente hauptsächlich Jupyter-Notebook, eine Open-Source-Webanwendung, mit der Dokumente erstellt und ausgetauscht werden können, die Live-Code, Gleichungen, Visualisierungen und narrativen Text enthalten.

Die Umsetzung der Algorithmen sowie die Programmierung der Datenabfragen und Analysen erfolgte in SQL und Python. Während die Verwendung dieser Sprachen keine Herausforderung darstellte, war eine Einarbeitung in das vorliegende Datenmodell sowie in die Prozessbedingungen und -abläufe des vor Ort implementierten Hadoop-Spark-Softwarestack zur Datenverarbeitung nötig.

Das anfängliche Fehlen einer zentralen Dokumentation des zur Verfügung gestellten Datensatzes bedingte häufigere Rückfragen zur Datenstruktur und Bedeutung von Variablenbezeichnungen sowie zu deren Entstehung und Hintergründe. Dies war notwendig, um deren richtige Verwendung und die Aussagefähigkeit der darauf basierenden Analysen einschätzen zu können. Das Analystenteam Scantias konnte hier im engen Austausch Hilfestellung und entsprechende Informationen geben.

Aufgrund von Weiterentwicklungen der hauseigenen Analysen und Verbesserungen an der Datenbasis waren auch auf ifeu-Seite häufige Änderungen notwendig. Dies bedingte anschließend häufig auch eine teilweise Anpassung der bisher entwickelten Codebasis, um die Analysen und Modellierungen weiterführen zu können. Hierdurch entstand ein erheblicher und unvorhergesehener Mehraufwand im Projekt.

Anfänglich bereitete auch die Einarbeitung in die zur Verfügung gestellte Arbeitsumgebung und deren Einrichtung für die vorgenommenen Tätigkeiten zusätzlichen Aufwand. Herausforderungen waren der Umgang mit Rechten zur Verwendung und Verfügbarkeit von Programmierbibliotheken in Python und die Sondierung der verschiedenen Möglichkeiten und Einschränkung der zur Verfügung gestellten Varianten zum Datenzugriff und zur Datenverarbeitung.

Da das Rechencluster von mehreren Analysten in verschiedenen Teams verwendet wurde, musste immer wieder mit Verzögerungen in der Ausführung des Codes gerechnet werden. Aufgrund einer Auslastungsübersicht und Rangfolgendarstellung war dies in Teilen planbar, musste aber im Arbeitsablauf berücksichtigt werden. So wurden einerseits Testläufe mit möglichst geringen Datenmengen durchgeführt und andererseits wurde darauf geachtet, die Abfragen und Analysen möglichst recheneffizient zu konzipieren, um sowohl den eigenen Arbeitsablauf zu optimieren, als auch möglichst wenig Rechenkapazität und -zeit hinsichtlich der anderen gleichberechtigten Arbeiten in Anspruch zu nehmen.

Hier war es sinnvoll weitestgehend parallelisierten Code zu programmieren, um die Berechnungszeiten aufgrund der Datensatzgrößen gering zu halten. Rechenabbrüche sowie stunden- bis tagelange Wartezeiten sollten damit vermieden werden.

Zur Nutzung der Daten musste zu Beginn der Kooperation eine strikte Geheimhaltungserklärung unterzeichnet und Sicherheitsschulung durchgeführt werden. Ferner besteht weiterhin die Verpflichtung vor Veröffentlichung und Verwendung von (Teil-)Ergebnissen außerhalb der Arbeitsumgebung Scantias eine Freigabe einzuholen. Die Diskussionsmöglichkeiten rund um die Modellbildung im ifeu-Projektteam waren dadurch eingeschränkt.

Im Rahmen der Zusammenarbeit konnten jedoch wichtige Ergebnisse und Erkenntnisse in einem sehr kollegialen und kooperativen Umfeld erarbeitet werden, die auch teilweise in den anderen APs des Gesamtvorhabens Berücksichtigung finden konnten. Zusätzlich wurde auch eine Methodik entwickelt, die nun in dieser und ähnlichen Fragestellung unmittelbar einsetzbar ist.

# 4 AP 5.1: Analyse realer Lkw-Einsatzdaten

## 4.1 Aufbereitung des Datensatzes

Die Fahrzeugflotte Scantias stellt ein Subsample der gesamten Lkw im Lieferverkehr dar. Der Marktanteil Scantias an den zugelassenen Nutzfahrzeugen der Klasse über 3,5 t ist seit 2010 gestiegen und betrug 2017 knapp 13 % (siehe Abbildung 1). Neben Lkw-Modellen von Scania finden sich u.a. Fahrzeuge der Hersteller DAF, Iveco, MAN, Mercedes-Benz, Renault Trucks und Volvo in den Flotten der Spediteure.

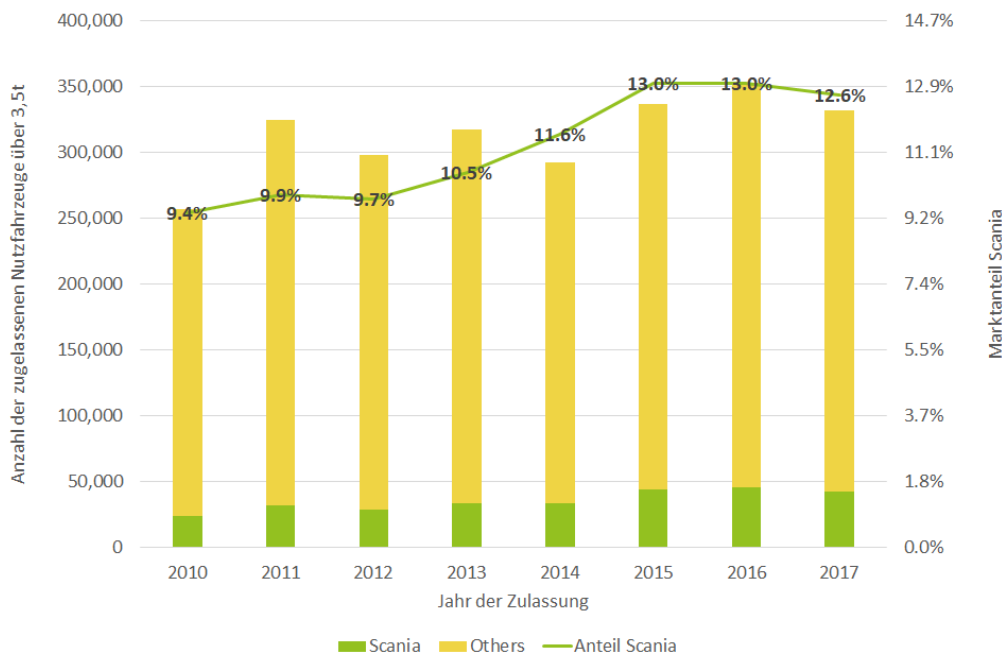


Abbildung 1: Marktanteile Scantias an den Nutzfahrzeugen der Klasse über 3,5t in den Jahren 2010 bis 2017 ohne Linien- und Reisebusse in der EU und den Ländern der Europäische Freihandelszone (EFTA). Eigene Darstellung basierend auf Daten der ("ACEA - European Automobile Manufacturers' Association" 2018)

Da die Floating-Vehicle-Daten damit zwar einen relevanten, aber dennoch nur begrenzten Anteil am Lkw-Gesamtmarkt darstellen, war es eingangs notwendig dieses Sample auf den Gesamtbestand und die damit einhergehenden Verkehre im Straßengüterverkehr zu skalieren. Die Basis der verfügbaren Daten erlaubt dann Rückschlüsse auf die Gesamtheit der Lkw-Verkehre.

### Methodische Herangehensweisen

Prinzipiell boten sich zwei methodische Herangehensweisen an:

- Die Verwendung von Marktdaten, die entweder als Gesamtzulassungszahlen, Marktanteilen des Herstellers oder Bestandsdaten in den betrachteten Regionen vorliegen.
- Die Verwendung von Verkehrszählungen oder -statistiken in den zu untersuchenden Gebieten.

Bei der **Verwendung von Marktdaten** ist es notwendig die Größenklassen der Nutzfahrzeuge zu berücksichtigen, um keine Verzerrungen zu erzeugen, da die Marktanteile zwischen den Gewichtsklassen der Nutzfahrzeuge variieren. So liegt beispielsweise der Marktanteil von Scania bei Lkw über 16 t zwischen 2010 und 2017 im Mittel bei knapp 15 % und somit höher als im Mittel der Lkw ab 3,5 t aufwärts (*THE SCANIA REPORT 2018 2019*). Ferner spielt das Fahrzeugalter eine Rolle, da dieses Einfluss auf den Verbrauch der Fahrzeuge aufgrund der fortschreitenden Effizienzsteigerung und deren regionale Nutzung (siehe Abbildung 2) hat. Während in Ländern wie Deutschland, Luxemburg und Dänemark ein Großteil der Fahrleistung im Transportsektor durch relativ neue Fahrzeuge erfolgt, sind in typischen Zweitmarktländern wie Griechenland, Zypern oder Tschechien die eingesetzten Fahrzeuge zu einem überwiegenden Teil älter als sechs Jahre.

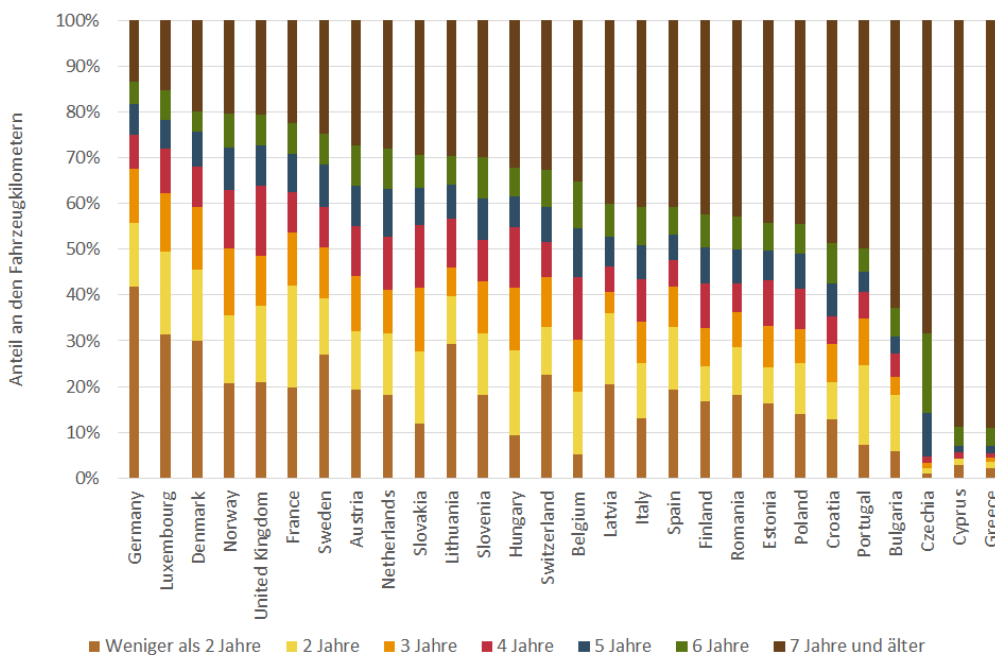


Abbildung 2: Anteil der Lkw-Alterklassen an den zurückgelegten Fahrzeugkilometern in ausgewählten europäischen Ländern im Jahr 2017. Eigene Darstellung basierend auf (EUROSTAT 2019a)

Bei der Verwendung von **Verkehrszählungen oder -statistiken** (siehe beispielsweise Abbildung 3) in den zu untersuchenden Gebieten lässt sich abschätzen, wie groß der Anteil am Transportverkehrsaufkommen ist, der durch das Sample repräsentiert wird.



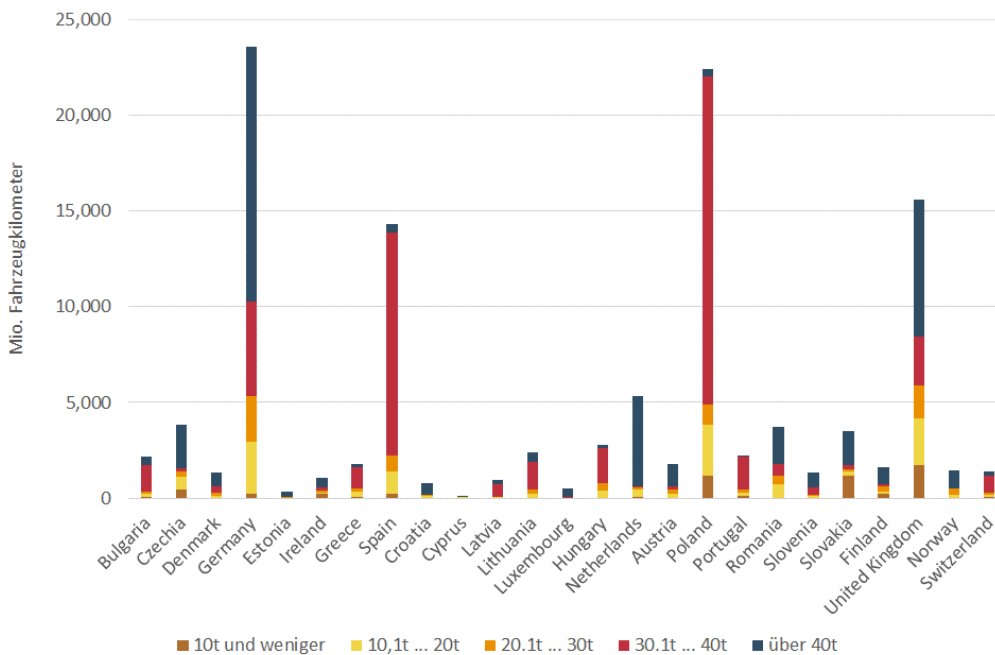


Abbildung 3: Fahrleistungen in Mio. Fahrzeugkilometern von Lkw (beladen) nach zulässigem Gesamtgewicht unterteilt in ausgewählten Ländern der EU im Jahr 2017. Eigene Darstellung basierend auf (EUROSTAT 2019a)

Verkehrszählungen lagen für die Regionen Europas in unterschiedlicher zeitlicher Auflösung vor und sind in ihrer größten Einteilung als nationale Fahrleistungs- oder Transportstatistiken für verschiedene Stichjahre verfügbar. Alternativ lagen auch Tagesmittelwerte im Jahresdurchschnitt (AADT – Average Annual Daily Traffic) für Straßenabschnitte vor oder waren feiner aufgelöst als Stundenwerte erhältlich. Die regionale Verfügbarkeit und Auflösung variierten hier jedoch ebenfalls. Während für die BRD eine relativ dichte Erfassung auf Bundesstraßen und Autobahnen existiert (BAST 2017) und frei verfügbar war, wie auch in einigen anderen EU-Ländern, lagen in anderen Ländern zum Teil nur Stichprobenzählungen und Hochrechnungen auf den Europastraßen vor (siehe ("Traffic Census 2015 - Transport - UNECE" 2015) und (Pettersson et al. 2018)).

### Vorgehensweise

Während die Marktdaten in Form von Zulassungs- und Bestandszahlen eine gute Basis zur Hochrechnung auf den Gesamtbestand darstellten, konnten dabei Ungenauigkeiten bezüglich der regionalen Fahrleistungen entstehen. Bei örtlich aufgelösten Verkehrsdaten wurde die Fahrleistungen auf einzelnen Straßenabschnitten besser repräsentiert, allerdings war hier der verfügbare Datenbestand bezüglich der örtlichen und zeitraumbezogenen Verfügbarkeit sehr heterogen.

Zur Anpassung des Datensatzes auf den Gesamtbestand bzw. auf die Fahrleistungen der Straßenabschnitte in den zu betrachtenden Ländern sollte schlussendlich ein Hochrechnungsfaktor für jedes Fahrzeug in der Datenbank unter Verwendung beider Ansätze bestimmt werden. Dies war ein sich über die gesamte Kooperation erstreckender Prozess, der maßgeblich von der Datenverfügbarkeit sowie der Übertragungs-

möglichkeit ins Datenmodell abhing. Insbesondere die Qualität (Vollständigkeit, zeitlicher Bezug, Übereinstimmung von statistischen Variablen und Parametern, etc.) der Quellen sowie die Erhöhung der Genauigkeit standen dabei im Vordergrund.

Die Hochrechnung auf Basis der Fahrleistung der Lkw in den einzelnen europäischen Ländern (siehe Abbildung 3) konnte noch im Projektverlauf durchgeführt werden. Die Daten zum Abgleich mit verfügbaren Verkehrszählwerten einzelner Länder befinden sich jedoch noch in der Aufbereitung durch Scania.

## 4.2 Analyse der internationalen Verkehre und deren Einfluss auf das Elektrifizierungspotenzial

In den bisherigen Arbeitspaketen lag der Betrachtungshorizont auf den inländischen Verkehren auf dem Gebiet der BRD. Um das zusätzliche Elektrifizierungspotenzial durch internationale Verkehre abschätzen zu können, wurde der Anteil relevanter internationaler Verkehre auf einzelnen Streckenabschnitten untersucht. Der Anteil internationaler Verkehre ist dabei in den Ländern Europas sehr unterschiedlich (siehe Abbildung 4): Für Deutschland liegt er unter Berücksichtigung unbeladener Fahrten und beladener Fahrten die im Land entweder be- oder entladen werden, bei knapp 10 % und ist damit relevant.

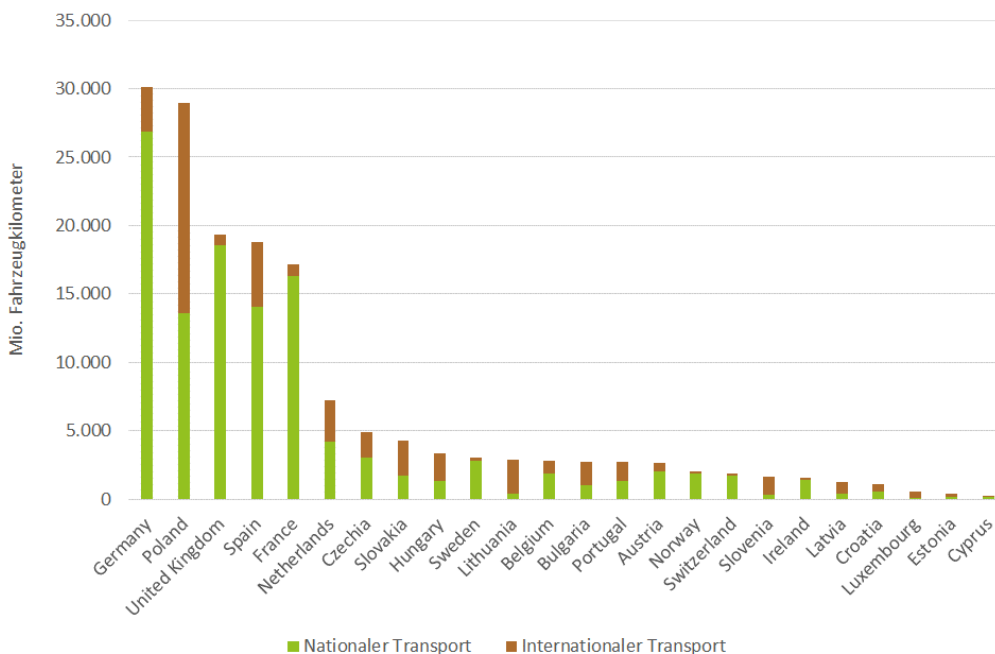


Abbildung 4: Fahrleistungen nationaler und internationaler Straßentransporte durch beladene und unbeladene Lkw in ausgewählten Ländern Europas im Jahr 2017. Eigene Darstellung basierend auf (EUROSTAT 2019a)

Im Falle Polens, ein direkter Nachbarstaat Bundesrepublik mit ebenfalls hohen Transportverkehrsleistungen, liegt der Anteil internationaler Verkehre dagegen mit etwa 50 % deutlich höher. Ebenso trägt Polen mit über 2,8 Mio. Fahrten durch die BRD im

Jahr 2017 den größten Anteil an den Transitfahrten durch das Bundesgebiet bei (siehe Abbildung 5), gefolgt von den Niederlanden mit über 900.000 Fahrten. Insgesamt be-  
laufen sich die Durchquerungen des Transitlandes Deutschland auf rund 6,9 Mio. Fahr-  
ten.

Die Auswirkungen der internationalen Transporte und der Transitverkehre auf das  
Elektrisierungspotenzial sollte auf Basis der hier beschriebenen Analyse untersucht  
werden.

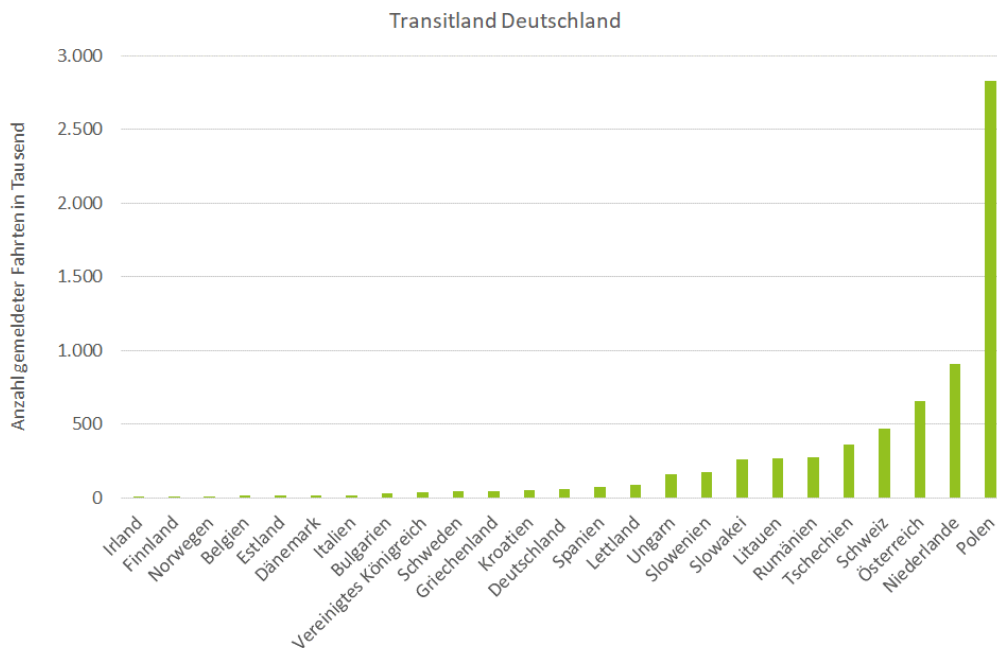


Abbildung 5: Transitfahrten in Tausend Fahrten von Lkw der Meldeländer durch Deutschland im Jahr 2017. Eigene Darstellung basierend auf (EUROSTAT 2019a)

Zur Durchführung der Analyse wurden fahrleistungsbezogene Vollkosten unter Berücksichtigung der unterschiedlichen Anschaffungspreise sowie der Differenz der Betriebskosten, z.B. dem Treibstoffverbrauch für jedes Fahrzeug an jedem untersuchten Streckenabschnitt, berechnet, abhängig davon, ob der Abschnitt elektrifiziert oder dieselbetrieben zurückgelegt werden kann. Hierbei wurden u.a. die Kostenannahmen der Modellrechnungen der Arbeiten im Basisprojekt zugrunde gelegt und die Werte zwischen den Modellen wurden abgeglichen.

Anschließend wurden die ermittelten Kosten mit denen rein dieselbetriebener Fahrzeuge, also ohne hybridisierten Antriebstrang, abgeglichen. Bei Kostenneutralität oder bei einem Kostenvorteil wurde der Lkw als potenzieller Oberleitungs-Lkw gesetzt. Als weitere Sensitivität wurden CO<sub>2</sub>-Vermeidungskosten von rund 0,25 und 0,5 € pro Liter Diesel in den Berechnungen eingepreist, um die Effekte eines CO<sub>2</sub>-Preises auf die Kosten und damit auf mögliche Umstellungspotenziale auf den Oberleitungsbetrieb zu analysieren. Eine Kopplung mit einem Infrastrukturausbaumodell konnte im Projektverlauf nicht mehr durchgeführt werden. Diese wurde daher im jeweiligen Betrachtungsgebiet auf den Strecken des Europäischen Fernstraßennetzes je nach Berechnungsschritt als gesetzt oder nicht vorhanden betrachtet.

- Im ersten Berechnungsschritt wurden die Umstellungspotenziale nur für Fahrzeuge berechnet, die sich innerhalb des betrachteten Staatsgebietes befinden, um diese Menge vom zusätzlichen Potenzial durch Einbeziehung der internationalen Verkehre abgrenzen zu können.
- Der nachfolgende Rechengang bezog die grenzüberschreitenden Verkehre mit ein, wobei die Straßenabschnitte der umliegenden Länder nicht elektrifiziert sind, um die Anzahl der Lkw zu ermitteln, die aufgrund genügend großem Anteil ihrer Fahrleistung unter Oberleitungsinfrastruktur im Betrachtungsland wirtschaftlich betrieben werden könnten.
- Zur Bestimmung des Gesamtpotenzials und des Einflusses der Strecken elektrifizierung der europäischen Mitgliedstaaten wurden in weiteren Rechenläufen zuerst die vorgesehenen Straßenabschnitte einzelner Länder umgestellt und zuletzt das Potenzial für oberleitungsbetriebene Lkw bestimmt, wenn alle europäischen Staaten ihre Haupttransportverkehrsstraßen für einen Oberleitungsbetrieb rüsten.

Bezüglich der Betriebskosten wurde vereinfachend angenommen, dass der Treibstoff in dem Land getankt wird, in dem er verbraucht wurde. Abbildung 6 zeigt exemplarisch die zugrunde gelegten Kosten für Strom und Diesel der einzelnen Länder für das Bezugsjahr 2018. Dies impliziert die Annahme, dass das Fahrzeug statistisch dort häufiger tankt, wo es am meisten betrieben wird, was aufgrund unterschiedlicher Dieselpreise in den verschiedenen Ländern nicht zwingend gegeben ist. Eine Analyse zur genaueren Verortung der Tankvorgänge wäre prinzipiell mithilfe des Datensatzes denkbar, konnte aber nicht mehr durchgeführt werden.

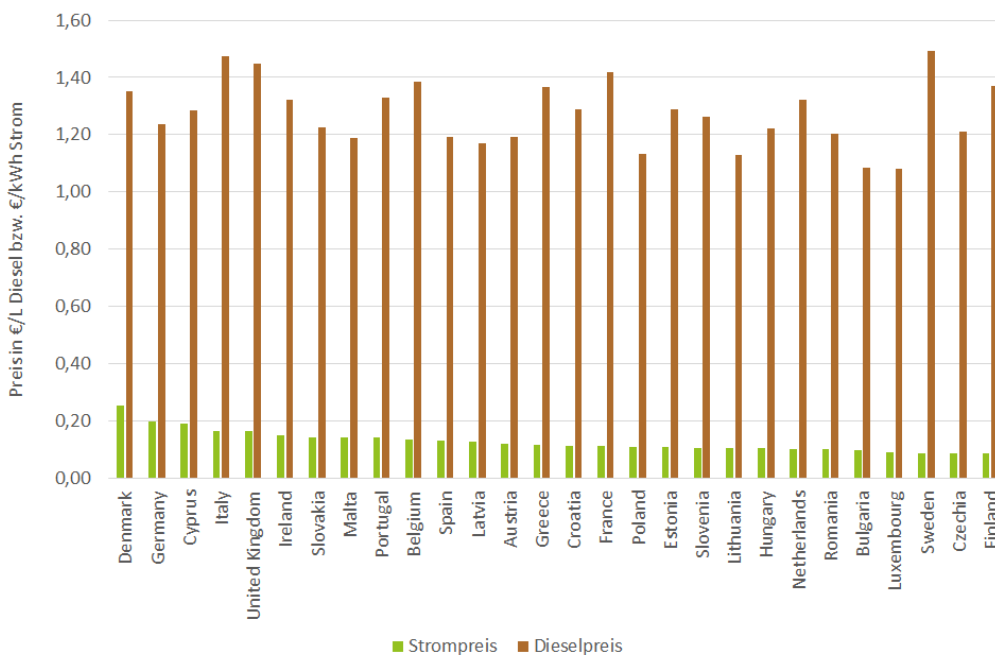


Abbildung 6: Mittlere Preise für Diesel und Strom (Gewerbekunden) inklusive Steuern und Angaben in ausgewählten europäischen Ländern im Jahr 2018. Eigene Darstellung basierend auf (European Environment Agency 2019) und (EUROSTAT 2019b)

Über die zusätzlichen Potenziale für elektrifizierten Lkw-Betrieb unter der Oberleitung durch Berücksichtigung inter- und transnationaler Verkehre können zusätzliche oberleitungsgebundene Fahr- oder Transportleistung im Straßengüterverkehr auch durch Strecken elektrifizierung in weiteren Ländern generiert werden. Der jeweilige Beitrag

wurde in einer Verflechtungsmatrix zwischen den Ländern dargestellt (siehe exemplarisches Beispiel in Abbildung 7). Leider können hier die bereits erzielten Ergebnisse nicht dargestellt werden, da aufgrund des Endes der Kooperation bisher noch keine Autorisierung seitens Scania vorliegt. Eine Validierung der Ergebnisse und Prüfung des Rechenmodells und seiner Eingangsparameter waren zudem noch ausstehend. Ein Abgleich der Umstellungspotenziale zwischen den Modellergebnissen mit den Ergebnissen aus dem Potenzialnetz 2030 für Deutschland erfolgte bisher ebenfalls nicht.

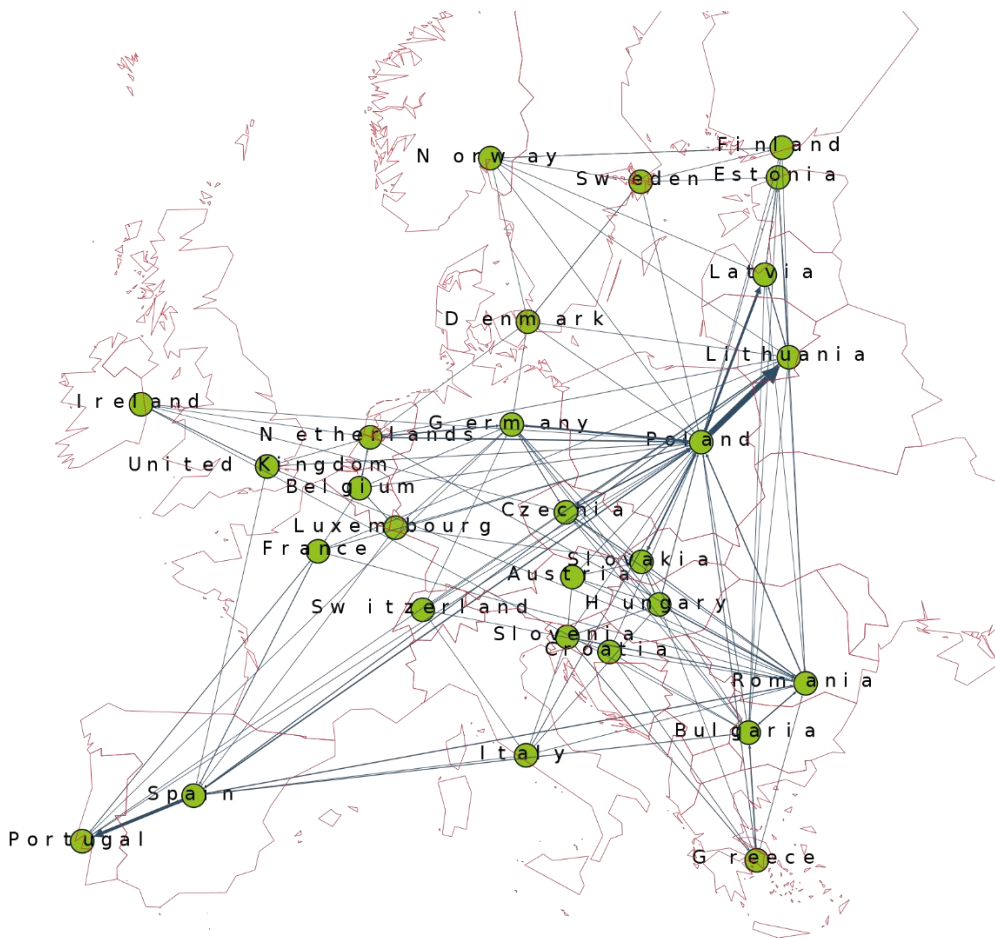


Abbildung 7: Zusätzliches Potenzial an elektrifizierter Fahrleistung aufgrund des Infrastrukturausbaus in den entsprechenden Nachbarländern und der Umstellung zusätzlicher Lkw der internationalen und Transitverkehre auf den Oberleitungsbetrieb. Exemplarische Darstellung mit Ersatzwerten, da die bisherigen Zwischenergebnisse nicht zugänglich sind.

## 4.3 Entwicklung einer Methodik zu Identifizierung von Pendelverkehrsanteilen

Diese Arbeiten hatten das Ziel, oberleitungsg geeignete Fahrprofile von einzelnen Nutzern zu identifizieren. Mit *Fahrprofilen* sind hier Regelmäßigkeiten der gefahrenen Routen gemeint - zum Beispiel

- ob ein Nutzer zwischen zwei Standorten hin und her pendelt,
- ob er gewisse Strecken sehr häufig befährt, andere aber kaum,
- ob er zwischen mehreren Orten im Kreis fährt oder
- ob er ohne erkennbare Regelmäßigkeiten unterwegs ist.

Fokus war die Unterscheidung in Pendler und nicht Pendler, da für Pendler ein besonders hohes Elektrifizierungspotenzial gesehen wird. Die erarbeiteten Methoden können aber auch verwendet werden, um Nutzer in andere Gruppen einzuordnen. Die Klassifikation basiert auf Fahrzeugdaten sowie den „Floating-Car-Data“.

Die Methodik hat zahlreiche Anwendungsmöglichkeiten. Beispielsweise leistet die Identifizierung von Pendelverkehren - gerade in Phasen des Markthochlaufs - einen wichtigen Beitrag zur Bestimmung der Elektrifizierungspotentiale von Fuhrparks und Fahrzeugflotten. Da insbesondere in der Frühphase die Verfügbarkeit der Infrastruktur limitiert ist, eignen sich Pendelverkehre besonders gut zur Umstellung auf neue Technologien. Sind die Strecken auf denen zahlreiche Pendler verkehren bekannt, kann die Infrastruktur effizient aufgebaut werden - sei es eine Ladeinfrastruktur für Batterie-Fahrzeuge oder Oberleitungen für oberleitungsgebundene Lkw.

Da die Nutzung und Darstellung von Scania-Daten nicht mehr möglich ist, und diese Daten mittlerweile auch nicht mehr zugänglich sind, wird die Methodik am Beispiel von Pkw-Daten aus dem My-eDrive-Projekt des ifeu entwickelt und dargestellt. Aufgrund der ähnlichen Datenstruktur kann die Methodik dann mit geringem Aufwand auf die Fahrdaten von Scania übertragen werden.

Es handelt sich bei den My-eDrive-Daten um 36.000 aufgezeichnete Fahrten (*Tracks*), wobei jeder Track einem User zugeordnet werden kann und aus Datenpunkten zu verschiedenen Zeitpunkten mit u.a. der genauen Position des Fahrzeugs, Uhrzeit, Geschwindigkeit besteht. Dieser Datensatz wurde in ein möglichst ähnliches Datenmodell, wie es bei den Analysen der Lkw-Daten von Scania eingesetzt wurde, überführt.

Hierzu wurden die Datensätze von GPS-Trajektorien in ein Gitterzellenmodell transferiert, die Datenpunkte der GPS-Koordinaten werden damit eindeutigen geographischen Zellen (Grid) zugeordnet. Treten mehrere direkt aufeinanderfolgende Datenpunkte eines Tracks in einer Gitterzelle auf, werden deren Informationen (z.B. Zeitpunkt, Geschwindigkeit, Beschleunigung, Verbrauch und Lenkwinkel) zu einem Mittelwert zusammengefasst. Bereits gewonnene Erkenntnisse flossen in die Ausarbeitung des Algorithmus und die weiteren Untersuchungen zur Performance des Algorithmus ein.

Der Datensatz wurde hierzu nach Nutzern gefiltert, um Aussagen über die Routenwahlfolge sowie über die Eigenschaften dieser Routen zu treffen. Dazu wurden *Features* definiert, also Merkmale, die das Verhalten eines Nutzers in einer Zahl quantifizieren<sup>1</sup>. Solche Merkmale sind zum Beispiel:

- Wie viele Tracks hat ein Nutzer aufgezeichnet?
- Wie viele Kilometer war ein Nutzer unterwegs?
- Wie groß ist das Gebiet in dem ein Nutzer
- Wie oft werden bestimmte Routenabschnitte durch den Nutzer durchfahren

Statt alle Tracks in Form einer Koordinatenfolge zu betrachten, wird dann jeder Nutzer für die folgende Analyse nur noch durch eine Reihe von Features und deren numerischen Wert beschrieben. Anhand dieser Features sollte dann ermittelt werden, ob der Nutzer bestimmte Strecken im Pendelverkehr fährt. Daraus ergaben sich zwei wesentliche Arbeitsschritte:

1. Definition geeigneter Features, die Pendler von Nicht-Pendlern unterscheiden.
2. Gruppierung der Nutzer, die hier Fahrzeuge repräsentieren, anhand dieser Features.

#### 4.3.1 Definition und Beschreibung der untersuchten Features

Nachfolgend werden die untersuchten Features dargestellt, die sich in vier Kategorien untergliedern lassen.

##### Häufigkeits-Features

Nutzer bzw. Fahrzeugeinsätze können dadurch charakterisiert werden, wie häufig sie bestimmte Orte besuchen. Um möglichen Ungenauigkeiten der GPS-Positionen entgegenzuwirken, wird ein Gitter aus Rechtecken (Raster) über die Karte gelegt. Anstelle der genauen Position wird betrachtet, in welcher Gitterzelle sich ein Nutzer gerade befindet. Aufeinanderfolgende Datenpunkte, bei denen sich der Nutzer innerhalb der gleichen Zelle befindet, werden zusammengefasst. Aufenthalte innerhalb einer Zelle werden also nur mehrfach gezählt, wenn der Nutzer zwischendurch eine andere Zelle besucht hat. Durch Repräsentation der Fahrten in einem solchen Raster lassen sich folgende Features ableiten:

- Durchschnittliche Häufigkeit der Durchfahrung der Zellen (Summe der Häufigkeiten aller Zellen dividiert durch Gesamtanzahl der besuchten Zellen; *num\_visits*)
- Median der Häufigkeiten für jede Zelle (robuster gegen Ausreißer als der Durchschnitt; *num\_visits\_med*)

---

<sup>1</sup> Da Literatur und Code-Dokumentationen alle auf Englisch sind, werden hier für manche Begriffe die englischen Worte verwendet:

- Track: eine aufgezeichnete Fahrt eines Nutzers
- Feature: Merkmal, welches das Nutzerverhalten in einer Zahl zusammenfasst
- Cluster: Gruppe von Nutzern, ermittelt anhand von Features
- Clustering: Gruppierung (also Gesamtmenge der Cluster) von Nutzern anhand von Features
- Dataframe: Datenstruktur in Python und anderen Programmiersprachen, im Prinzip eine Tabelle

- Varianz der Häufigkeiten für jede Zelle (*var\_counts*)

Es wird im Falle von Pendelstrecken ein relativ hoher Median der Häufigkeiten erwartet, da viele Zellen mehrfach besucht werden. Die Varianzen der Häufigkeiten sollten für Pendler dagegen eher niedriger liegen als für Nicht-Pendler.

### Ähnlichkeits-Features

Grundannahme ist die Ähnlichkeit der Tracks für Nutzer im Pendelverkehr. Der Begriff *ähnlich* wird hier definiert wie in (Mariescu-Istodor / Fränti 2017): Für zwei Tracks wird die Anzahl der Zellen berechnet, welche von beiden Tracks besucht werden und diese wird mit der Anzahl der in mindestens einem der beiden Tracks besuchten Zellen normiert. Das Ergebnis wird *Similarity-Score* genannt. Dieser ist 1, wenn die Tracks identisch sind und 0, wenn sie keine Zelle gemeinsam haben.

Da Pendler viele identische oder zumindest sehr ähnliche Tracks haben, wird vermutet, dass Pendlern hier hohe Werte aufweisen.

### Geometrische Features

Dabei wird die Geometrie der besuchten Gitterzellen analysiert indem die konvexe Hülle um die Positionspunkte berechnet wird. Es wird vermutet, dass Pendler auf einem kleineren Gebiet unterwegs sind und womöglich die konvexe Hülle eher kleinere Flächen oder geringere Komplexitäten aufweisen als bei Nutzern, deren Routen unregelmäßiger sind.

### Graphen-Features

Pendler können auch mithilfe von Methoden aus der Graphentheorie identifiziert werden. Dafür wird zunächst ein gerichteter Multigraph aus den besuchten Orten des Nutzers erstellt mit Knoten für Start- und Endpunkte jedes Tracks und Orte, an denen der Nutzer eine gewisse Zeit lang gestanden ist.

Zeitlich aufeinanderfolgende Knoten werden dann jeweils mit einer Kante verbunden. Damit wird ein gerichteter Graph für den Nutzer generiert. Es wird erwartet, dass dieser Graph für Pendler andere Strukturen hat als für Nicht-Pendler. Beispielsweise werden bei Pendlern vermutlich einige Kanten recht häufig vorkommen oder viele Kreise auftreten (diese entstehen durch Hin- und Herfahren).

### Winkel-Features

Schlussendlich können auch Lenkwinkel analysiert werden. Diese werden jeweils aus drei aufeinanderfolgenden Datenpunkten berechnet, dessen Varianz, als Maß dafür wie stark und häufig der Fahrer seine Richtung ändert, betrachtet werden kann.



### 4.3.2 Gruppierung mittels Clustering-Verfahren

Im vorigen Abschnitt wurden zahlreiche mögliche Features vorgestellt, mit denen sich aus großen Datensätzen an Fahrten potenziell Nutzer im Pendelverkehr identifizieren lassen. Ein exemplarisches Ergebnis dieser Feature-Zuweisung ist in Abbildung 8 dargestellt: Ergebnis ist eine Tabelle mit einer Zeile pro Nutzer und den jeweiligen Features als Spalten.

	id_user	num_visits_med	sim_score	ch_area_norm	mfe_un
0	10161	1.00000	0.32411	0.58387	13.00000
1	11998	12.00000	1.00000	1.15003	12.00000
2	11268	1.00000	0.35334	13.48194	19.00000
3	11246	1.00000	0.12511	1.55784	12.00000
4	10158	1.00000	0.29910	1.28527	14.00000
5	12404	3.00000	0.64708	0.68670	12.00000
6	11348	13.00000	0.82689	1.22492	12.00000
7	11208	2.00000	0.29550	0.50249	13.00000
8	11193	12.00000	0.86911	0.77255	12.00000
9	10306	2.00000	0.73755	1.12506	12.00000
10	11758	12.00000	0.88612	1.37641	12.00000

Abbildung 8: Auszug einer Tabelle mit Nutzern und einem zugeordneten Featureset.

Anhand des Featuresets wurde dann untersucht, ob die Nutzer anhand der Differenz zwischen den Werten aller Features zugleich gruppieren lassen. Liegen die Featurewerte nahe beieinander, werden diese Nutzer einer Gruppe (hier Pendler oder Nicht-Pendler) zugeordnet.

#### Algorithmische Umsetzung

Für die genaue Aufteilung in verschiedene Gruppen stehen verschiedene Algorithmen zur Verfügung. Jeder dieser Algorithmen gibt als Ergebnis eine Gruppierung der Datenpunkte bzw. Nutzer, ein *Clustering*, aus. Bei einer solchen Gruppierung wird jedem Datenpunkt ein *Label* zugewiesen, welches die Clusterzugehörigkeit des Datenpunktes angibt. Hierzu können für zukünftige Untersuchungen u.a. folgende Algorithmen angewendet werden:

- Der **KMeans**-Algorithmus bildet Cluster, indem er die Summe der quadrierten Abstände von den Cluster-Zentren minimiert. Als Parameter wird die Anzahl der Cluster verwendet.
  - Vorteile: Schnell, leistungsfähig, einfach zu verstehen und benötigt nur einen Parameter
  - Nachteile: Cluster müssen konvexe Form haben; Ausreißer beeinflussen Ergebnis stark
- Der **DBSCAN**-Algorithmus trennt Bereiche hoher Dichte von Bereichen mit niedrigerer Dichte. Er arbeitet mit dem Begriff eines „core samples“, dies sind Punkte in Bereichen mit hoher Dichte. Cluster sind Ansammlungen solcher „core samples“. Der Parameter ist hier der maximale Abstand zwischen zwei Samples (Nutzern), und die Mindestanzahl an Samples innerhalb dieses maximalen Abstandes untereinander, um ein gültiges Cluster zu bilden.
  - Vorteile: Cluster können beliebige Form haben, Algorithmus erkennt Ausreißer
  - Nachteile: Bei der Gruppierung mit Clustern verschiedener Dichte

### 4.3.3 Bewertung der Clusteringverfahren

Zur Evaluation der Eignung der beiden Clusteringmethoden gibt es zwei grundsätzliche Ansätze:

- **Supervised Setting:** Kennt man die wahre Gruppenzugehörigkeit der Datenpunkte, so kann das Ergebnis des Algorithmus mit der wahren Gruppierung verglichen werden. Die wahre Gruppierung kann beispielsweise durch händische Betrachtung und Einordnung der einzelnen Datenpunkte erstellt werden. Da dies mühsame Arbeit und nur für kleine Datensätze machbar ist, ist dieses Verfahren in der Praxis oft nicht umsetzbar.
- **Unsupervised Setting:** Hier kennt man die wahre Gruppenzugehörigkeit nicht. Gruppierungen können trotzdem beurteilt werden, zum Beispiel abhängig davon, wie nah Punkte des gleichen Clusters beieinanderliegen und wie weit Punkte unterschiedlicher Cluster auseinanderliegen oder wie viel Überschneidung die Cluster im Raum haben.

Für beide Ansätze gibt es jeweils eine Reihe von Metriken. Mithilfe dieser werden verschiedene Kombinationen aus der großen Menge an möglichen Features einem den beiden Clustering-Verfahren zugeführt und die Ergebnisse bewertet, einerseits hinsichtlich der Eignung der Features zur Gruppierung der Daten und andererseits hinsichtlich der Rechenperformance. Für jeden Clustering-Algorithmus wurde dabei wie folgt vorgegangen:

1. Berechnung aller Features für jeden Nutzer
2. Für verschiedene mögliche Kombinationen aus Features und für verschiedene Parameter der Clustering-Algorithmen:
  - a. Durchführung des Clusterings
  - b. Bewertung des Clusterings
3. Vergleich der Clusterings und dadurch dann Auswahl geeigneter Features

Zur Vereinfachung der Bewertung der Zuordnungsperformance und visuellen Zuordnung der Tracks eines Nutzers wurden Häufigkeits-Plots erstellt (siehe Beispiel in Abbildung 9). Dabei werden alle durchfahrenen Gitterzellen eines Nutzers gezeigt und mittels Farbgebung die Häufigkeit der Durchfahrung einer Zelle bzw. eines Straßenabschnitts angezeigt.

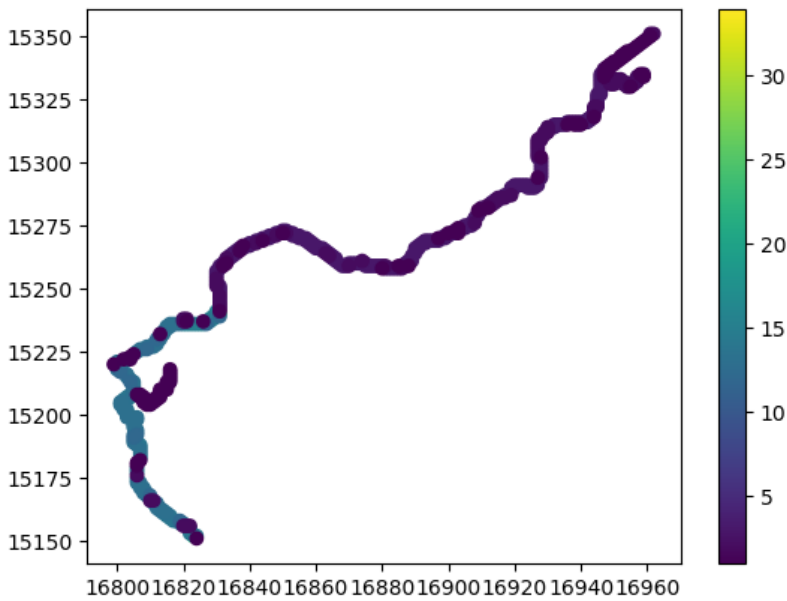


Abbildung 9: Exemplarische Darstellung der Durchfahungshäufigkeit einzelner Abschnitte durch einen Nutzer

Als weitere Visualisierungsmöglichkeit wurden Graphen-Plots erstellt, die die Anzahl der Befahrungen an allen Kanten antragen (siehe Beispiel in Abbildung 10).

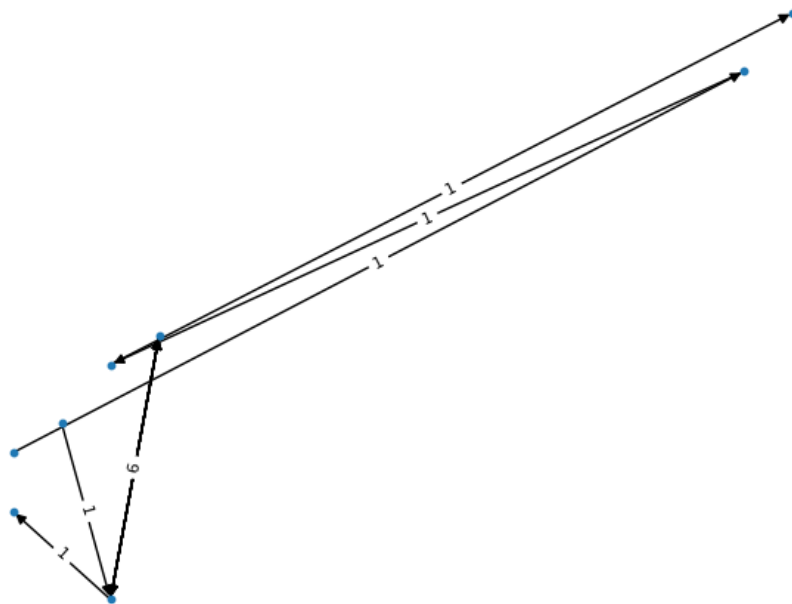


Abbildung 10: Exemplarischer Graphenplot eines Nutzers, der die Anzahl der Durchfahrungen von Routen (Kanten) darstellt.

Für die Bewertung der Clusterings wurden mithilfe der beschriebenen Visualisierungen manuelle Labels für jeden Nutzer eines reduzierten Datensatzes zugewiesen. Dabei wurde nicht nur zwischen Pendlern und Nicht-Pendlern, sondern zwischen mehrere Gruppen unterschieden, wie „eindeutige Pendler mit vielen Tracks“, „Nutzer, die auf Teilen der Strecke pendeln“ und „Nutzer mit Pendel-Tendenzen, aber wenigen Tracks“.

Da für diesen Datensatz damit eine Einteilung auf Basis manueller Labels zur Verfügung stand, konnten sowohl Methoden aus dem *Supervised Setting* als auch aus dem *Unsupervised Setting* zur Bewertung der Algorithmen und der verwendeten Labels eingesetzt werden:

- Adjusted Rand Score, supervised (Hubert / Arabie 1985)
- V Measure, supervised (Rosenberg / Hirschberg 2007)
- Silhouette Score, unsupervised (Rousseeuw 1987)
- Davies-Bouldin Index, unsupervised (D. L. Davies / D. W. Bouldin 1979)

All diese Metriken wurden so angepasst, dass 1 der bestmögliche Wert ist und bedeutet, dass die vom Algorithmus zugeordneten Labels genau mit den manuell erstellten Labels übereinstimmen. Im *unsupervised setting* deutet ein Wert nahe bei 1 auf sehr dichte, deutlich voneinander separierte Cluster hin. Die Scores wurden summiert und die Clusterings mit ihren unterschiedlichen Featuresets nach ihrer Gesamtbewertung sortiert (siehe Beispiel in Abbildung 11).

	n_features	features	n_clusters	sum_both
2360	3	['mfe_un', 'num_visits_med', 'var_counts']	5	3.56356
1544	3	['mfe_un', 'num_visits_med', 'var_counts']	4	3.52852
1728	3	['ch_area_norm', 'mfe_un', 'num_visits_med']	5	3.50583
9160	4	['ch_area_norm', 'mfe_un', 'num_visits_med', 'var_counts']	5	3.49313
912	3	['ch_area_norm', 'mfe_un', 'num_visits_med']	4	3.47129
9142	4	['ch_area_norm', 'mfe_un', 'n_2cycles', 'var_counts']	5	3.46110
1526	3	['mfe_un', 'n_2cycles', 'var_counts']	4	3.45614
1231	3	['cycle_per_node', 'mfe_un', 'var_counts']	4	3.45339
1724	3	['ch_area_norm', 'mfe_un', 'n_2cycles']	5	3.44416
8847	4	['ch_area_norm', 'cycle_per_node', 'mfe_un', 'var_counts']	5	3.44384
908	3	['ch_area_norm', 'mfe_un', 'n_2cycles']	4	3.42836
851	3	['ch_area_norm', 'cycle_per_node', 'mfe_un']	4	3.40220
5780	4	['ch_area_norm', 'cycle_per_node', 'mfe_un', 'n_2cycles']	4	3.39830
30323	5	['ch_area_norm', 'long_cycles_per_edge', 'mfe_un', 'n_2cycles', 'var_counts']	5	3.39368

Abbildung 11: Ausschnitt der Rangfolge hinsichtlich der Unterscheidungs- bzw. -Vorhersageperformance von Pendelfahrten verschiedener Featuresets.

Da viele Feature- und Parametersets eine ähnlich gute Performance zeigen, wurde anschließend untersucht, welche Features in den am besten bewertenden Klassifizierungsläufen vorkommen, um anschließend mit dieser Unterauswahl Rechenläufe mit variierenden Featuresets mit dem Gesamtdatensatz durchzuführen. Die unter den 200 besten Clusteringläufen des reduzierten Datensatzes am häufigsten vorkommenden Features sind unter Verwendung des K-Means-Algorithmus in Tabelle 1 dargestellt.

Feature	Beschreibung
<i>num_visits_med</i>	Median der Häufigkeiten an Durchfahrten aller durchquerten Zellen eines Fahrzeugs/Nutzers.
<i>var_counts</i>	Varianz der Häufigkeiten an Durchfahrten aller durchquerten Zellen eines Fahrzeugs/Nutzers.
<i>ch_area_norm</i>	Fläche der konvexen Hülle der durchquerten Zellen, normiert mit der Anzahl der besuchten Gitterzellen (ohne Wiederholungen d.h. jede Zelle zählt einfach unabhängig von der Häufigkeit, mit der sie besucht wurde)
<i>cycle_per_node</i>	Anzahl der zyklischen Wiederholung von Knoten (Hin- und Herfahren) im Graphen, normiert mit der Anzahl der Knoten.
<i>mfe_un</i>	Auftretenshäufigkeit der am meisten auftretenden Kante im ungerichteten Graphen.
<i>sim_score</i>	Ähnlichkeitskennwert, der über die Anzahl der gleichbesuchten Zellen verschiedener Tracks (Fahrten) und der insgesamt besuchten Zellen gebildet wird.

Tabelle 1: Geeignetste Features im Falle des K-Means-Algorithmus zur Klassifizierung von Pendelverkehren.

Im Falle des DBSCAN-Algorithmus stellt sich die Rangfolge wie in Tabelle 2 dar.

Feature	Beschreibung
<i>num_visits_med</i>	Median der Häufigkeiten an Durchfahrten aller durchquerten Zellen eines Fahrzeugs/Nutzers.
<i>n_2_cycles</i>	Anzahl der zyklisch wiederholten Knotenfolge über 2 Knoten hinweg.
<i>long_cycles_per_edge</i>	Anzahl der zyklisch wiederholten Knotenfolge über mehrere Knoten hinweg, normiert an der Anzahl der Kanten.
<i>sim_score</i>	Ähnlichkeitskennwert, der über die Anzahl der gleichbesuchten Zellen verschiedener Tracks (Fahrten) und der insgesamt besuchten Zellen gebildet wird.
<i>var_counts</i>	Varianz der Häufigkeiten an Durchfahrten aller durchquerten Zellen eines Fahrzeugs/Nutzers.

Tabelle 2: Geeignetste Features im Falle des DB-Scan-Algorithmus zur Klassifizierung von Pendelverkehren.

Um eine Visuelle Bewertung der Zuordnungperformance zu ermöglichen wurden Scatterplots erstellt (siehe Abbildung 12 und Abbildung 13), die die Distanz zwischen den Nutzern und den Gruppierungen darstellen und es so ermöglichen Ausreißer zu identifizieren. Zudem wurden Slideshows der Häufigkeitsplots für jede identifizierte Gruppe erstellt, um die Ähnlichkeit der Routenführung und Streckenbefahrungshäufigkeiten visualisieren und bewerten zu können.

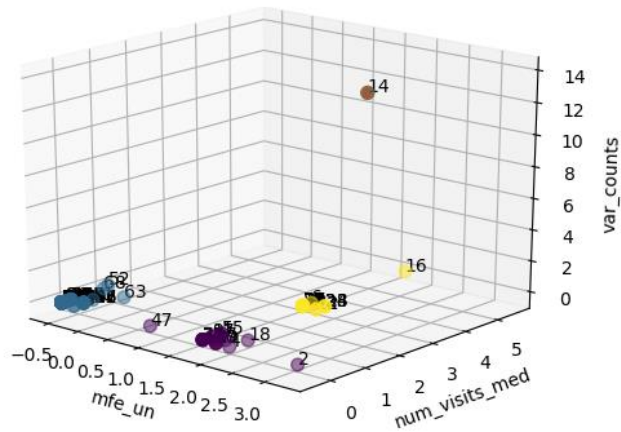


Abbildung 12: Clustering mit KMeans und den Features *num\_visits\_med*, *var\_counts* und *mfe\_un*

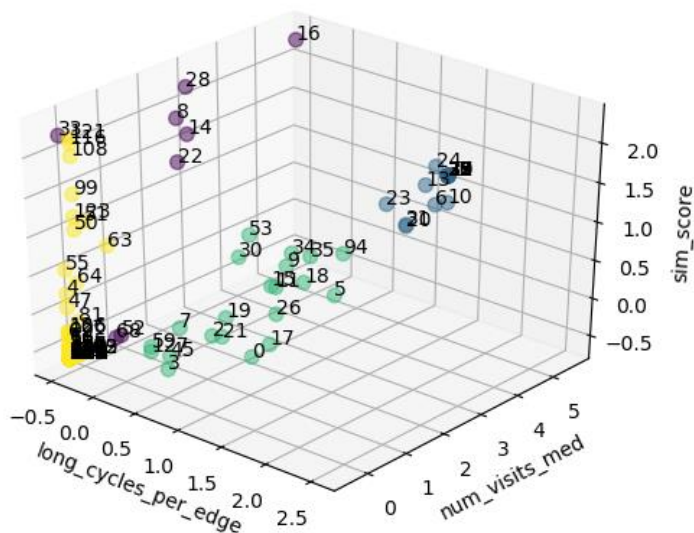


Abbildung 13: Clustering mit DBSCAN und den Features *num\_visits\_med*, *long\_cycles\_per\_edge*, *sim\_score*, *min\_samples = 7*, *eps = 0.8*

Bei der Anwendung der Algorithmen auf den gesamten Datensatz ergaben sich mit diesen Methoden folgende Erkenntnisse:

- Die Suche nach Kreisen, also Hin- und Rückfahrten zwischen zwei Knoten im Graphen dauert zu lange für große Datensätze. Daher sollten alle Features, welche sich auf die Anzahl der Kreise beziehen, nicht mehr betrachtet.

- Die Dichte der Cluster variiert stark. Es gibt deutlich weniger Nicht-Pendler als Pendler und die Pendelverkehre unterscheiden sich stärker untereinander. Dementsprechend ordnet DBSCAN relevante Datenpunkte ggf. keinem Cluster zu.
- Es entstehen Ausreißer durch das Feature *var\_counts*. Dies führt dazu, dass in Clusterings mit KMeans oft dieser einzelne Punkt ein einzelnes Cluster bildet und dieses Feature für alle restlichen Datenpunkte nicht zur Differenzierung beiträgt.
- Auch bei *ch\_area\_norm* sind die Werte sehr ungleichmäßig verteilt, was u.U. ebenfalls Probleme bei der Zuordnung durch den DBSCAN-Algorithmus bereitet.

Algorithmus	Vorteile	Nachteile
KMeans	Parameter (Anzahl der Cluster) sehr leicht festzusetzen; Cluster können unterschiedliche Dichten haben	Einzelne Ausreißer können Ergebnis stark verzerren; Algorithmus nimmt an, dass Cluster konvex sind
DBSCAN	Struktur der Cluster kann beliebig sein; Ausreißer werden auch als solche gekennzeichnet und tragen nicht zur Cluster-Bildung bei	Parameterwahl komplizierter; Algorithmus hat Schwierigkeiten, Cluster unterschiedlicher Dichte gleichzeitig zu finden: relevante Datenpunkte werden auch als Ausreißer eingestuft

Tabelle 3: Gegenüberstellung der Vor- und Nachteile des K-MEANS- und DB-SCAN-Algorithmus.

Insbesondere wegen des zuletzt genannten Nachteils des DBSCAN Algorithmus wurde dieser für den großen Datensatz als weniger geeignet angesehen. Der Fokus der weiteren Untersuchung lag damit auf dem KMeans Algorithmus. Mit diesem wurde ein Clustering für den gesamten Datensatz durchgeführt, dass in Abbildung 12 nach den Features *num\_visits\_med*, *sim\_score* und *mfe\_un\_norm* visualisiert ist.

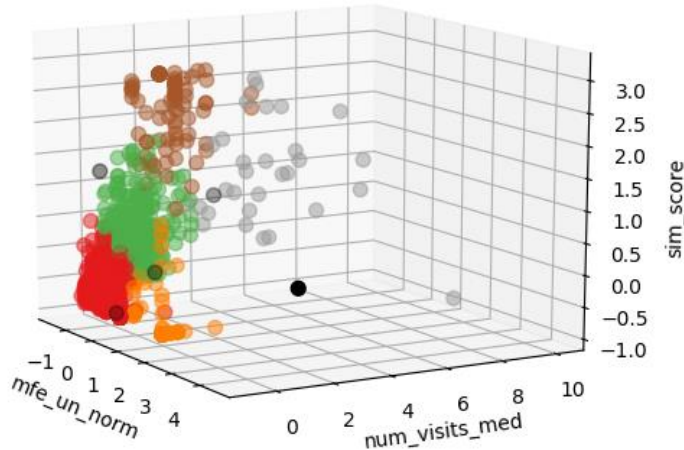


Abbildung 12: Clustering des gesamten Datensatzes mit KMeans: num\_visits\_med, sim\_score, mfe\_un\_norm. Gefundene Cluster sind farblich differenziert.

Folgende Tendenzen zeigen sich in den identifizierten Clustern:

- Nutzer im roten Cluster zeigen keine Pendel-Muster. Häufig gibt es nur einzelne Orte, die oft besucht werden – dies könnte der Heimatort des Nutzers sein.
- Nutzer im grünen Cluster haben einen begrenzten Pendleranteil. Oft gibt es kleine Teilstrecken, die mehrfach befahren werden. Dies erklärt höhere Werte bei *mfe\_un\_norm* und *sim\_score*.
- Das braune Cluster enthält Nutzer mit klaren Pendel-Tendenzen, allerdings mit eher wenigen Tracks – daher die geringeren Werte für *num\_visits\_med*.
- Das orangene Cluster besteht dagegen aus Ausreißern: Der Graph besteht dort oft nur aus einem oder sehr wenigen Knoten. Hier ist also die betrachtete Gittergröße vermutlich zu groß.
- Nutzer im grauen Cluster sind entweder Pendler mit mehr Tracks als im braunen Cluster oder Nutzer mit Teilstrecken, die sehr häufig befahren werden. Hier könnte evtl. sogar noch eine weitere Aufteilung sinnvoll sein.
- *mfe\_un\_norm* und *sim\_score* scheinen eng korreliert zu sein.

## 4.4 Ergebnisse

Tabelle 4 fasst die im vorherigen Abschnitt gezeigten Vor- und Nachteile der untersuchten Features zusammen. Nicht mit in die Tabelle aufgenommen wurden Features die Kreise im Graphen benötigen, da diese für große Datensätze vermutlich nicht umsetzbar sind. Die Kombination aus *num\_visits\_med*, *mfe\_un\_norm* und *sim\_score* hat



sich dabei als besonders erfolgsversprechend dargestellt und kann für weitere Analysen angewendet werden.

Tabelle 4: Vergleich der performantesten Features zur Selektion von Pendelverkehren

Feature	Vorteile	Nachteile
<i>num_visits_med</i>	Gute Identifikation von Pendlern, da diese viele Gridzellen sehr häufig besuchen	Unterscheidet schlecht Nicht-Pendler von Teil-Pendlern
<i>var_counts</i>	Differenzierung von Nicht-Pendler (niedrige Varianz) und Teil-Pendlern (höhere Varianz)	In der Anwendung evtl. schwieriger Umgang mit Ausreißern
<i>mfe_un_norm</i>	Identifikation von Pendeltendenzen möglich, wenn für einen Nutzer wenige Tracks vorhanden sind	Sehr abhängig von Gittergröße für Erzeugung des Graphen
<i>sim_score</i>	Korreliert mit <i>mfe_un_norm</i> , daher s. oben	Stuft Pendler mit wenigen und vielen Tracks gleichermaßen ein
<i>ch_area_norm</i>	Geometrie ist für einige Fälle in Kombination mit Häufigkeitsfeatures aussagekräftig	Umgang mit Ausreißern ist unklar und müsste noch weiter untersucht werden

Bei den weiteren untersuchten Features und deren Kombinationen ist die Bewertung weniger günstig, weil entweder die Rechenzeit erhöht wird (Graphenfeatures, die nach zyklischen Wiederholungen suchen) oder deren Werte für Pendler und Nicht-Pendler zu geringe Differenzen aufweisen. Die visuelle Beurteilung der Häufigkeitsplots für verschiedene Cluster zeigt eine zufriedenstellende Identifizierung von (Teil-) Pendelverkehren. Die hier gewonnenen Erkenntnisse und Algorithmen könnten nun direkt auf den Scania-Datensatz oder ähnliche Anwendungen übertragen werden, um Gebiete mit vermehrt auftretenden Pendelfahrten zu identifizieren.

# 5 AP 5.2: Abgleich des Netzausbaumodells mit realen Lkw-Einsatzdaten

---

## 5.1 Methodik

Um den realitätsnahen Anteil an Pendelverkehren auf den Relationen des Infrastrukturausbaumodells zu bestimmen, sollte die Anzahl der realen Pendelfahrten der Lkw auf dem Ausbaunetz 2030 (siehe Abbildung 14) mit der Anzahl der angenommenen Pendelfahrten im Infrastrukturausbaumodell verglichen werden. Hierzu sollten die mit der in Abschnitt 0 entwickelten Methodik identifizierten Lkw im Pendelverkehr auf einen möglichen wirtschaftlichen Betrieb im Zielnetz hin untersucht und als Vergleichsgröße herangezogen werden.

Zudem sollten alle weiteren Lastkraftwagen, die keine Pendelbewegungen auf dem Ausbaunetz 2030 aufweisen, dahingehend untersucht werden, ob ihr Fahrtanteil im Oberleitungsnetz groß genug ist, um trotzdem wirtschaftlich elektrifiziert werden zu können. Dann sollten diese den entsprechenden Ausbauabschnitten in der Befahrungsmatrix der Optimierungsrechnung zugeordnet werden, möglichst zu Gruppen mit ähnlichen bzw. gleichen Abschnittsbefahrungen gebündelt, um die Berechnungszeit des Optimierungsmodells möglichst wenig zu beeinträchtigen.

Um die realen Kraftstoffverbräuche von aktuell eingesetzten Lkw auf den Abschnitten des Ausbaunetzes 2030 zu ermitteln, war geplant die mittleren Verbräuche der Fahrzeuge des Scania-Datensatzes nach Lkw-Gewichtsklassen sortiert zu ermitteln und mit den simulierten Verbräuchen, die in die Optimierungsrechnung zur Ausbaupfadermittlung eingingen, zu vergleichen. Zur Sensitivitätsbetrachtung sollten darüber hinaus weitere statistische Größen, wie Minima, Maxima und Mediane in den Vergleich einbezogen werden. Ggf. müssten die Verbrauchswerte dann weiter angepasst werden, wobei hier die einbezogenen Altersklassen der Lkw im Scania-Datensatz berücksichtigt werden sollten, um den Effizienzentwicklungen der Fahrzeuge Rechnung zu tragen.

Die gewonnenen Erkenntnisse und Zahlenwerte sollten dann zur Anpassung der bisherigen Eingangsparameter dienen, um eine weitere Optimierungsrechnung zur optimalen Netzausbaufolge der Abschnitte durchzuführen. Deren Resultat könnte dann mit den bisher generierten Ergebnissen Streckenabschnittsweise abgeglichen werden, um den Einfluss realer Fahrprofile abzuschätzen und die Aussagegenauigkeit der bisherigen Eingangsdatensätze zu beleuchten.

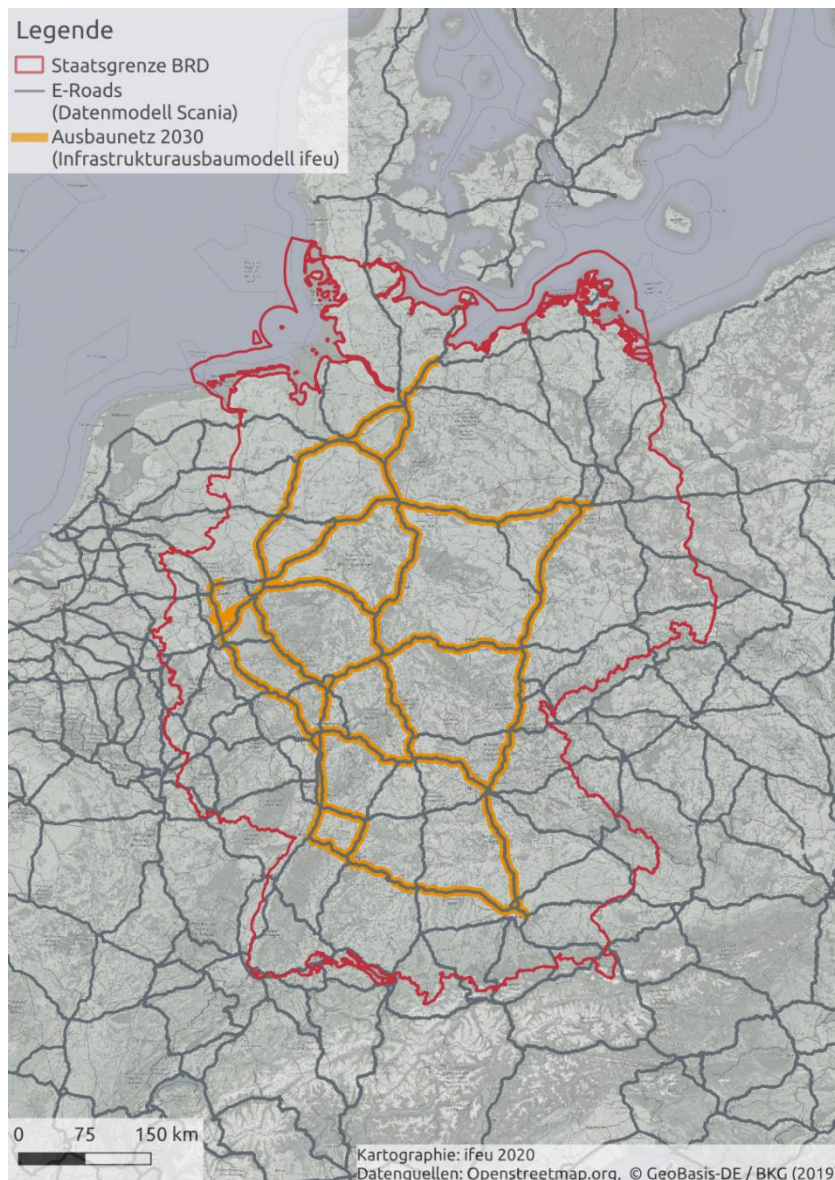


Abbildung 14: Ausbaunetz 2030 des Infrastrukturausbaumodells (ifeu) und Straßennetz (E-Roads) des Datenmodells bei Scania.

## 5.2 Ergebnisse

Erste Gegenüberstellungen des Scania-Datensatzes, die vor Beendigung der Kooperation durchgeführt wurden, zeigten eine gute Übereinstimmung der Befahrungsstärken auf den Straßenabschnitten zwischen (a) den hochgerechneten Flottendaten und (b) den bisherigen Ausbaunetzmodellaten, die zu Teilen auf den Simulationen des Verkehrsmodells Validate von PTV basieren. Weitere Kalibrierungen auf Basis der Verkehrszählungen aus (BASt 2017) könnten die Genauigkeit und Präzision der Hochrechnung der auf Telematik basierten Flottendaten für lokale Straßenabschnitte des Untersuchungsgebietes (Ausbaunetz 2030) weiter erhöhen. Zu berücksichtigen wären weitere Nebenbedingungen, wie unter anderem die Gesamtfahrleistungen der Lkw-Größenklassen, die jeweilige Anzahl zugelassener Lkw und ggf. den Befahrungsmessgrößen auf weiteren Nebenstrecken (siehe Abbildung 15)..



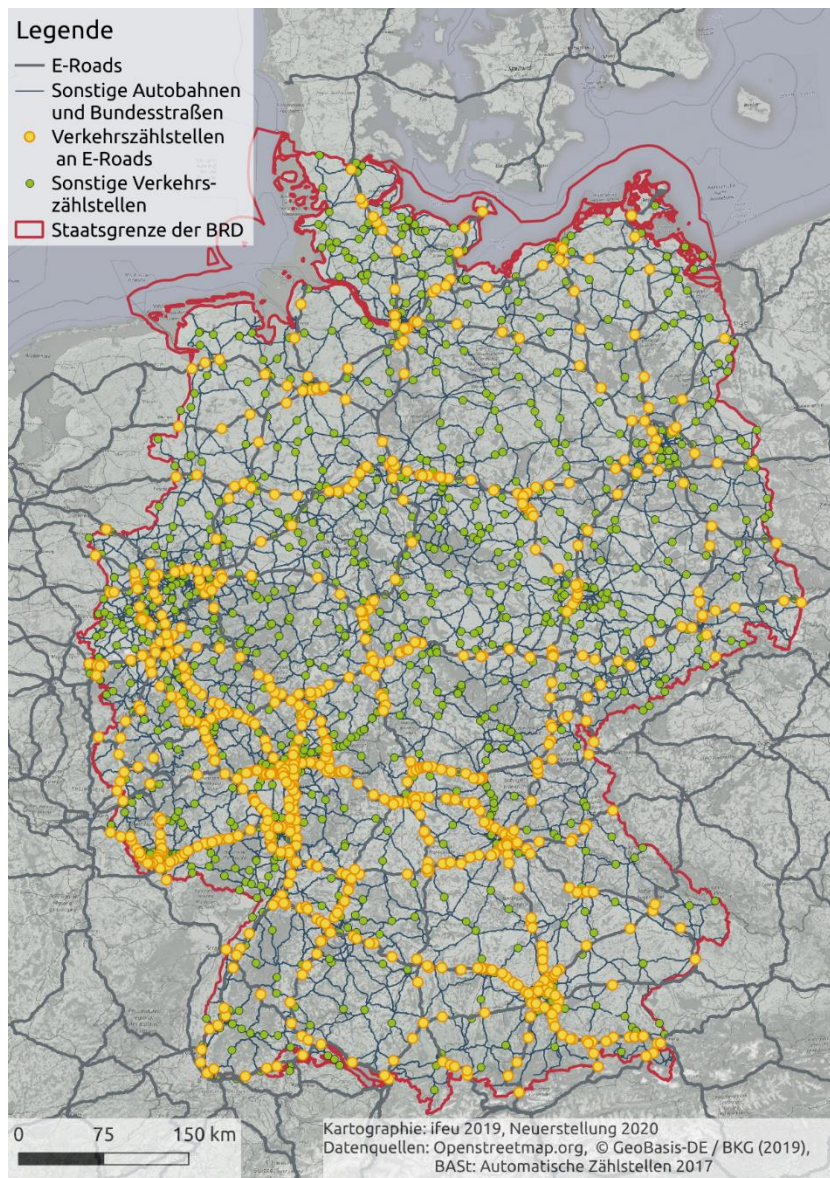


Abbildung 15: Automatische Zählstellen entlang der E-Roads sowie auf weiteren Autobahnen und Bundesstraßen innerhalb Deutschlands.

Auf Basis des entwickelten Algorithmus sollten dann Straßenabschnitte mit vermehrt auftretenden Pendelverkehren im Scania-Datensatz mit dem bisher identifizierten Datensatz verglichen werden, um die Auftrittshäufigkeit von Pendelverkehren auf den untersuchten Relationen des Ausbaumodells zu identifizieren. Dieser Schritt konnte im Rahmen der Kooperation jedoch nicht mehr durchgeführt werden. Eine Untersuchung der Eignung von Fahrzeugen ohne signifikantes Pendelverhalten auf dem Ausbaumodell 2030 erfolgte ebenfalls nicht.

Bezüglich der Realverbräuche von Lkw auf verschiedenen Abschnitten konnten lediglich Testläufe zur Bestimmung regional eingegrenzter Verbrauchsmittelwerte für verschiedene Lkw-Klassen durchgeführt werden, die sich jedoch prinzipiell auf die zu untersuchenden Abschnitte anwenden lassen würden.

## 6 AP 5.3: Schlussfolgerungen hinsichtlich des Road Mappings

---

Aufgrund des Bearbeitungsstandes können noch keine endgültigen Schlussfolgerungen für die Erstellung der Roadmap gezogen werden. Allerdings ergaben sich während der Arbeiten mit den Daten bereits folgende Beobachtungen und Indizien:

- Die reale Nutzung einzelner Lkw unterscheidet sich in Teilen deutlich von der angenommenen Routenführung auf Basis der Güterströme und den daraus abgeleiteten Lkw-Transportfahrten des Verkehrsmodells von PTV. Bei dieser wurden die Lkws so verteilt, dass möglichst viele Pendelbeziehungen innerhalb realistischer Lenkzeiten und Nutzungsbedingungen entstehen.
- So kommt es in der Realität anscheinend in Teilen häufiger vor, dass ein Lkw Waren von A nach B transportiert, um anschließend Transportaufträge von B oder C zum Zielort D zu transportieren, anstatt weitere Güter von B nach A zu befördern oder leer zurückzufahren während die mögliche Fahrt von B nach A von einem weiteren Lkw übernommen wird. Eine Quantifizierung der Unterschiede vom Transportmodell für das Zielnetz 2030 und den abgeleiteten Pendelfahrten aus den realen Lkw-Einsatzdaten konnte jedoch nicht mehr durchgeführt werden.
- Die Verkehrsverflechtungen durch internationale und Transitverkehre innerhalb Europas zeigen nach ersten Analysen, dass der Anteil wirtschaftlicher Oberleitungs-Lkw durch weiteren koordinierten Ausbau der Oberleitungsinfrastruktur in den Nachbarländern zunimmt.
- Abhängig ist die Zunahme im Wesentlichen von den Betriebsbedingungen und dem Anteil internationaler und durchgehender Verkehre auf der entsprechenden Strecke. Auch die geographische Lage der elektrifizierenden Länder spielt eine Rolle: Je zentraler gelegen und je geringer die Fläche eines Staates ist, desto eher profitiert dieser von einem zusätzlichen Infrastrukturausbau in Nachbarstaaten. Während Großbritannien kaum durch Elektrifizierung europäischer Nachbarn beeinflusst wird, profitieren Staaten wie Polen, Niederlande oder Luxemburg in höherem Maße.
- Im Fall Deutschlands ließen sich auf Basis erster Analysen, die eine Bepreisung der CO<sub>2</sub>-Emissionen im Verkehr berücksichtigen, signifikante zusätzliche Verkehre verzeichnen, vor allem auf Streckenabschnitten mit höherem internationalen Transportanteil. Die abschließende Validierung des Rechen- und Kostenmodells steht jedoch noch aus. Auch der Einbezug verschiedener Förderinstrumente zur Umstellung von Lkw auf den Oberleitungsbetrieb ist noch nicht erfolgt.



# Literaturverzeichnis

---

ACEA - European Automobile Manufacturers' Association (2018): In: *Consolidated Registrations - By Manufacturer*. <https://www.acea.be/statistics/tag/category/by-manufacturer-registrations>. (16.09.2020).

BAST (2017): Automatische Zählstellen auf Autobahnen und Bundesstraßen. In: *BAST 2017 - Fachthemen – Verkehrstechnik - Automatische Zählstellen auf Autobahnen und Bundesstraßen*. [https://www.bast.de/BAST\\_2017/DE/Verkehrstechnik/Fachthemen/v2-verkehrsaehlung/Verkehrsaehlung.html](https://www.bast.de/BAST_2017/DE/Verkehrstechnik/Fachthemen/v2-verkehrsaehlung/Verkehrsaehlung.html). (17.09.2020).

D. L. Davies; D. W. Bouldin (1979): A Cluster Separation Measure. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. PAMI-1, No.2, S. 224–227.

EUROSTAT (2019a): Road freight transport by vehicle characteristics. In: *Road freight transport by vehicle characteristics - Statistics Explained*. [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road\\_freight\\_transport\\_by\\_vehicle\\_characteristics#Vehicle\\_fleets](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road_freight_transport_by_vehicle_characteristics#Vehicle_fleets).

EUROSTAT (2019b): Energy prices 2019. In: *Energy prices in the EU, 2019*. <https://ec.europa.eu/eurostat/news/themes-in-the-spotlight/energy-prices-2019>.

Hubert, L.; Arabie, P. (1985): Comparing partitions. In: *Journal of Classification*. Vol. 2, No.1, S. 193–218.

Mariescu-Istodor, R.; Fränti, P. (2017): Grid-Based Method for GPS Route Analysis for Retrieval. In: *ACM Transactions on Spatial Algorithms and Systems*. Vol. 3, No.3, S. 1–28.

Petterson, J.; Resch, S.; thor Straten, E.; Miettinen, H.-M. (2018): Trans-European Road Network, TEN-T (Roads) 2017 Performance Report. <https://www.cedr.eu/download/Publications/2018/TEN-T-Performance-report-2017.pdf>.

Rosenberg, A.; Hirschberg, J. (2007): V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure. S. 11.

Rousseeuw, P. J. (1987): Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. In: *Journal of Computational and Applied Mathematics*. Vol. 20, S. 53–65.

THE SCANIA REPORT 2018 (2019): Annual and Sustainability Report [https://www.volkswagenag.com/presence/investorrelation/publications/annual-reports/2019/scania/Scania\\_Annual\\_Report\\_2018.pdf](https://www.volkswagenag.com/presence/investorrelation/publications/annual-reports/2019/scania/Scania_Annual_Report_2018.pdf) (16.09.2020).

Traffic Census 2015 - Transport - UNECE (2015): [https://www.un-ece.org/trans/main/wp6/e-roads\\_census\\_2015.html](https://www.un-ece.org/trans/main/wp6/e-roads_census_2015.html).

