BMVI-Förderprojekt im Rahmen der Forschungsinitiative mFUND – das Startkapital für die Mobilität 4.0

mobileVIEW

Verbesserung der Kurzfristvorhersage von Niederschlagsereignissen mittels Fahrzeugsensoren

Gemeinsamer Schlussbericht des Verbundvorhabens

Förderkennzeichen: 19F2024A-C
Projektlaufzeit: 01.10.2017 bis 31.12.2020

GEFÖRDERT VOM

Bundesministerium für Verkehr und digitale Infrastruktur

mFUND
Das Startkapital für die Mobilität 4.0

BMVI-Förderprojekt
im Rahmen der Forschungsinitiative mFUND –
das Startkapital für die Mobilität 4.0

mobile^{VIEW}

Verbesserung der Kurzfristvorhersage von
Niederschlagsereignissen mittels Fahrzeugsensoren

19F2024A-C

mobile^{VIEW}

IMPRESSUM

Verbundkoordination

Forschungsinstitut für Wasser- und Abfallwirtschaft an der RWTH Aachen e. V. (FiW)
Mark Braun, M. Sc.
Dr. Frank-Andreas Weber
Kackertstraße 15 – 17
52072 Aachen

GEFÖRDERT VOM

Bundesministerium für Verkehr und digitale Infrastruktur

mFUND
Das Startkapital für die Mobilität 4.0

VDI | VDE | IT

Projektpartner

Forschungsinstitut für Wasser- und Abfallwirtschaft an der RWTH Aachen (FiW) e. V.
Kackertstraße 15 – 17
52072 Aachen

Mark Braun, M. Sc.
Dr. sc. Dipl.-Ing. Frank-Andreas Weber
Dr.-Ing. Natalie Palm
Dr.-Ing. Gesa Kutschera
Magnus Hoffmann, M. Sc.
Philipp Knollmann, M. Sc.
Sajjad Tabatabaei, M. Sc.

Emschergenossenschaft / Lippeverband (EGLV)
Kronprinzenstraße 24
45128 Essen

Dipl.-Geogr. Angela Pfister
Dipl.-Geogr. Adrian Treis
Dimitri Falk, M. Sc.

IAV GmbH Ingenieurgesellschaft Auto und Verkehr
Carnotstraße 1
10587 Berlin

Dr. Etienne Costa-Patry
Dr. Kay-Jochen Langeheinecke
Dr.-Ing. Matthias Pätsch

Gestaltung und Satz

Forschungsinstitut für Wasser- und Abfallwirtschaft an der RWTH Aachen (FiW) e. V.
Tanja Dohr

Zitierhinweis

Braun, M., Hoffmann, M., Knollmann, P., Tabatabaei, S., Palm, N., Weber, F.-A., Kutschera, G., Pfister, A., Treis, A., Falk, D., Costa-Patry, E., Pätsch, M., Langeheinecke, K.-J. (2021): mobile^{VIEW} – Verbesserung der Kurzfristvorhersage von Niederschlagsereignissen mittels Fahrzeugsensoren. Abschlussbericht des BMVI-Vorhabens mobile^{VIEW}, Förderkennzeichen 19F2024A-C, FiW e. V., Aachen.

Förderkennzeichen

Das diesem Dokument zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Verkehr und digitale Infrastruktur im Rahmen der Forschungsinitiative mFUND unter dem Förderkennzeichen 19F2024A-C gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Aachen, März 2021

INHALT

Impressum **4**

Abkürzungsverzeichnis **6**

1 Aufgabenstellung **9**

2 Voraussetzungen **10**

3 Planung und Ablauf des Verbundvorhabens **11**

3.1 TP1: Datensammlung 11

3.2 TP2: Datenverarbeitung 12

3.3 TP3: Datenverwertung 12

3.4 TP4: Projektmanagement & Qualitätssicherung 13

4 Stand der Technik **14**

4.1 Projekt RAINCARS 14

4.2 Windshield Wipers on connected Vehicles produce high-accuracy Rainfall Maps (Bartos et al. 2019) 14

4.3 Weitere Projekte 15

4.4 Patentlage 15

4.5 Verwendete Informationsdienste und Datenquellen 15

5 Bisherige Arbeiten des Konsortiums **16**

5.1 Forschungsinstitut für Wasser- und Abfallwirtschaft an der RWTH Aachen (FiW) e. V. 16

5.2 Emschergenossenschaft und Lippeverband 17

5.3 IAV GmbH Ingenieurgesellschaft Auto und Verkehr 18

6 Ergebnisse **19**

6.1 Detaillierte Ergebnisdarstellung 19

6.1.1 Teilprojekt TP 1: Datensammlung 19

6.1.2 Teilprojekt TP 2: Datenverarbeitung 31

6.1.3 Teilprojekt TP 3: Datenverwertung 60

6.1.4 Teilprojekt TP 4: Projektmanagement und Qualitätssicherung 87

7 Fazit und Ausblick **89**

8 Literatur **93**

Abbildungsverzeichnis **94**

Tabellenverzeichnis **96**

Bildnachweise **96**

ABKÜRZUNGSVERZEICHNIS

ALKIS ..... Amtliches Liegenschaftskatasterinformationssystem (ALKIS)
ATKIS ..... Amtliches Topographisch-Kartographisches Informationssystem (ATKIS)
BMVI ..... Bundesministerium für Verkehr und digitale Infrastruktur
CAN ..... Controller Area Network
CDC ..... Climate Data Center des DWD
CET ..... Central European Time (auch MEZ – Mitteleuropäische Zeit)
DSB ..... Datensammelbox
DSGVO ..... Datenschutzgrundverordnung der Europäischen Union
DWA ..... Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall e. V.
DWD ..... Deutscher Wetterdienst
DX-D ..... Dämpfungskorrigiertes DX-Niederschlagsradarprodukt
EGLV ..... Emschergenossenschaft und Lippeverband
FEWS ..... Flood Early Warning System
FiW ..... Forschungsinstitut für Wasser- und Abfallwirtschaft an der RWTH Aachen (FiW) e. V.
FFNN ..... Feedforward Neural Network
FKZ ..... Förderkennzeichen
FTP ..... File Transfer Protocol
GPU ..... Graphics Processing Unit
HOWIS ..... Hochwasserinformationssystem
IAV ..... IAV GmbH Ingenieurgesellschaft Auto und Verkehr
IoT ..... Internet of Things
JSON ..... JavaScript Object Notation
Kfz ..... Kraftfahrzeug
KI ..... Künstliche Intelligenz
KNN ..... Künstliches neuronales Netz (auch NN oder ANN)
LSTM ..... Long short-term memory
MLP ..... Multi-Layer Perceptron
MSE ..... Mean-squared error
MSIC ..... Multi-Source-Information-Cluster
MSPM ..... Multi-Source-Precipitation-Map
MULNV ..... Ministerium für Umwelt-, Landwirtschaft-, Natur- und Verbraucherschutz NRW
MQTT ..... Message Queuing Telemetry Transport
OBD ..... On-Board-Diagnose
OEM ..... Original Equipment Manufacturer

QoS ..... Quality of Service

REST ..... Representational State Transfer

RLS ..... Regen-Licht-Sensor

RMSE ..... Root-mean-square error

RNN ..... Recurrent neural networks

SRI ..... Starkregenindex

TOM ..... Technisch-organisatorische Maßnahmen

TP ..... Teilprojekt

UTC ..... Coordinated Universal Time

UTM ..... Universal Transverse Mercator

1 AUFGABENSTELLUNG

Das Ziel des Forschungsprojektes mobile^{VIEW} war die Kombination von Sensordaten aus fahrenden Kfz mit Daten von stationsgebundenen Erhebungsquellen sowie Radar- und Satellitendaten zur Erfassung von Niederschlag. Durch diese Kombination sollten kurzfristige Warnungen vor Starkregenereignissen im urbanen und ruralen Raum zukünftig ermöglicht werden.

Als Projektgebiet wurde das Einzugsgebiet der Emschergenossenschaft und des Lippeverbands in Nordrhein-Westfalen ausgewählt, in dem sowohl urbane als auch rurale Strukturen existieren und das eine hohe Dichte an terrestrischen Niederschlagsstationen aufweist. Die für diese Region bestehenden Daten des Deutschen Wetterdienstes (DWD), eigene Niederschlagsradarprodukte von Emschergenossenschaft und Lippeverband (EGLV) sowie das terrestrische Messnetz an Pluviometern sollten um in Kfz erhobene Sensordaten erweitert werden. Für die Datenerhebung wurden 100 Fahrzeuge der EGLV-Fahrzeugflotte mit Sensoren und der notwendigen Übertragungstechnik ausgestattet. Die erhobenen Daten wurden anschließend an einen zentralen Server zur sicheren und datenschutzkonformen Speicherung und Weiterverarbeitung übermittelt.

Zentraler Aspekt der Datenverarbeitung war die Fragestellung, ob in Kfz erhobene Sensordaten generell für die Ableitung von Niederschlagsinformationen geeignet sind. Ist diese Voraussetzung gegeben, sollten die veredelten Daten hinsichtlich ihrer Verwendung im Warnmanagement bei Starkregen und im hydrologischen Kontext untersucht werden. Darauf aufbauend können weitere Verfahren wie Kürzestfristvorhersagen im Nowcastingbereich abgeleitet werden. Um ein belastbares Niederschlagsprodukt zu erhalten, war eine Kombination mit den Pluviometern und dem Niederschlagsradar erforderlich. Im Projektverlauf wurden zudem weitere Echtzeit-Drittdaten auf ihre Anwendbarkeit geprüft.

Die wesentliche Aufgabe bestand darin, aus einer großen Anzahl von Einzelinformationen (Big Data) eine Gesamtinformation bzw. einen Mehrwert (Smart Data) zu generieren, der der Öffentlichkeit zugutekommt. Die Datenerhebung in den Kfz sollte keinen einseitigen Informationsfluss darstellen. Die generierten Niederschlagsinformationen sollten nach Projektende an die Nutzer zurück übertragen werden, z. B. in Form einer Verkehrsinformation oder Warnung. Auch weitergehende Nutzungsmöglichkeiten wurden im Zuge des Projektes untersucht.

Die Zielgruppe des Forschungsprojektes waren somit diverse Akteure. Behörden und Verbände im Bereich der Wasserwirtschaft sollten in die Lage versetzt werden, zielgerichtet über die aktuellen Niederschlagsverhältnisse zu informieren, um entsprechende Maßnahmen einleiten zu können. Beispielhaft ist hier die Steuerung von wasserwirtschaftlichen Anlagen (Regenbecken etc.) zu nennen. Auch andere administrative Einheiten wie die Feuerwehr, der Katastrophenschutz oder Verkehrsleitzentralen können von verdichteten Niederschlagsinformationen profitieren. Weitere Profiteure der Informationsverdichtung sind juristische Personen und Einzelpersonen. Firmen, die durch Überschwemmungen Produktionsausfälle zu befürchten haben und Eigenheimbesitzer oder Verkehrsteilnehmer, könnten so entsprechend früher Maßnahmen zum Schutz ihrer Güter und eigenen Sicherheit in die Wege leiten.

Das Forschungsprojekt mobile^{VIEW} lieferte somit einen Beitrag zur Digitalisierungsstrategie der deutschen Wasserwirtschaft und verfolgte dabei folgende übergreifende Ziele:

1. Durch eine Verdichtung der in der Fläche erhobenen Daten sollte eine Verbesserung der Echtzeit-Flächeninformation über den Niederschlag erreicht werden.
2. Mittels der Möglichkeit zur Einbindung der Bevölkerung in die Datenerhebung sollte das Systemverständnis der wasserwirtschaftlichen Aufgabenstellungen erhöht werden. Die in den Kfz und anderer im Verkehrsbereich erhobenen Daten verdeutlichten die Relevanz von digitalen Schnittstellen zwischen verschiedenen Wirtschaftssektoren, hier der Wasser- und Verkehrswirtschaft.
3. Durch die Verdichtung der Eingangsdaten aus mobilen Sensorträgern (Kfz) für die hydrologische Modellierung sollte eine Verbesserung der Abflussmodellierung sowie die Starkregenvorhersage erreicht werden.

2 VORAUSSETZUNGEN

Die deutsche Wasserwirtschaft befindet sich in vielen ihrer Bereiche in einem kontinuierlichen Digitalisierungsprozess. Die Erhebung, Vernetzung und Verfügbarkeit digital erhobener Daten ermöglichen eine ständige Verbesserung der Analyse wasserwirtschaftlicher Prozesse. Durch die erhöhte zeitliche und räumliche Verdichtung von relevanten Informationen kann der Detaillierungsgrad von Eingangsdaten für hydrologische Modelle sowie die Genauigkeit der Prozessabbildung verbessert werden.

Auch die Niederschlagsvorhersage bietet vor allem bei kleinräumigen Niederschlagsereignissen ein großes Potenzial für Verbesserungen in der Prozessabbildung. Die aktuelle Datenerfassung mit terrestrischen Niederschlagsstationen und 16 Niederschlagsradarstandorten für ganz Deutschland weist aufgrund der variierenden Dichte des Messstellennetzes und der spezifischen Fehlerquellen von Radarprodukten (siehe Kapitel 4) in Bezug auf Starkregenereignisse gewisse Unsicherheiten auf. Hier besteht der Bedarf an einer weiteren unabhängigen Datenquelle, die vor allem die Erfassung von kleinräumigen und hinsichtlich ihrer Intensität und Zugrichtung stark variablen Starkregenereignissen in Echtzeit ermöglicht.

Im Zuge des Klimawandels ist zukünftig mit einer Zunahme an Starkregenereignissen zu rechnen (IPCC 2013). Gleichzeitig besitzen Starkregenereignisse aufgrund ihres Gefahrenpotenzials und potenziell ökologischen, ökonomischen sowie sozialen Auswirkungen eine hohe Relevanz für Privatpersonen, Gewerbe/Industrie und administrative Einheiten in Behörden und Verwaltung.

Das Projekt mobile^{VIEW} kann an dieser Stelle dazu beitragen, durch die Verdichtung von Niederschlagsinformationen eine frühere Information über aktuelle Niederschlagsverhältnisse und infolgedessen einen verbesserten Schutz von menschlichem Leben sowie die Reduzierung ökonomischer Schäden durch Starkregenereignisse zu erzielen.

3 PLANUNG UND ABLAUF DES VERBUNDVORHABENS

Das Forschungsprojekt wurde über drei Jahre im Zeitraum von Oktober 2017 bis Dezember 2020 durchgeführt. Die Struktur des Vorhabens bestand aus vier einzelnen Teilprojekten, die zahlreiche gegenseitige Schnittstellen zueinander aufweisen. Dennoch verfolgte jedes Teilprojekt individuelle spezifische Ziele und Ergebnisse, die in Form von Meilensteinen den Gesamtprojektverlauf skizzieren. Das Vorhaben gliederte sich in vier Hauptarbeitspakete bzw. Teilprojekte:

- TP1: Datensammlung
- TP2: Datenverarbeitung
- TP3: Datenverwertung
- TP4: Projektmanagement & Qualitätssicherung

3.1 TP1: Datensammlung

Im ersten Teilprojekt erfolgte die Entwicklung und Zusammenstellung der für das Projekt benötigten Hard- und Software. Die Hardware umfasste dabei die Komponenten, die in den Kfz für die Datenerhebung (u. a. Regen-Licht-Sensoren) und Datenübertragung (Datenboxen) erforderlich waren. Bezüglich der Software ist vorrangig die Einrichtung und Programmierung der Datenboxen mittels einer Steuerungssoftware zu nennen, die die Datenübertragung gemäß dem festgelegten zeitlichen Intervall ermöglichte.

Nach der Definition der Anforderungen an die Datenerhebung, -übertragung, -speicherung und an die Ermöglichung des Zugriffs im Sinne einer weiteren wasserwirtschaftlichen Verwendung wurde ein Hardwarekonzept für die Übertragungstechnik entwickelt und in 100 Kfz von Emschergenossenschaft und Lippeverband (EGLV) eingebaut. Die Datenerhebung und Datenübertragung sollten dabei über eine Datenbox ausgeführt werden. Die Datenspeicherung und die Ermöglichung des Zugriffs wurden auf einem geschützten Server durchgeführt. Weiterhin wurden datenschutzrechtlich unkritische Daten für eine Veröffentlichung in der mCLOUD des BMVI vorbereitet.

Parallel dazu wurden Methodenpakete entwickelt, um die Verfahren zur Datengewinnung festzulegen. Bis zum Projektende wurden die entwickelten Komponenten ständig überprüft und ggf. angepasst. Regelmäßige Workshops rundeten das Teilprojekt ab.

Gemäß der Datenschutzgrundverordnung der Europäischen Union (DSGVO EU) ist die Verarbeitung von personenbezogenen Daten an strenge Vorgaben geknüpft. Auf das Projekt mobile^{VIEW} bezogen, war die Erhebung, Speicherung und Verarbeitung der in den Kfz erhobenen Daten eine Verarbeitung von personenbezogenen Daten. Der Schutz dieser Daten musste zu jedem Zeitpunkt gewährleistet sein, die Nachweispflicht der technisch-organisatorischen Maßnahmen (TOM) lag hier beim Konsortium. Weiterhin musste das Konsortium die Rechtmäßigkeit der Datenverarbeitung gemäß Art. 6 DSGVO belegen.

Folgende Meilensteine wurden in TP1 erreicht:

- MS 1: Anforderungskatalog an Hard- und Software erstellt
- MS 2: Datenschutz personenbezogener Daten gemäß DSGVO gewährleistet
- MS 3: Hard- und Software funktionieren und liefern Daten aus diversen Kfz
- MS 4: Datenübertragung in die Datenbank funktioniert im Testbetrieb

3.2 TP2: Datenverarbeitung

In Teilprojekt 2 wurden die in den Kfz erhobenen Daten und die recherchierten Drittdaten hinsichtlich ihrer Verwertbarkeit für wasserwirtschaftliche Anwendungen analysiert und verarbeitet. Die dafür notwendige Hardware (Rechner-Workstation), die Software (u. a. Datenbanksystem) sowie die Methoden wurden in diesen Arbeitspaketen beschafft, entwickelt bzw. programmiert und zusammengestellt.

Für die Bereitstellung der Daten wurde ein Datenbanksystem aufgesetzt, in dem die erhobenen Daten, Drittdaten und kombinierte Produkte über eine Schnittstelle ausgetauscht werden konnten. Die Vereinheitlichung von Datenattributen war an dieser Stelle eine zentrale Herausforderung. Außerdem war der Standort des zentralen Servers festzulegen.

Anschließend fand eine Plausibilisierung der Daten im Hinblick auf ihre Verwendungsfähigkeit und Qualität statt. Ein Kernstück dieses Teilprojektes war in diesem Zusammenhang die Konzeptionierung der Informationsmodellierung und die Nutzung kognitiver Technologien für die Verarbeitung, anschließende Plausibilisierung und Qualitätskontrolle der Daten. Zu Beginn wurden die technischen Herausforderungen der Datenspeicherung anhand artifizieller und bereits verfügbarer Daten gelöst. Nachdem die in den Kfz erhobenen Daten verfügbar waren, wurden diese mitberücksichtigt. Eine bedeutende Herausforderung lag in der Entwicklung von Algorithmen, die eine Filterung vornehmen und in der Lage sind, die erhobenen Daten mit möglichst kurzer Latenzzeiten für Vorhersagemodelle in Teilprojekt 3 zur Verfügung zu stellen. In Teilprojekt 2 bestand daher ein hoher Abstimmungsbedarf in seiner Funktion als Schnittstelle zwischen den Teilprojekten 1 und 3.

Ein weiterer Bestandteil des Teilprojektes 2 war die Aufbereitung der Daten für eine potenzielle Bereitstellung über eine öffentlich zugängliche Schnittstelle. Hierbei wurden verschiedene Möglichkeiten, bspw. eine Smartphone-App, ein Web-Interface oder die Veröffentlichung über einen Cloud-Dienst in Erwägung gezogen.

Folgende Meilensteine wurden in TP2 erreicht:

- MS 5: Die Plausibilisierung der Rohdaten genügt den Qualitätsansprüchen
- MS 6: Die Verwandlung der Daten in Informationen liefert die gewünschten Ergebnisse
- MS 7: Die automatisierte Datenübertragung zu einer öffentlich zugänglichen Schnittstelle funktioniert
- MS 8: Die Informationen genügen dem Qualitätsmanagement der Vorhersage

3.3 TP3: Datenverwertung

In Teilprojekt 3 wurden die Daten der Verwendung in der Meteorologie und Hydrologie zugeführt. Dazu mussten die in den Teilprojekten 1 und 2 verarbeiteten Daten mit Informationen aus bestehenden Datenquellen verknüpft werden. Die bestehenden Erhebungsquellen wurden zunächst recherchiert, in Echtzeit verfügbar gemacht und in Hinblick auf ihre Qualität kontinuierlich geprüft. Als Produkt der Datenkombination sollten Abbildungen des Gebietsniederschlags für das ausgewählte Untersuchungsgebiet entstehen, in denen eine möglichst gute Nachbildung der tatsächlichen Niederschlagsverhältnisse und eine Verbesserung der Vorhersage angestrebt wurden.

Im ersten Schritt wurden die Anforderungen an die über eine Schnittstelle transferierten Daten aus den Teilprojekten 1 und 2 definiert. Anschließend wurden Methoden entwickelt, um die gewonnenen Informationen in bestehende Modelle integrieren zu können. Dazu wurden ausgewählte Softwarepakete für die Modellierung und für die Vorhersage eingesetzt und teilweise weiterentwickelt, bspw. mussten Algorithmen zur Identifikation von Mustern, Korrelationen und Abhängigkeiten programmiert werden. Entscheidende Zielaspekte waren an dieser Stelle die räumliche und zeitliche Verdichtung von Niederschlagsinformationen und die Erhöhung der Prognosegenauigkeit bestehen-

der Modelle. Konkrete Einsatzmöglichkeiten ergaben sich in den bestehenden Hochwasservorhersagemodellen von EGLV. In diesem Zusammenhang war eine Plausibilisierung und Verifizierung der eingespeisten Daten durchzuführen.

Darüber hinaus dienten die Modelle der historischen Ereignisanalyse und waren damit das optimale Werkzeug für die Abschätzung der Anwendungsmöglichkeit anderer Echtzeitdaten für die Vorhersage, die sich aus bislang unbekannten Zusammenhängen zwischen einzelnen Kfz-Sensorinformationen ergaben. Abgerundet wurde das Teilprojekt über die Entwicklung einer Möglichkeit zur Veröffentlichung der erhobenen Daten und gewonnenen Erkenntnisse.

In Teilprojekt 3 wurde der folgende finale Meilenstein erreicht:

- MS 9: Erfolgreiche Integration der Daten in die Vorhersage

3.4 TP4: Projektmanagement & Qualitätssicherung

Als kontinuierliches und begleitendes Teilprojekt hat das FiW das zielgerichtete Initiieren, Steuern, Kontrollieren und Abschließen der einzelnen Meilensteine und des Gesamtergebnisses gewährleistet. Für die Arbeitspakete wurden Qualitätssicherungsvereinbarungen für jeden Partner getroffen, die auch durch den jeweiligen Projektleiter des Arbeitspaketes zu verantworten waren.

Im Rahmen des Arbeitspaketes wurden die einzelnen Teilprojekte, Teilaufgaben und Arbeitspakete abgearbeitet und die Verknüpfung zwischen den Arbeitspaketen sichergestellt. Dazu wurden geeignete Kommunikationsmethoden angewendet, um den Projektfortschritt ständig kontrollieren und bei Bedarf anpassen zu können. Durch die Verantwortung jedes Projektpartners für ein Teilprojekt wurden die Schnittstellen innerhalb des Konsortiums sichergestellt. Ein weiterer Bestandteil dieses Teilprojektes war neben der Aufgabenplanung auch die Kosten- und Zeitplanung über den gesamten Projektverlauf.

4 STAND DER TECHNIK

4.1 Projekt RAINCARS

Die bewusste dauerhafte Nutzung von mobilen Kfz-Sensoren für die Erfassung von Niederschlags- und Klimadaten war zum Projektstart 2017 nur theoretisch untersucht und mit Laborversuchen zur Genauigkeit der Erfassung der Regenmenge eines Regensensors im Projekt RAINCARS (Rabiei et al. 2013) vertieft worden. Das von 2011 bis 2013 durchgeführte Projekt hat anschließend praktische Versuche mit zwölf Kfz durchgeführt, zu denen keine einschlägigen Ergebnisse bekannt sind. Es existierte weder ein Gebrauchsmuster noch ein Patent.

Die Problematik, mit der sich das Projekt RAINCARS auseinandersetzte, fokussierte sich auf die Ermittlung von Regenmengen auf Basis der Wischfrequenz unter Berücksichtigung der Fahrtgeschwindigkeit sowie der Neigung der Windschutzscheibe.

Das Projekt hatte herausragenden Wert in dem Erkennen der Problematik der nicht flächendeckenden Informationen und der Möglichkeit, dies durch mobile Sensoren auszugleichen. Es bot somit eine Reihe von Ansätzen, die das Projekt mobile^{VIEW} mit anderen Ansätzen weiterverfolgte und zu einer praktischen Anwendung bringen wollte. Die Integration von mobile^{VIEW} in bestehende Niederschlagsmessungen, kombiniert mit hydrologischen Modellen und Vorhersagen, sollte dabei die Relevanz der Genauigkeit der erfassten Niederschlagsmengen neu bewerten. Der von den Projektpartnern IAV, FiW und EGLV formulierte Lösungsansatz setzte dagegen nicht an den Endergebnissen des Projektes RAINCARS an, sondern fokussierte von Beginn an einen anderen Lösungsweg, um ausgewählte Daten aus Sensoren aus fahrenden Fahrzeugen nutzbar zu machen. Zum einen sollten Daten einer großen Anzahl von Fahrzeugen ausgewertet werden, zum anderen wurden zahlreiche Abgleiche zwischen den vorhandenen stationsgebundenen Niederschlagsdaten von EGLV und den Sensordaten durchgeführt. Dabei kamen Methoden zum Einsatz, die der besonderen Struktur und Charakteristik der gesammelten Daten gerecht werden. Es wurde davon ausgegangen, dass die gesammelten Daten den Kriterien für die Einordnung als Big-Data-Kollektiv entsprachen. Sie bildeten Datenmengen, die zu groß für traditionelle Datenbanksysteme sind, eine geringe Halbwertszeit besitzen und in ihrer Form nicht den Richtlinien herkömmlicher Datenbanksysteme entsprechen. Ein wesentlicher Bestandteil des Forschungsansatzes lag daher in der Aufbereitung der Daten. Dazu wurden entsprechende Ansätze aus dem Werkzeugkasten der Big-Data-Aufbereitung, bspw. statistische Modelle und künstliche Intelligenz, angewendet.

4.2 Windshield Wipers on connected Vehicles produce high-accuracy Rainfall Maps (Bartos et al. 2019)

In den USA wurde in den letzten Jahren ein Projekt zur Einbindung von mobilen Sensordaten in die Niederschlagsprognose durchgeführt. Am 17.01.2019, und somit während der Projektlaufzeit von mobile^{VIEW} haben Bartos et al. (2019) einen Online-Artikel in den Scientific Reports des Nature-Magazins veröffentlicht mit dem Titel „Windshield Wipers in connected Vehicles produce high-accuracy Rainfall Maps". Die Autoren befassten sich in der beschriebenen Studie mit der Kombination von Niederschlagsradarprodukten mit Scheibenwischerfrequenzen in einer Testflotte von 70 Fahrzeugen. Zielsetzung der Untersuchung war die Verdichtung von Niederschlagsinformationen in Bezug auf eine Verbesserung der Vorhersagen von Starkregen und Überschwemmungen. Die Messfahrten wurden durch Kameraaufnahmen der Windschutzscheiben begleitet, sodass mit dieser, in Deutschland womöglich datenschutzrechtlich kritischen, Datenquelle eine Validierung der tatsächlich auf den Boden auftreffenden Niederschläge durchgeführt werden konnte. Bartos et al. (2019) kamen zu dem Ergebnis, dass Scheibenwischerfrequenzen eine bessere binäre Information über den auf den Boden treffenden Niederschlag ermöglichen als Niederschlagsradare. Diese Erkenntnis wurde in den Überlegungen von mobile^{VIEW} unter Verweis auf Bartos et al. (2019) berücksichtigt. Zur Verschneidung

mehrerer Datenquellen nutzten Bartos et al. (2019) ein bayessches Verfahren. Auch dieser Ansatz wurde vom mobile^{VIEW}-Konsortium kritisch geprüft, um eine Eignung bei der eigenen Verschneidung von Datenquellen zu erwägen. Letztendlich wurde in mobile^{VIEW} aber ein eigenes Verfahren zur Verschneidung von Kfz-, Pluviometer- und Radardaten entwickelt.

4.3 Weitere Projekte

Eine Nutzung von Kfz-Sensordaten zum Zweck der Verbesserung von Starkregenvorhersagen oder anderen wasserwirtschaftlichen Nutzungen, vor allem im Bereich der Niederschlagsinformation, waren zum Projektstart nicht bekannt.

4.4 Patentlage

In Vorbereitung des Projektes wurde im März 2016 eine Patentrecherche durchgeführt. Dabei wurden u. a. die Dokumente DE 10 2011 002 802 A1 (Bosch) und DE 10 2012 004 998 A1 (Daimler-Benz) ermittelt, die beide die Erstellung von Wetterprognosen auf Basis von Wetterdaten von dezentral verteilten Informationsquellen (z. B. Kraftfahrzeuge) beschreiben. Die Idee an sich war damit bekannt und somit nicht mehr patentwürdig. Darüber hinaus war die Idee auch aus dem zitierten Projekt RAINCARS bekannt. Die daran beteiligten Institute oder Mitarbeiter haben erkennbar kein Schutzrecht zu den RAINCARS angemeldet.

Das Dokument DE 10 2012 004 998 A1 von Daimler-Benz war bereits erloschen und nicht mehr relevant. Das Schutzrecht DE 10 2011 002 802 A1 von Bosch war dem Antrag anhängig, zu diesem Zeitpunkt aber noch nicht erteilt. Der Einschätzung der beantragenden Partner nach war die Erteilungswahrscheinlichkeit eher gering, was sich jedoch erst im Prüfungsverfahren herausstellen konnte, das Bosch zum Projektstart noch nicht beantragt hatte. Außerdem bezogen sich die Patentansprüche, die den eigentlichen Schutzumfang angeben, auf stationäre Anlagen zur Energieerzeugung, aus deren Betriebsparametern Wetterinformationen erzeugt werden.

4.5 Verwendete Informationsdienste und Datenquellen

Im Rahmen der Projektbearbeitung griff das mobile^{VIEW}-Konsortium auf diverse Informationsdienste und Datenquellen zurück.

Für die Kombination der in mobile^{VIEW} erhobenen Kfz-Daten wurde als Radarprodukt das DX-Niederschlagsradarprodukt des DWD verwendet. Dieses wurde über das Open-Data-Angebot des DWD (https://opendata.dwd.de/) bereitgestellt und an dieser Stelle abgegriffen. Der Projektpartner EGLV nahm für die Verwendung in den eigenen hydrologischen Modellen eine Anpassung dieses Produktes vor, u. a. eine Dämpfungskorrektur. Das daraus entstehende aufbereitete Produkt, bezeichnet als dämpfungskorrigiertes DX-Produkt, wurde von EGLV in mobile^{VIEW} zur Verfügung gestellt.

Weiterhin brachte EGLV die Daten aus dem verbandseigenen terrestrischen Niederschlagsmessnetz mit in das Projekt ein.

5 BISHERIGE ARBEITEN DES KONSORTIUMS

5.1 Forschungsinstitut für Wasser- und Abfallwirtschaft an der RWTH Aachen (FiW) e. V.

Das Forschungsinstitut für Wasser- und Abfallwirtschaft an der RWTH Aachen (FiW) e.V. wurde 1979 als unabhängiges An-Institut an der Rheinisch-Westfälisch Technischen Hochschule (RWTH) Aachen gegründet. Die Arbeiten mit ca. 50 Mitarbeitern konzentrieren sich auf themenübergreifende Forschungsaktivitäten im Spannungsfeld von Wasserwirtschaft, Infrastruktur, Energie und Klima. Das FiW betreibt anwendungsorientierte Forschung und Entwicklung inkl. Wissenstransfer und ist regelmäßig als Mitinitiator und Gesamtkoordinator an Verbundforschungsprojekten beteiligt. Das überwiegend aus Drittmitteln finanzierte FiW ist Gründungsmitglied der Johannes-Rau-Forschungsgemeinschaft NRW und Mitglied der Deutschen Industrieforschungsgemeinschaft Konrad Zuse e. V.

Mit Bezug zu dem aktuellen Vorhaben kann das FiW Erfahrungen in der Durchführung komplexer Verbundvorhaben (z. B. DynAKlim, RiverView®) ebenso aufweisen wie fachliche Erfahrungen im Umgang mit Radardaten und deren Echtzeitverwendung (z. B. Projekt „Reduktion der Gewässerbelastung durch Steuerung des Kanalnetzes"), Oberflächendaten und Oberflächenabflusssimulationen zur Erfassung von Starkregeneinflüssen (z. B. DynAKlim) und der Erfassung und Auswertung von Massendaten durch schwimmfähige Drohnen (z. B. RiverView®).

Im Rahmen des Netzwerk- und Verbundvorhabens DynAKlim wurde im Zeitraum von 2009 bis 2015 die dynamische Anpassung regionaler Planungs- und Entwicklungsprozesse an die Auswirkungen des Klimawandels in der Emscher-Lippe-Region (Ruhrgebiet) untersucht. Das FiW übernahm dabei neben der Gesamtkoordination des Vorhabens auch das Teilprojekt „Potenziale einer anpassungsfähigen Siedlungswasserwirtschaft", in dem an drei Pilotstandorten die Umsetzung einer wassersensiblen Stadtentwicklung durchgeführt wurde. Dabei lag der Schwerpunkt in der integralen Betrachtung des Entwässerungssystems und möglicher Transport- und Speicherplätze für Wasser auf der Oberfläche. Dabei spielte insbesondere die multifunktionale Flächennutzung von Verkehrsflächen zur kurzfristigen Speicherung und Abführung von Regenwasser eine große Rolle. Daneben war auch die Niederschlagsverteilung in den Untersuchungsgebieten relevant.

Das vom BMBF geförderte Verbundvorhaben RiverView® befasste sich mit der Erfassung, Auswertung und Visualisierung umfangreicher Gewässerdaten über und unter Wasser. Dazu kam ein autonomes Messfahrzeug zum Einsatz, das mit einer Vielzahl von Messgeräten ausgestattet ist. So wurden Bildinformationen über Wasser kombiniert mit Aufnahmen der Gewässersohle und Wasserqualitätsdaten. Das FiW übernahm die Gesamtkoordination des von 2015 bis 2020 laufenden Vorhabens sowie wesentlich die Aufnahme und die Auswertung der raumzeitlichen Gewässerdaten, wobei die erfasste Rohdatenmenge eine Größenordnung von 1 TB pro Tag erreichte.

Im vom MULNV NRW geförderten Projekt „Reduktion der Gewässerbelastung durch Steuerung des Kanalnetzes", das von 2013 bis 2016 durchgeführt wurde, stand die Erprobung eines Online-Kanalnetzsteuerungs-Werkzeugs für die Reduktion der Gewässerbelastung im Vordergrund. Hierbei wurden Radardaten des DWD und hydrodynamische Modelle für eine Niederschlagsprognose und eine Echtzeitsteuerung von bis zu zehn Regenbecken verwendet. Die Aufgaben des FiW lagen in der Koordination, in der Datenauswertung und in der Qualitätssicherung der Berechnungsergebnisse und der Steuerung.

5.2 Emschergenossenschaft und Lippeverband

Die Emschergenossenschaft wurde am 14. Dezember 1899 als erster deutscher sondergesetzlicher Wasserwirtschaftsverband gegründet. Dieses Modell stand Pate für eine ganze Reihe weiterer Wasserverbände, darunter für den am 18. Januar 1926 gegründeten Lippeverband. Emschergenossenschaft und Lippeverband haben von Anfang an eng kooperiert. Ihre rund 1.600 Mitarbeiter arbeiten unter dem Dach einer einheitlichen Organisationsstruktur als NON-PROFIT-Unternehmen in den ca. 4.200 km² großen Einzugsgebieten der Emscher und der Lippe zusammen, in denen mehr als 4 Mio. Einwohner leben.

Die unterschiedlichen Nutzungsansprüche, die Wirtschaft, Mensch und Natur an die Emscher haben, in ein nachhaltig vertretbares Gleichgewicht zu bringen – das ist im Wesentlichen die Aufgabe der Emschergenossenschaft als Flussmanager der Emscher. Seit mehr als 100 Jahren ist die Emschergenossenschaft für alle Belange rund um die Emscher und ihre Nebenläufe verantwortlich. Dazu gehören die Gewässerunterhaltung, der Hochwasserschutz, die Abwasserbeseitigung und das Regen- und Grundwassermanagement. Ebenso die naturnahe Umgestaltung der Emscher und die allgemeine Planung, Beratung und Durchführung aller Maßnahmen, die für den Emscher-Umbau erforderlich sind. Dabei werden laufend Innovationen maßgeblich von der Emschergenossenschaft mit entwickelt – als Beispiele seien hier die automatischen Inspektions- und Reinigungssysteme für den Abwasserkanal Emscher angeführt, oder die Koppelung moderner Ansätze wie die Radarerfassung flächenhafter Niederschläge und deren Nutzung in der operativen Online-Vorhersage von Abflüssen im Emscher- und Lippegebiet.

Mit Bezug zu dem aktuellen Vorhaben kann EGLV ebenso Erfahrungen in der Durchführung komplexer Verbundvorhaben (z. B. DynAKlim, Future Cities) aufweisen. Die fachlichen Erfahrungen von EGLV im Umgang mit Radardaten und deren Echtzeitverwendung (z. B. Reduktion der Gewässerbelastung durch Steuerung von Regenrückhaltebecken) gehen bereits in die 1990er Jahr zurück, in denen im Rahmen des F & E-Projektes „Abflusssteuerung unter Verwendung radargemessener Niederschläge“ ein eigenes X-Band Radargerät aufgebaut und betrieben sowie diese Niederschlagsflächeninformationen in Abflusssimulationen eingesetzt wurden. Heute ist im Technischen Hochwassermanagement bei EGLV ein operationelles Hochwasserinformationssystem Stand der Technik, das auf den C-Band-Radardaten des DWD im Rahmen einer Kooperation der Wasserverbände NRW mit dem DWD zur Katastrophenvorsorge basiert. Seit mehr als fünf Jahren bündelt dieses Informationssystem online Niederschlagsmessungen und meteorologische Vorhersagen und bindet diese in die Niederschlag-Abfluss-Modellierung ein. Damit dient es den internen Betriebsabteilungen zur Erfüllung ihrer Aufgaben im Hochwassermanagement.

Die Erfahrung zeigt, dass auch der kommunalen Daseinsvorsorge gegenüber Schäden aus lokalen Überflutungen eine immer größere Bedeutung zukommt. Dies haben nicht nur die Ereignisse im Juni 2016 am Niederrhein (NRW) und in Braunsbach im Mai 2016 (Baden-Württemberg) oder auch im Juli 2014 in Münster (NRW) plakativ vor Augen geführt. Die Beteiligung von EGLV im aktuellen Vorhaben stellt eine intensive Einbeziehung kommunaler und regionaler Anliegen sicher. Darüber hinaus ist aufgrund von umfangreichen Erfahrungen aus europäischen Forschungsprojekten (z. B. Future Cities mit „Stark gegen Starkregen“ oder SIC Adapt) zur Anpassung an den Klimawandel und zum Umgang mit Starkregen ein vielschichtiger Blick auf das Thema sichergestellt.

5.3 IAV GmbH Ingenieurgesellschaft Auto und Verkehr

Die IAV GmbH Ingenieurgesellschaft Auto und Verkehr ist mit mehr als 6.500 Mitarbeitern einer der weltweit führenden Engineering-Partner der Automobilindustrie. Das Unternehmen entwickelt seit über 30 Jahren innovative Konzepte und Technologien für zukünftige Fahrzeuge. Die Kernkompetenzen liegen in serientauglichen Lösungen in allen Bereichen der Elektronik-, Antriebsstrang- und Fahrzeugentwicklung. Zu den Kunden des Unternehmens zählen weltweit alle namhaften Automobilhersteller und Zulieferer.

Der Transfer von Technologien, Prozessen und Prozeduren zur Entwicklung innovativer Produkte, Verfahren und Dienstleistungen ist Teil der Zukunftsstrategie von IAV. Kompetenzen und Erfahrungen aus dem Automotive Engineering werden daher zunehmend auch in anderen Geschäftsfeldern, wie z. B. der Energiewirtschaft oder der Wasserwirtschaft, erfolgreich eingesetzt. Hierbei sind es insbesondere die Themenbereiche Monitoring, Smart Metering, Steuerung und Regelung, Predictive Maintenance, sowie die Technologisch-Ökonomische Optimierung von technischen Anlagen, die IAV im Portfolio hat.

Speziell für die Aspekte des vorliegenden Forschungsvorhabens verfügt IAV über langjährige Erfahrung in der Erfassung, Speicherung, Transferierung, Auswertung und Verwertung von Daten aus Kfz-Sensoren. Die Entwicklung und Nutzung von Datenplattformen ist ebenfalls ein wichtiger Baustein in diesem Zusammenhang.

Wir denken in Systemen – egal ob bei der Fahrzeugentwicklung oder im Bereich der Wasserwirtschaft. Und wir setzen auf Synergien: Wir wollen unsere Erfahrungen aus der Automobilentwicklung auch für die Themen „Digitalisierung“ und „Automatisierung“ nutzen. Es gibt viele Parallelen zwischen beiden Bereichen: Die Regelungstechnik und das Datensatzmanagement unterscheiden sich bei Fahrzeugen und wasserwirtschaftlichen Anlagen kaum voneinander.

In zahlreichen Forschungsvorhaben mit den unterschiedlichsten Forschungsträgern und in unterschiedlichen Themenbereichen hat IAV überdies seine Fähigkeit nachgewiesen, Forschungsprojekte technisch und administrativ qualifiziert und mit Erfolg auszuführen.

6 ERGEBNISSE

6.1 Detaillierte Ergebnisdarstellung

Zu Beginn jedes Teilprojektes wurde eine strukturierte Planung der zu bearbeitenden Aufgaben vorgenommen. Dies wurde in den Projektstatustreffen und Videokonferenzen, die das Projekt im regelmäßigen Turnus begleiteten, gewährleistet.

6.1.1 Teilprojekt TP 1: Datensammlung

TA 1.2 Hardware für Aufnahme, Speicherung und Versand der Kfz-Daten

Im ersten Arbeitspaket des Teilprojektes wurde das für die Datenaufnahme, -speicherung und -übertragung notwendige Hardwarekonzept entwickelt. In Vorbereitung der Entwicklung des Hardware- und dazugehörigen Softwarekonzeptes durch die IAV regte das FiW einen wissenschaftlichen Austausch mit dem mFUND-Projekt FlottenWetterKarte (FloWKar, FKZ 19F2045A-B) an, mit dem während der Projektlaufzeit immer wieder Zwischenstände und Zwischenergebnisse ausgetauscht wurden. Auch aus anderen Forschungsprojekten konnten Erkenntnisse berücksichtigt werden.

Auswahl der Fahrzeugflotte

Um Fahrzeuge ausstatten zu können, musste zunächst eine Auswahl der bei EGLV verfügbaren Fahrzeuge vorgenommen werden. Hierfür holte EGLV in Rücksprache mit dem Fahrzeugmanagement eine Liste aller verfügbaren Dienst-, Betriebs- und Poolfahrzeuge ein. Fahrzeugfabrikate, die nicht zur Volkswagen AG gehörten, mussten nach Rücksprache mit IAV aufgrund von abweichenden Codierungen der Datenströme im Fahrzeug für das Forschungsvorhaben ausgeschlossen werden. Da es sich hierbei zudem größtenteils um auf drei Jahre geleaste Fahrzeuge handelt, mussten diese

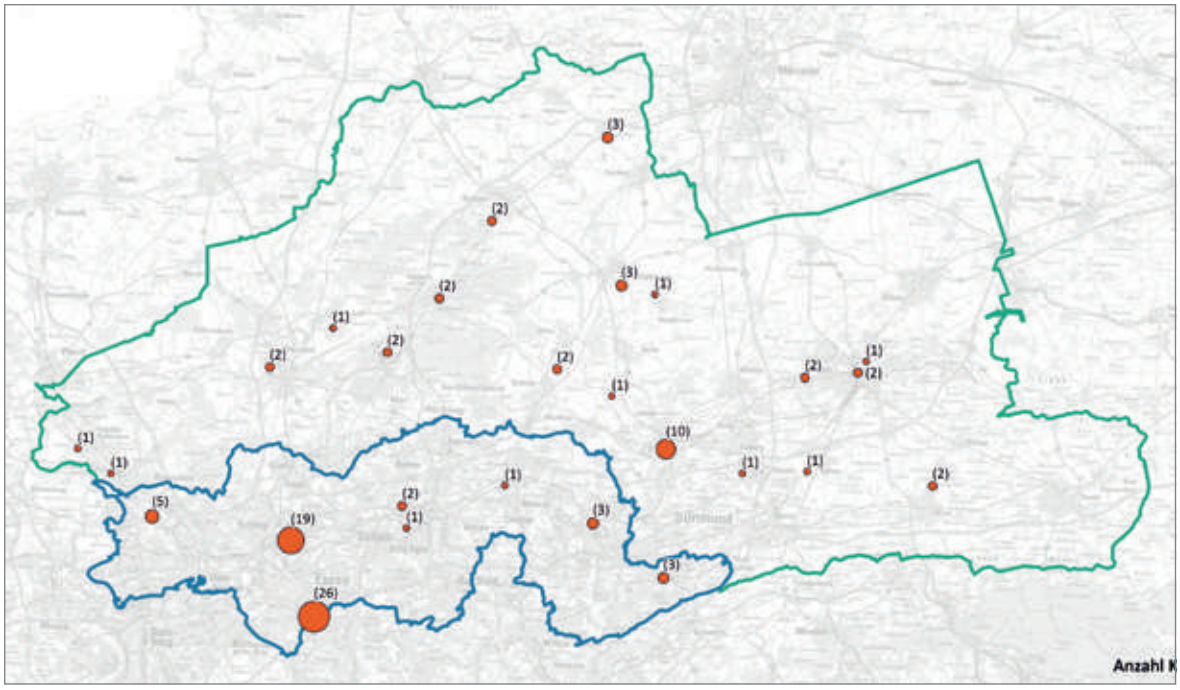


Abbildung 1: Standorte der im Rahmen des Forschungsvorhabens ausgerüsteten Fahrzeuge im Verbandsgebiet von EGLV. © EGLV

zunächst vor dem Hintergrund der organisatorischen Rahmenbedingungen bewertet werden, sodass ein für das Forschungsvorhaben unverhältnismäßiger und zeitlich ungünstiger Fahrzeugwechsel relativ kurz nach Einbau vermieden werden konnte. Über die bereitgesellten Fahrgestellnummern aller in Frage kommenden Kraftfahrzeuge war IAV zudem in der Lage, fahrzeugindividuelle Ausstattungsmerkmale abzurufen und somit sowohl jene Fahrzeuge zu identifizieren, die sich grundsätzlich eigneten, jedoch über keine native Regen-Licht-Sensorik verfügten und somit die Nachrüstung eines nachträglichen Sensors erforderten. Weitere Dezimierungen des Fahrzeugbestandes ergaben sich aus Fahrzeugen mit einem Baujahr vor 2015 aufgrund abweichender Informationscodierungen. Zudem war eine repräsentative Abdeckung des Untersuchungsgebietes hinsichtlich städtischer Zentren und eher ländlicher Regionen ebenfalls Gegenstand der Überlegungen zur Fahrzeugauswahl. Die Standorte der Fahrzeuge sind in Abbildung 1 dargestellt.

Hard- und Softwarekonzept für die Sensorik zur Datenaufnahme und -übertragung

Für die Entwicklung des Hard- und Softwarekonzeptes wurde ein Anforderungskatalog definiert, der sich zunächst an den Anforderungen der Wasserwirtschaft an die zu erhebenden Daten orientierte. Bezüglich der zu erhebenden Rohdaten wurden vor allem das GPS-Signal (Ort- und Zeitstempel), der Regensensor und die Scheibenwischerfrequenz als wichtige Parameter identifiziert. Doch auch weitere Sensordaten mit Bezug zu wasserwirtschaftlichen Fragestellungen sollten mit aufgenommen werden. Im Rahmen der technischen Umsetzbarkeit wurde die Datenübertragung nur für fahrende Fahrzeuge vereinbart. Das Anbringen einer Back-Up-Batterie wäre zwar technisch möglich gewesen, durch das ohne eingeschaltete Zündung fehlende automatisierte Überwischen des Regensensors auf der Windschutzscheibe hätte es jedoch keine validen Mehrinformationen gegeben. Zudem hätten ein überdachter Abstellort (Garage, Parkhaus, Baumallee) oder eine aktive Datenübertragung in einer Waschanlage zu fehlerhaften Messwerten geführt.

Im Gegensatz zu den üblichen Entwicklungsfahrzeugen von IAV handelte es sich bei der Flotte von EGLV nicht um Kfz mit einer Prototyp-Zulassung, welche es den Original Equipment Manufacturer (OEM) erlaubt, weitreichende Änderungen an den Fahrzeugen durchzuführen und sie trotzdem im Straßenverkehr führen zu dürfen. Vielmehr bestand die EGLV-Fahrzeugflotte aus regulären, straßenzugelassenen Fahrzeugen, bei denen etwaige invasive Eingriffe, wie die Abisolierung eines Kabels und das Anbringen einer T-Steckverbindung zum Mitschneiden gewisser Datenströme im Fahrzeug, zu einer Erlöschung der Allgemeinen Betriebserlaubnis geführt hätte. Die hierbei für das Forschungsvorhaben vorliegenden Informationen im Kfz lagen darüber hinaus nicht an den OBD2-Schnittstelle vor, was einen barrierefreien Abgriff ermöglicht hätte, sondern am Komfort-CAN. Vor diesem Hintergrund wurden verschieden technische Lösungen ausgiebig untersucht. Dabei durften weder die Sicht des Fahrers behindert werden noch ein invasiver Eingriff in das elektrische System erfolgen.

Die vorgesehene Datenübertragung über die Bluetooth-Schnittstelle des Fahrzeugs und ein verbundenes Mobiltelefon wurde nicht weiterverfolgt. Dies ist zum einen durch die dadurch resultierende Blockade der Bluetooth-Schnittstelle für den Fahrzeugführer als auch durch die gewählte Kfz-Sensorik begründet. Nicht alle Kfz der EGLV Fahrzeugflotte waren standardmäßig mit einem Regensensor ausgestattet. So wurden die entsprechenden Fahrzeuge mit einem nachträglich verbauten Regensensor ausgerüstet. Bei einem Abgriff der Daten über die Bluetooth-Schnittstelle hätten die Daten des extern eingerichteten Regensensors nicht abgegriffen werden können.

Als Ergebnis kam ein induktiver CAN-Reader zum Einsatz, welcher ein Mitlesen der Datenströme ermöglichte, ohne die Kabelisolierungen in den Fahrzeugen zu verletzen. Zudem war es auf diese Weise möglich, erhebliche Kosten einzusparen, da die Fahrzeuge zum Leasingende hin in ihren Auslieferungszustand hätten rückgeführt werden müssen, was somit noch einmal die Notwendigkeit der reversiblen Einbringung von Hardware-Komponenten unterstreicht.

Die 100 ausgewählten Fahrzeuge waren in vier Plattformen unterteilt (MQBA1, PQ26, PQ35 und MLBevo). Je nach Plattform waren unterschiedliche serienmäßige Sensoren (Regen-Licht-Sensoren – RLS) verbaut, die jeweils auch unterschiedliche Signalkodierungen verwendet haben. Daraus ergab sich die Notwendigkeit der Konzeptionierung und Programmierung mehrerer Softwarekonfigurationen. Für einen Teil der EGLV-Flotte, die über keinen eigenen RLS verfügte, stellte die Audi AG im Rahmen des wissenschaftlichen Austausches zwischen den Projekten FloWKar und mobile^{VIEW} kostenneutral RLS zur Verfügung.

Jedes Fahrzeug wurde somit mit einer Rechen- und Übertragungseinheit ausgestattet, an die die folgenden Anforderungen gestellt wurden:

- Fahrzeugsignale lesbar entsprechend der Spezifikationen der Messkonfiguration
- Externe Sensoren lesbar entsprechend der Spezifikationen der Messkonfiguration

- Echtzeitübertragung von Messdaten
- Temporäre Speicherung von Messdaten bei fehlerhafter Übertragung
- Herunterladen neuer Messkonfigurationen möglich
- Flexible Schnittstellen
- Unterstützung Arduino/RasberryPI-basierter Systeme

Um diese Anforderungen zu erfüllen wurde ein Übertragungssystem (Mobile Device) der Firma Axotec (IGX560) ausgewählt. Die Messkette ist in Abbildung 2 beschrieben. Um die Übertragung über Mobilfunk zu testen, wurden vor dem Einsatz der technisch hochwertigen Datensammelboxen 20 GPS-Tracker mit SIM-Karten ausgestattet und in den Fahrzeugen angebracht.

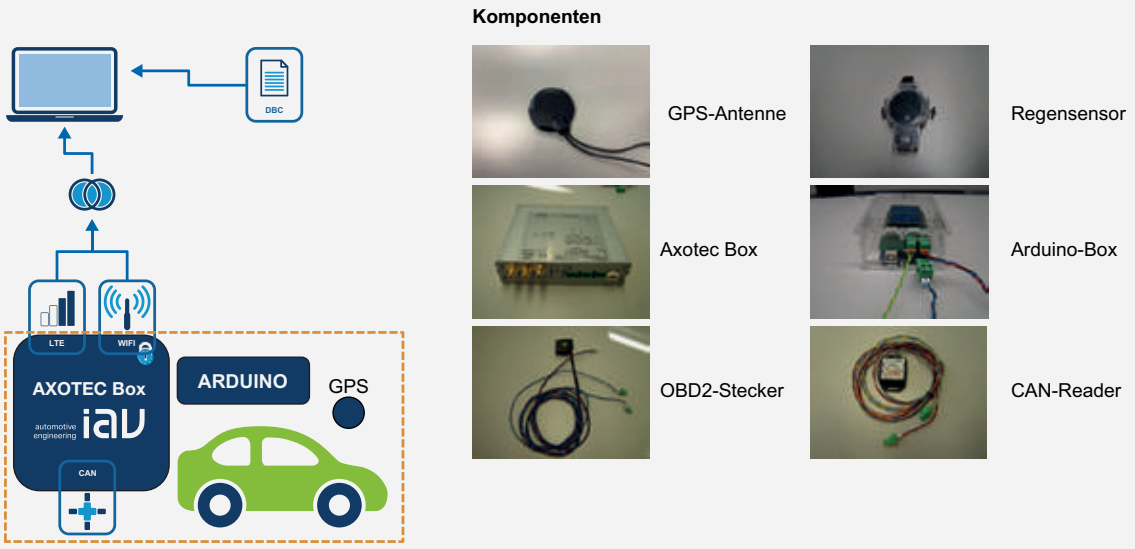
Die Software für die Datenaufnahme aus Fahrzeugen wurde in der Programmiersprache Python geschrieben und ausführlich getestet (Abbildung 3). Die Anforderungen an die Software auf Mobile Devices waren:

- Möglichkeit eines externen Zugriffs über das Internet auf die Konfigurationsoberfläche (SPA) entweder durch feste IP-Vergabe vom Provider oder Tunnel-Integration
- Login erforderlich bei Zugriff auf Konfigurationsoberfläche
- Automatisches Starten des Sendens der MQTT-Nachrichten bei Kl-15 ein
- Herunterfahren bei Kl-15 aus
- Integration von Hardware- und Software-Watchdogs (Automatisches Rebooten und Erkennen von Fehlern)
- MQTT-Nachrichten im JSON-Format mit Inhalt aus zuvor definierter Messdatenliste
- Deaktivierung der Sendemöglichkeit von CAN-Nachrichten zur Umsetzung der Sicherheitsanforderungen (Read-Only Modus)
- Verschlüsselte Kommunikation (TLS)

Die Softwarekonfigurationen benötigte eine Entschlüsselungsdatei, um die Fahrzeugsignale interpretieren zu können. Über die Zusammenarbeit mit dem Projekt FloW-Kar wurden die nötigen Entschlüsselungsdateien von der Audi AG an IAV übergeben.

Sicherstellung einer DSGVO-konformen Datenverarbeitung

Neben den fachlichen Anforderungen musste die Datenerhebung zudem auch die Vorgaben der DSGVO zur Verarbeitung personenbezogener Daten erfüllen, da die Datenerhebung von Zeit- und Ortstempel der jeweiligen Fahrzeuge einer Verarbeitung von personenbezogenen Daten entspricht. Unter anderem musste das Konsortium die Rechtmäßigkeit der Datenverarbeitung gemäß Art. 6 DSGVO belegen.



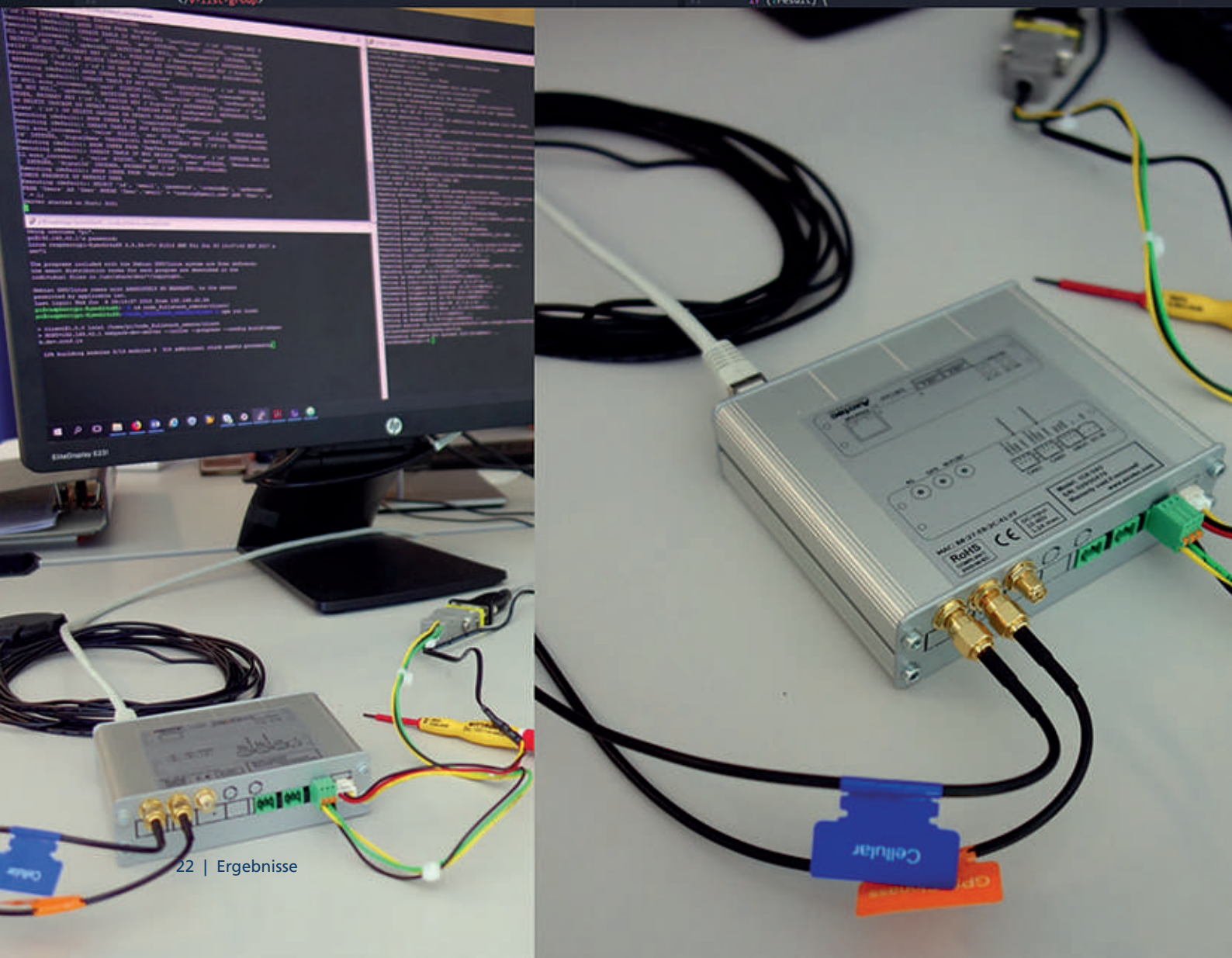
Messkette MK1: GPS + vorhandener Regensensor + IGX560 = Übertragung an Server
Messkette MK2: GPS + nachgerüsteter Regensensor + IGX560 + Arduino-Box = Übertragung an Server

Abbildung 2: Messkette der Übertragungstechnik mit einer Datenbox der Firma Axotec. © IAV

```
<template>
  <div id="app">
    <v-app id="inspire">
      <v-navigation-drawer
      fixed
      v-model="drawer"
      app>
        <v-list dense>
          <v-list-tile v-for="item in items" :key="item.title"  @click="goTo(item)">
            <v-icon>{{ item.icon }}</v-icon>
          <v-list-tile-action>
            <v-list-tile-content>
              <!-- <v-list-tile-title :style="`${(!item.icon) ? 'margin-left: 1.4em;' : ''}`">{{
              <v-list-tile-title>
                {{ item.title }}
                  <!-- <v-icon>expand_more</v-icon></v-btn> -->
              </v-list-tile-title>
            </v-list-tile-content>
          </v-list-tile>
          <v-list-group
            v-model="item.active"
            v-for="item in expandableItems"
            :key="item.title"
            :prepend-icon="item.icon"
            no-action
          >
            <v-list-tile slot="activator">
              <v-list-tile-content>
                <v-list-tile-title>{{ item.title }}</v-list-tile-title>
              </v-list-tile-content>
            </v-list-tile>
            <v-list-tile
              v-for="subitem in item.subitems"
              :key="subitem.title"
              :to="{
                  name: 'dataLogging',
                  params: {
                    canParamsId: subitem.config.id,
                    canParamsName: subitem.config.name
                  }
                }">
            <!-- <v-list-tile v-for="subitem in item.subitems" :key="subitem.title" @click="goTo(
              <v-list-tile-content class="subitem">
                <v-list-tile-title>{{ subitem.title }}</v-list-tile-title>
              </v-list-tile-content>
              <v-list-tile-action>
                <v-icon>{{ subitem.icon}}</v-icon>
              </v-list-tile-action>
            </v-list-tile>
          </v-list-group>
```

```
// const express = require('express')
const bodyParser = require('body-parser')
const cors = require('cors')
const morgan = require('morgan')
const config = require('./config/config')
const {sequelize, User} = require('./models')

const app = require('express')()
const http = require('http').Server(app)
const io = require('socket.io')(http)
const fs = require('fs')
const exec = require('sync-exec')
const users = require('../seed/users.json')

app.use(morgan('combined')) // Zeigt Logs zu den Eigenschaften der clients - Browser der
app.use(bodyParser.json())
app.use(cors()) // erreichbar von außen

require('./routes')(app)
require('./sockets')(io)
// sequelize.sync({force: true}) // drop all

// Create necessary directories
for (storage in config.files) {
  // try to open directory for read and write
  let storageDirectory = config.files[storage]()
  try {
    fs.accessSync(storageDirectory, fs.constants.R_OK | fs.constants.W_OK)
  } catch (err) {
    console.log(`no r/w access to ${storageDirectory}`)
    console.log(`trying to create ${storageDirectory} ...`)
    let result = exec(`mkdir -p ${storageDirectory}`)
    if (!result.status) continue
    console.error(`failed to create ${storageDirectory}`)
    console.error(result)
    process.exit(1)
  }
}

sequelize.sync({force: false})
  .then(async function () {
    // Check if default user exists
    console.log('CHECK PRESENCE OF DEFAULT USER')
    let result = await User.findOne({
      where: {
        email: 'testing@gmail.com',
        id: 1
      }
    })
    // create default User if not already existent
    if (!result) {
```

Abbildung 3: Embedded Software Test Station. © IAV



Zu diesem Zwecke wurde innerhalb des Konsortiums durch die Konsortialpartner eine grundsätzliche datenschutzrechtliche Verpflichtungserklärung ausgearbeitet, mit dem Zweck, für alle Betroffenen detailliert darzulegen, welche Daten wo zu welchem Zweck verarbeitet werden. Zudem musste in der Frage des Datenschutzes im Arbeitnehmer-Kontext auch der Personalrat von EGLV beteiligt werden, dem das ausgearbeitete datenschutzrechtliche Konzept zur Abstimmung vorgelegt wurde. Unter Einbeziehung der Datenschutzbeauftragten von EGLV wurde auf dieser Basis von EGLV eine datenschutzrechtliche Einwilligungserklärung entworfen, die allen betroffenen Fahrzeugführern – die meisten hiervon Mitarbeiter des Betriebes auf den dezentralen wasserwirtschaftlichen Anlagen der Verbände – nach einer persönlichen Vorstellung des Projektes zur Unterschrift vorgelegt wurde. Grundsätzlich bestand jederzeit die Möglichkeit einer Rücksprache im Hinblick auf datenschutzrechtliche Fragen.

Ebenfalls Unterstützung seitens der Datenschutzbeauftragten von EGLV bekam das Konsortium bei der Dokumentation des Verzeichnisses von Verarbeitungstätigkeiten gemäß EU-DSGVO Art. 30, Abs. 1. Zuletzt wurde die Entwicklung des sog. „Privacy Switches“, einer technischen Maßnahme zur temporären Deaktivierung der Datenübertragung, begleitet und anschließend auf alle betroffenen Kraftfahrzeuge verteilt. Im Falle von Betriebs- und Dienstfahrzeugen mit einem überschaubaren Kreis an potenziellen Fahrzeugführern, denen allesamt eine datenschutzrechtliche Einwilligungserklärung vorgelegt wurde, wurde der Privacy Switch im „opt-out“-Modus realisiert, um eine größtmögliche Menge an Messdaten für das Projekt zu gewinnen. In dieser Einstellung findet eine standardmäßige Erhebung von GPS- und Messdaten statt, welche vom Fahrzeugführer manuell deaktiviert werden muss. Im Falle von drei Pool-Kfz, bei denen sich der Kreis möglicher Fahrzeugführer auf alle knapp 1.600 Beschäftigten von EGLV erstreckt, war der hohe Aufwand des Einholens von Einwilligungen jedoch nicht zu rechtfertigen, sodass hier abweichend der „opt-in“-Modus umgesetzt wurde, bei dem eine explizite Einwilligung der betroffenen Fahrzeugführer in die Datenaufzeichnung vor Beginn einer jeden Fahrt notwendig ist. Über QR-Codes aufrufbare Links wurden auf alle im Forschungsvorhaben gebundenen Fahrzeuge in Form von Aufklebern (Abbildung 4) sowie Fahrtenbuch-Einlagen im Format DIN A5 verteilt.

Zum Betätigen der Schalter wurde für alle dienstlichen Smartphones der Mitarbeiter von EGLV eine entspreche QR-Code-Reader-App seitens der IT-Abteilung bereitgestellt. Bei der Aktivierung eines „Privacy Switch“ wurde die Information dem Server kommuniziert. Alle eingehenden Informationen aus den spezifischen Fahrzeugen wurden für die angegebene Dauer unwiderruflich entfernt (s. Abbildung 5).

Abbildung 4:
Muster eines im Fahrzeug angebrachten „Privacy Switches“ zur temporären Unterbrechung der Datenaufzeichnung.
© EGLV

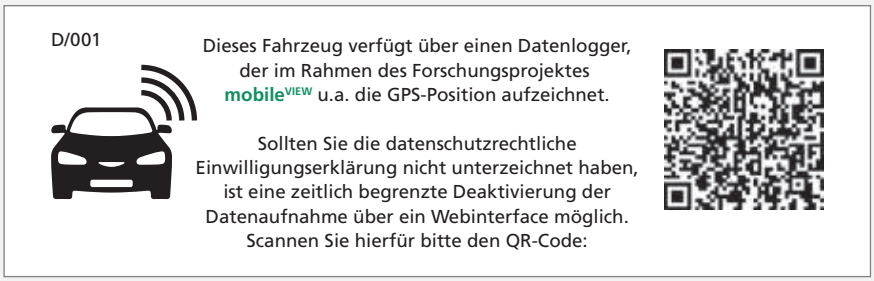
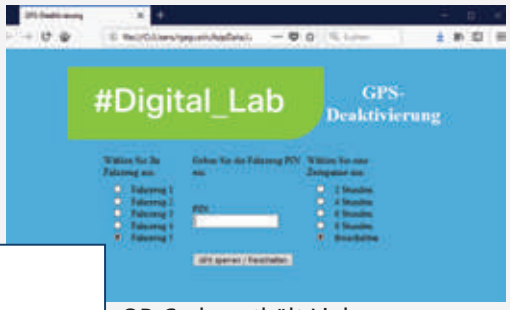


Abbildung 5:
Web-Interface bei Aktivierung des „Privacy Switch“ über ein Smartphone. © IAV



Fahrzeug E-X 4711

Dieses Fahrzeug nimmt Daten auf.

Eine Deaktivierung ist über ein Webinterface möglich.
Folgen sie bitte dem QR-Code

Fahrzeug-PIN: 0815

QR-Code enthält Link zum Webinterface

Darüber hinaus wurde innerhalb des Konsortiums ein Geofencing-Konzept vereinbart, nach dem die aufgezeichneten Messungen verworfen wurden, wenn sich Fahrzeuge über eine definierte Grenze hinaus entfernten. Dies geschah vor dem Hintergrund des Konzeptes der Datensparsamkeit im Sinne der DSGVO.

Für die Datenübertragung vergab das FiW einen Unterauftrag an einen Mobilfunkanbieter, bei dem die Upload-Kapazitäten, die im Rahmen der Anforderungsdefinition festgelegt wurden, für die große Datenmenge ausreichend hoch waren.

Im ersten Arbeitsschritt wurde somit die nach den o.a. Überlegungen konzipierte, programmierte und angefertigte Soft- und Hardware vom FiW beschafft und von IAV in Testläufen implementiert, in Testfahrzeugen verschiedener Plattformen verbaut und die Funktionsumfänge wurden in Versuchsfahrten validiert. Weiterhin wurde der Datenschutz gemäß DSGVO bei der Datenverarbeitung sichergestellt.

Für den Rückbau der Sensorik und zur Rückführung in den Auslieferungszustand wurde mit der Werkstatt eine entsprechende Regelung getroffen, die für die ausgewählten Fahrzeuge zur Anwendung kommt.

TA 1.3 Software für den Sensorbetrieb und die Datenübertragung

Nach Anschaffung der Hardware zur Datenübertragung wurde im nächsten Arbeitspaket die Software für den Sensorbetrieb und die Datenübertragung ausgewählt. Eine erste Definition der Anforderungen zu Datenqualität, Prüfung, Speicherung und Versand von Daten wurde beim IT-Workshop Anfang 2019 in Gifhorn in einem Lastenheft festgehalten. Als Übertragungsmedium wurde eine Datensammelbox (DSB) der Firma Axotec GmbH ausgewählt.

Jedes Fahrzeug verfügte nach der Einbauphase über drei bzw. vier Informationsquellen: die Axotec-Box, den GPS-Tracker, das Fahrzeug selbst und den nachgebauten RLS (nur in Fahrzeugen, die nicht mit einem Seriensensor ausgestattet waren). Vor Beginn der ersten Messfahrten wurde keine Bewertung der späteren Praxistauglichkeit einzelner Sensoren vorgenommen, um keine später ggf. relevanten Informationen im Vorfeld auszuschließen.

Um die Qualität der über die Box übertragenen Daten bewerten zu können, wurde als geeigneter Parameter das GPS-Signal identifiziert. Dieses ermöglichte einen Vergleich der ermittelten Bewegungsprofile, die durch die GPS-Tracker aufgezeichnet wurden. Die über die DSB übertragenen GPS-Daten wurden auf diese Weise als plausibel bewertet. Die aufgezeichneten Bewegungsprofile flossen u. a. in die Auswahl zur Ausstattung weiterer Fahrzeuge ein, um ein möglichst großes Gebiet abdecken zu können.

Definition und Erfassung relevanter Sensordaten zur Qualitätssicherung

Für die Zusammenführung der Sensordaten und gleichzeitigen gegenseitigen Überprüfung war es vorteilhaft, dass sich die Information aus den o. a. unterschiedlichen Quellen ergänzten und teilweise doppelt vorlagen. So verfügte z. B. die Axotec-Box über eine interne Zeitmessung, deren Genauigkeit jedoch je nach Box variierte und im Worst-Case einen maximalen Offset von bis zu 60 Sekunden gegenüber der GPS-Zeit aufwies. Deshalb wurde für die Datenablage die GPS-Zeit bevorzugt. Der GPS-Tracker lieferte neben Breiten- und Längengrad sehr genaue Zeitangaben.

Fahrzeuge nutzten zur Datenübertragung zwischen mehreren Steuergeräten ein CAN-Bus-System[1]. Die relevanten Fahrzeugdaten im Projekt wurden über den sogenannten Komfort-CAN übertragen. Je nach Plattform befanden sich auf dem Komfort-CAN zwischen 4.000 und 6.000 Signale. Das Auswahlverfahren für relevante Daten fokussierte sich darauf, nur Signale zu selektieren, die einen direkten oder indirekten Bezug zur Typisierung „Niederschlag" hatten. Als „direkt" wurden dabei z. B. Niederschlagssignale aus den Serien-RLS typisiert, wobei Informationen aus den nachgebauten RLS fehlende Serien-RLS ersetzten. Als „indirekt" wurden z. B. Außentemperatur oder Lichtsensorsignale bewertet.

Die gesamte Liste ausgewählter Signale ist in Tabelle 1 zusammengefasst.

1 CAN- Bus = Controller-Area-Network. Ein serieller Datenbus, der gleichberechtigte Stationen miteinander verbindet. CAN wird zur Vernetzung von Steuergeräten eingesetzt. Einfach ausgedrückt können die angeschlossenen Steuergeräte über die Leitungen „kommunizieren" und gegenseitig Informationen austauschen.

Tabelle 1: Auflistung der ausgelesenen Fahrzeugsignale.

Box und GPS-Tracker	
name	Box-Bezeichnung
timestamp	Time aus der Axotec Box
longitude	GPS-Longitude
latitude	GPS-Latitude
readableDate	timestamp in UTC-Format
Fahrzeugsignale	
ESPLatBesc	Beschleunigungswert in y-Achse
ESPLonBesch	Beschleunigungswert in x-Achse
ESP	Winkelgeschwindigkeit um die z-Achse
ESPGesch	Geschwindigkeitssignal vom ABS/ASR oder ESP
RadRPS	Radumdrehungen pro Sekunde
NaviLat	Breitengrad des Ortungsergebnisses nach WGS84
NaviLatRich	Richtung des Breitengrads des Ortungsergebnisses nach WGS84
NaviLon	Längengrad des Ortungsergebnisses nach WGS84
NaviLonRich	Richtung des Längengrads des Ortungsergebnisses nach WGS84
NaviQuaGesch	Gesamtgenauigkeit bezüglich Geschwindigkeit am Erdboden
NaviQuaRich	Gesamtgenauigkeit bezüglich der Ausrichtung (Kurswinkel)
NaviQuaWR	Gesamtgenauigkeit bezüglich der 2D-Position (horizontal)
NaviQuaSR	Gesamtgenauigkeit bezüglich der 1D-Position (vertikal)
NaviHoh	Höhe über dem Meeresspiegel
NaviZeit	UTC-Zeit, Sekunden seit dem 01.01.1970 00:00
FeuchPrzt	Feuchtewerte des Sensors. Signalwerte > 100 % bedeuten relative Feuchte = 100 %
AnzGesch	Fahrzeuggeschwindigkeit
TempSt	Status Temperatursensor
Temp	Außentemperatur
LichtTop	Lichtsensor über dem Fahrzeug
LichtVorn	Lichtsensor vor dem Fahrzeug
LichtSnrsStFeh	Fehlererkennung Lichtsensor
LichtSnrSt	Verbauinfo Lichtsensor (ja/nein)
RegSnr	Regenintensität über Reflektivität des opto-elektronischen RLS in %
WischTipp	Wischerhebel: Tippwischen
WischIntSet	Wischerhebel: Stellung Intervall (Empfindlichkeit)
WischSt1	Wischerhebel: Stufe Wischen 1
WischSt2	Wischerhebel: Stufe Wischen 2

WischVorn	Wischerhebel: Stellung Wisch Wasch vorne
WischHeckInt	Wischerhebel: Stellung Wisch Wasch hinten
WischHeck	Wischerhebel: Stellung Heckintervall
WischInt	Wischerhebel: Stellung Intervallstufen
RegSnrStFeh	Verbauinfo Regensensor (ja/nein)
RegSnrSt	Fehlererkennung Regensensor (ja/nein)
WischGes	Wischergeschwindigkeit (Hübe/min)
Signale des nachgebauten RLS	
SonIntRe	Sonnenintensität rechts (2D Sensor)
SonIntLin	Sonnenintensität links (2D Sensor)
FSPrzt	Relative Luftfeuchte
FSTemp	Außentemperatur am Sensor
TempScheibe	Außentemperatur an der Scheibe
FeuchTauPkt	Aus der Luftfeuchte rechnerisch ermittelter Wert
LSVorn	Lichtsensor vor dem Fahrzeug
RegMen	Regenintensität über Reflektivität des opto-elektronischen RLS in %
LSTop	Lichtsensor über dem Fahrzeug
LichtDun	Regen- / Lichtsensor
RegWischGes	Wischergeschwindigkeit (Hübe/min)

Für die Übertragung der Daten aus dem Fahrzeug wurde gemäß dem höchstmöglichen Messtakt ein zeitlicher Abstand von einer Sekunde ausgewählt.

Überlegung zur einer ereignisgesteuerten Daten-Aggregierung und -versendung

Es wurde keine ereignisgesteuerte Übertragung konzipiert, da sich eine kontinuierliche Übertragung als vorteilhafter herausstellte. So war es zum Beispiel möglich, Daten aufzunehmen, während ein Fahrzeug im Betrieb keinen Niederschlag verzeichnete, im gleichen Zeitraum das zur Auswertung angesetzte Niederschlagradar jedoch ein Niederschlagsereignis anzeigte bzw. für die aufgenommenen Daten eine Niederschlagsinformation abgeleitet wurde. Das bei einer kontinuierlichen Übertragung ermittelte Daten-Übertragungsvolumen konnte problemlos über den ausgewählten Mobilfunkanbieter realisiert werden.

Aus praktischen Erwägungen wurde entschieden, die Datenverarbeitung auf der Axotec-Box zu minimieren. Programm-Anpassungen auf der Fahrzeugbox eines Fahrzeugs durchzuführen, das evtl. seinen Betriebszustand häufiger auch über einen kleinen Zeitraum verändert, sind komplexer als auf einem zentralen Server. Deshalb wurde keine Aggregierung in der Axotec-Box durchgeführt und die Daten wurden direkt nach ihrer Erhebung über Mobilfunk an einen zentralen Server gesendet. Das verwendete Übertragungsprotokoll (MQTT)[2] eignete sich dabei besonders gut für die kleinen Datenpakete (2 – 3 kB pro Eintrag).

Um die Sicherheit und ständige Verfügbarkeit des Servers zu gewährleisten, wurden die Einrichtung einer geeigneten IT-Infrastruktur und der Serverbetrieb in einem Unterauftrag vergeben. Nach eingehender Recherche und dem folgenden Vergabevorgang beauftragte das FiW die CPU 24/7 GmbH, die mit der Sicherheitsstufe „Tier 4“ eine hohe Sicherheit garantieren konnte, mit dem Aufbau der geforderten IT-Infrastruktur und einer Außenanbindung an das Internet. Weiterhin stellte CPU 24/7 die Installation, Konfiguration und den Betrieb eines MQTT-Brokers zur sicheren Datenübertragung

2 **MQTT:** Message Queuing Telemetry Transport = offenes Netzwerkprotokoll für Machine-to-Machine-Kommunikation (M2M). Ermöglicht die Übertragung von Telemetriedaten in Form von Nachrichten zwischen Geräten, trotz hoher Verzögerungen oder beschränkter Netzwerke.

aus den Kfz zur Verfügung und leistete dem Konsortium Applikationssupport während der Projektlaufzeit. Gegenüber Cloud-Diensten wie z. B. der Amazon Cloud hatte CPU 24/7 u. a. den Vorteil, dass der Serverstandort in Deutschland liegt und somit im Geltungsbereich europäischer Gesetzgebung, wie z. B. der DSGVO.

Für die Steuerung der Datenaufnahme wurde eine webbasierte Schnittstelle bereitgestellt. Die Boxen konnten entweder über Mobilfunk oder die Wifi-Schnittstelle erreicht werden und die Änderungen bei der Datenaufnahme durchgeführt werden. Wenn nötig, wurden Änderungen der ganze Flotte mittels Server übernommen.

Zur Sicherstellung der Funktionsfähigkeit wurde die Serverauslastung bei CPU 24/7 am 17.09.2019 unterschiedlichen Tests unterzogen. Hierzu wurde eine Reihe von API-Aufrufe teilweise parallel durchgeführt und große Datenmengen gleichzeitig abgefragt. Zusätzlich wurde die Internetbandbreite und Datenverfügbarkeit einigen Tests unterzogen. Die Tests waren erfolgreich, sodass das gewählte Serverkonzept beibehalten wurde.

TA 1.4 Big Data Handlung – Methoden zur Datenaufnahme und Datenübertragung

Im dritten Arbeitspaket wurden die Methoden zur Datenaufnahme und Datenübertragung, also praktisch dem Umgang mit Big-Data-Mengen, entwickelt. Zu Beginn stand auch hier eine Definition der Anforderungen und Randbedingungen der zu sammelnden Daten. Auch potenzielle Drittdaten (s. Kapitel 6.1.3) wurden dabei berücksichtigt.

Die Anforderungen enthalten dabei Definitionen der Datenstruktur, der raumzeitlichen Variabilität, der Regelmäßigkeit der Datenaufnahme und der Identifikation einer möglichen Normallast und Hochlast. Eine Überlastung der Mobilfunkkapazitäten durch zu große Datenmengen konnte dabei zu keinem Zeitpunkt festgestellt werden. Das Datenaufkommen war verständlicherweise zu den Betriebszeiten von EGLV erhöht, von einem abgegrenzten Wechsel von Normal- und Hochlastphasen konnte aber nicht gesprochen werden.

Folgende Informationen konnten als wesentliche Anforderungen an die Big-Data-Hard/-Software definiert werden:

- Speichermenge für Fahrzeugdaten: jedes Fahrzeug sendet bis zu 500 MB pro Monat.
- Anzahl MQTT-Pakete:
  - Auf Basis der Größe jeder Messung (4 kB) = 300.000.000 Pakete
  - Auf Basis einer Fahrzeugnutzung von 20 Tagen pro Monat = 165.000.000 Pakete
  - Bis zu 625.000 Pakete pro Tag
- Die „Privacy Switch"-Funktion musste vor der Datenspeicherung eine Filterung der Daten vornehmen.
- Das System musste 24 Stunden pro Tag verfügbar sein.

Die schnelle Verarbeitung der Datenmenge sowie die Privacy Switch Funktion benötigten eine Verarbeitung der Daten in Echtzeit (Streamprocessing). Um eine ständige Verfügbarkeit zu erreichen, mussten die Systeme weitestgehend automatisiert sein.

Um die Qualität der erhobenen Sensordaten zu sichern, wurden Prüfroutinen implementiert, die auf ihre Sinnhaftigkeit überprüft wurden. Beispielsweise stellten Sensorwerte ohne Übertragung eines gültigen GPS-Signals keinen Mehrwert dar und wurden direkt nach der Übertragung an den Server eliminiert. Zu Beginn der Datenübertragung wurden für einige Sensoren häufig Fehlerwerte übermittelt. Durch eine kontinuierliche Überprüfung und Anpassung der Übertragungstechnik durch Updates konnten diese Fehlerwerte minimiert werden.

Überlegung zur gegenseitigen Kalibrierung von aneinander vorbeifahrenden Kfz

Auf eine Entwicklung von Methoden zur automatisierten gegenseitigen Prüfung von Sensordaten aus verschiedenen Fahrzeugen in Abhängigkeit der Entfernung der Fahrzeuge voneinander wurde aus praktikablen Gründen verzichtet. Die Fahrzeugdichte der 100 Kfz auf ein Betrachtungsgebiet von ca. 4.200 km² bot dabei kein Potenzial für einen Vergleich von Sensoren verschiedener Fahrzeuge, da die Entfernung der Fahrzeuge zueinander nur in statistisch nicht relevanten Ausnahmefällen gering genug war. Auch eine Integration von automatisierten Kalibrierungsroutinen bei Annäherung eines Fahrzeugs an eine stationäre Messstelle wurde nicht weiterverfolgt. Grund dafür sind die geringe Dichte an Kfz-Datensätzen in der Nähe von stationären Messstellen sowie die Unterschiede in den Messverfahren von Fahrzeugen und Pluviometern. Die Pluviometer und die Niederschlagsradarmessungen wurden stattdessen als eine mögliche Validierungsquelle für das Endprodukt der Informationsmodellierung identifiziert.

Berücksichtigung von Drittdaten

Im Rahmen der Berücksichtigung von Drittdaten wurde von EGLV eine detaillierte Auswertung, sowohl von Produkten des Deutschen Wetterdienstes als auch von EGLV-eigenen Entwicklungen durchgeführt, woraus weitere Übertragungsintervalle und Datenvolumina abgeleitet werden konnten. Hierzu wurden folgende Eigenschaften zusammengetragen:

- räumliche und zeitliche Auflösung,
- Erstellungs- und Bereitstellungszeit,
- Anzahl an Dateien,
- zugehörige Dateigrößen,
- Datenformate,
- Quellen,
- Produktbeschreibungen,
- Kurzbeschreibung des fachlichen Inhaltes für
  - sechs Produkte aus der Gruppe der Niederschlagsradarprodukte,
  - drei Produkte aus der Gruppe der deterministischen numerischen Wettervorhersagen,
  - zwei Produkte aus der Gruppe der probabilistischen numerischen Wettervorhersagen,
  - drei Produkte aus der Gruppe der Niederschlagsstationsaufzeichnungen und
  - vier Produkte aus der Gruppe der verfügbaren Geodaten.

Im Zuge dessen wurden auch die Nutzungsrechte der in Betracht gezogenen Drittdaten geprüft. Da es sich mit Novellierung des DWD-Gesetzes vom 25.07.2017 größtenteils um frei zugängliche Daten handelte, war an dieser Stelle keine gesonderte Genehmigung zur Verwendung der Daten im Rahmen des Forschungsvorhabens einzuholen. Für gesondert bereitgestellte Distrometerdaten wurde eine Vereinbarung mit dem Deutschen Wetterdienst getroffen, welche eine entgeltfreie Bereitstellung der Datensätze zur projektgebundenen Nutzung im Rahmen der Unterstützung von Forschung und Lehre für mobile[VIEW] erlaubte.

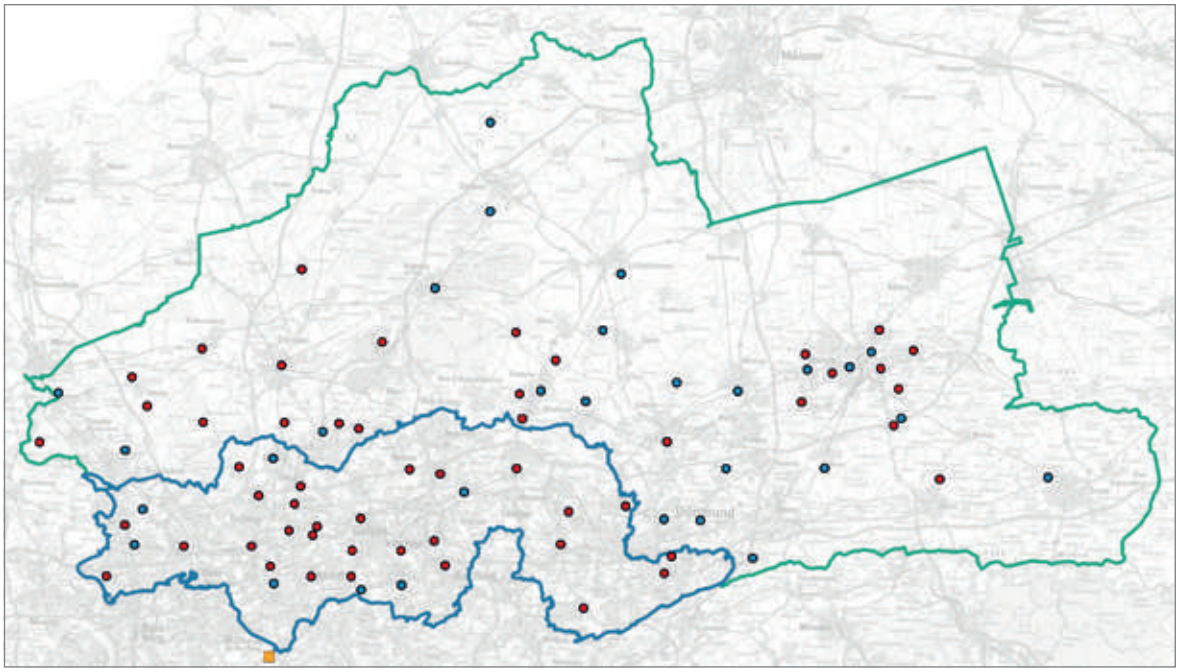


Abbildung 6: Übersicht über die räumliche Verteilung von EGLV-Niederschlagsstationen in den Verbandsgebieten nach Bereitstellungsintervall (rot = minütlich, blau = stündlich). © EGLV

Für diese Produkte wurde seitens EGLV eine Bereitstellung für das Forschungsvorhaben auf Basis eines FTP-Transfers auf einem hauseigenen Server implementiert, wozu eine zeitgetaktete Übertragung der Daten in gesonderten Austauschverzeichnissen sowie die Einrichtung von Nutzerzugängen für die Konsortialpartner zwecks Möglichkeit eines Datenabrufs zu zählen ist. Darüber hinaus konnten von 82 Niederschlagsstationen insgesamt 60 Stationen aufgrund der technischen Rahmenbedingungen im Rahmen einer stündlichen Übertragung in „near real-time" für die Konsortialpartner bereitgestellt werden. Hiervon war es bei 54 Stationen sogar möglich, eine zweiminütliche Bereitstellung zu realisieren (vgl. Abbildung 6).

TA 1.5 Hardwareintegration und Testfahrten

Beim Einbau der Messsysteme in die Fahrzeuge gab es unterschiedliche Anforderungen zwischen nachträglich eingebauten Sensoren sowie bereits serienmäßig vorhandener Sensorik. Die Anzahl belief sich hierbei auf 41 Fahrzeuge mit vorhandener Sensorik und entsprechend 59 Fahrzeuge ohne vorhandenen RLS. Im Falle von Kfz mit nachgerüsteter Sensorik musste noch ein zusätzlicher Mikrocontroller zwecks Signalverarbeitung und Kommunikation des nachträglich installierten Sensors mit dem eigentlichen Datenlogger verbaut werden. Die Montageklemmen der RLS konnten hierbei mit punktuell aufgebrachtem handelsüblichem Silikon in Verbindung mit Streifen doppelseitigen Klebebandes an den Kontaktstellen zur Windschutzscheibe angebracht werden (vgl. Abbildung 8). Abgesehen von wenigen Fahrzeugen hielt die Befestigungskonstruktion über die Projektlaufzeit im Wesentlichen stand. Lediglich bei wenigen Fahrzeugen kam es aufgrund von starken Temperaturwechseln im Winter und Sommer zu einer Ablösung.

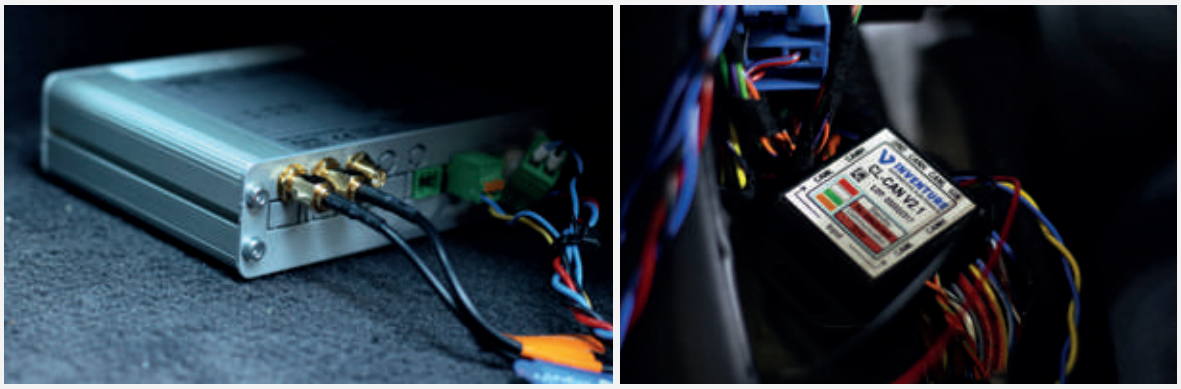


Abbildung 7: Verwendeter Datenlogger in Form einer Axotec-Datenbox (links) und verwendete induktive CAN-Bus-Klemme zum nicht-invasiven Mitlesen der Datenströme im Fahrzeug (rechts). © EGLV

Abbildung 8: Serienmäßig in den Rückspiegelfuß integrierter Regen-Licht-Sensor (links) und nachträglich montierte Montagefassung für das Forschungsvorhaben (rechts). © EGLV

Der Einbau der Hardware vollzog sich in folgenden Schritten:

- Einbau des CAN-Readers
- Einbau und Verdrahtung der Axotec-Box (vgl. Abbildung 7)
- Sicherstellung der Stromversorgung (über die OBD-Buchse)
- Einbau der RLS (wenn kein Serien-RLS vorhanden)
- Einbau einer zusätzlichen Steuerungsbox bei nachgebauten RLS
- Systemtest

Nachdem alle Komponenten in einem Fahrzeug eingebaut wurden, wurde die Inbetriebnahme über die Box-Web-Schnittstelle vor Ort sichergestellt. Danach wurde die Mobilfunkverbindung überprüft, indem der Eingang von Messdaten am Server bestätigt wurde.

In Anlehnung an die Methodik eines Forschungsprojektes in den USA (Bartos et al. 2019) wurden durch das FiW beschaffte Kameras zur Dokumentation von Messfahrten anlassbezogen in den Fahrzeugen angebracht, um eine unabhängige und zeitlich hochaufgelöste Datenquelle für die Benetzung der Windschutzscheibe zu erhalten und diese anschließend mit den rohen Sensormessungen abgleichen zu können. Bei Bartos et al. ging es im Vergleich mit mobile^{VIEW} jedoch nur um eine binäre Aussage über das Niederschlagsgeschehen und nicht um eine Quantifizierung.

Da die Dichte von Kfz-Datensätzen in der Nähe von Pluviometern sehr gering war, wurden auch geplante Messfahrten durchgeführt und protokolliert. Ziel dieser Messfahrten war die Erhöhung der Kfz-Datensätze in der Nähe von Pluviometern sowie eine Kalibrierung der Sensordaten an den Pluviometern. Testfahrten im direkten Umfeld des Niederschlagsradarstandortes in Essen sowie nahegelegener terrestrischer Niederschlagsstationen wurden in Abhängigkeit vielversprechender Niederschlagslagen innerhalb der Betriebszeiten am 01.10.2019, 01.07.2020, 09.07.2020 sowie am 21.10.2020 zusammen mit dem FiW durchgeführt. Die Testfahrten fanden zu verschiedenen Zeitpunkten statt und wurden mit unterschiedlichen Zielen durchgeführt. Übergeordnetes Ziel war der Abgleich von aufgezeichneten Kfz-Messdaten mit weiteren Datenquellen. So wurden beispielsweise Messungen am Niederschlagsradar Essen sowie an ausgewählten Pluviometerstandorten in den Verbandsgebieten durchgeführt. Hier konnten auch Einflüsse der Fahrtgeschwindigkeit auf die Messung nachvollzogen werden. Zu einigen Zeitpunkten war es auch möglich, parallele Messungen mit zwei Kfz durchzuführen.

Im Projektzeitraum war das Untersuchungsgebiet geprägt durch drei außerordentlich trockene Jahre mit ausgeprägten Dürreperioden im langfristigen Vergleich. Darunter fiel sowohl der Rekord-Sommer 2018 als auch der Dürre-Sommer 2019. Auch im Jahr 2020 setzte ab April eine durchgehende Trockenperiode ein, sodass gerade in den ansonsten starkregenträchtigen Monaten eine große Trockenheit und Dürre herrschte. In Kombination mit Ereignissen, die vielfach am späten Nachmittag bzw. Abend auftraten, was eine häufige Entwicklung bei konvektiven Sommerereignissen darstellt, sowie Ereignissen am Wochenende, konnte im Endeffekt nur eine begrenzte Anzahl an Messfahrten durchgeführt werden.

Die Begleitung der Auswertung erhobener Messdaten in der Testphase sowie die Prüfung der aufgenommenen Mess- und Metadaten bzw. Plausibilisierung der Ergebnisse fand hierbei fortlaufend statt. So wurde die korrekte Datenerfassung, -übermittlung und -ausgabe fortwährend im Hinblick auf räumliche Plausibilität der Fahrtstrecken, zugeordnete Messzeiten sowie übertragene Parameter begutachtet. Zudem wurde regelmäßig kontrolliert, ob die implementierten „Privacy Switches" sowohl im „opt-out"- als auch „opt-in"-Modus korrekt arbeiteten und ob die Anzahl der datenübermittelnden Fahrzeuge plausibel war und einen erwarteten Tagesgang aufwies. Im weiteren Projektverlauf wurden im Rahmen einer zweiten Iteration zudem alle QR-Codes umgestaltet, um die „Privacy Switches" fortan über codierte URLs aufzurufen und somit unter stärkerer Berücksichtigung von Aspekten der IT-Sicherheit eine Fremdbedienung der Schalter bei bekannter Kfz-ID zu unterbinden.

Im Rahmen der Datenübertragung prüften alle Partner die Übertragungsmöglichkeit von einem zentralen Server über eine REST (Representational State Transfer) – Schnittstelle zum jeweiligen Partner. Die Daten konnten auf diese Weise erfolgreich übermittelt und beim jeweiligen Partner gespeichert werden.

Das Resultat dieses Arbeitspaketes war somit die Ausstattung von 100 Kfz der EGLV-Fahrzeugflotte sowie der Nachweis über die Funktionsfähigkeit der Messtechnik. Die Datenaufnahme und -sicherung konnte durchgeführt und die Dienstgüte der Rohdaten, aus der Sicht der Anwender (Quality of Service), als geeignet für eine weitere Analyse bewertet werden.

6.1.2 Teilprojekt TP 2: Datenverarbeitung

TA 2.2 Aufbau eines Datenbanksystems zur Big Data Verwaltung

Zu Beginn des Teilprojektes zur Datenverarbeitung wurde ein Datenbanksystem zur Verwaltung von Big-Data-Mengen benötigt. Hierfür war eine Definition der Anforderungen an das System erforderlich, die gemeinsam von FiW und EGLV erstellt wurden.

Ein wichtiges Kriterium war dabei die Integration von Drittdaten. Das Datenbanksystem wurde so konzipiert, dass ein späteres Hinzufügen weiterer Produkte möglich ist. Für die Drittdaten mussten ebenfalls Definitions- und Integrationsbestimmungen getroffen werden. Die entsprechende Datenrecherche wurde im Rahmen von Teilprojekt 3 durchgeführt. Als relevante Drittdaten wurden das dämpfungskorrigierte DX-Niederschlagsradarprodukt (DX-D) von EGLV sowie die Pluviometerdaten terrestrischer Messstationen von EGLV ausgewählt. Das Datenbanksystem sollte außerdem die Fahrzeugdaten über das MQTT-Protokoll empfangen und eine Batchverarbeitung der in Echtzeit ankommenden Datenpakete ermöglichen. Zudem sollte der Zugang für alle Projektpartner gewährt werden. IAV setzte anschließend anhand der Anforderungen des Lastenhefts die Erstellung des Datenbanksystems um:

Für die Integration der Drittdaten von EGLV wurde ein FTP-Connector verwendet. Der Zugang für die Projektpartner wurde über eine webbasierte REST-Schnittstelle realisiert und mit individuellen Zugangsdaten geschützt.

Die Dateiformate waren produktspezifisch. Für Niederschlagsradardaten eignen sich generell andere Dateiformate (z. B. Raster) als für die durch die Kfz generierten Daten (objektorientierte Formate). Die Sensordaten der Kfz wurden als menschenlesbares Datenaustauschformat in *JSON* (JavaScript Object Notation) übertragen.

Die Pluviometerdaten wurden durch EGLV im *zrxp*-Format bereitgestellt. Das dämpfungskorrigierte Niederschlagsprodukt (DX-D) wurde von EGLV als ebenfalls menschenlesbar im *ESRI-Grid-Format (*.asc)* bereitgestellt. Beide Produkte wurden über einen durch die EGLV betriebenen ftp-Server bereitgestellt. Das Endprodukt, die Multi-Source-Precipitation-Map (MSPM), wurde durch die IAV zunächst im *JSON*-Format bereitgestellt, da dies im Gegensatz zum *asc*-Format die Übermittlung von Metadaten (z. B. Versionsnummer, Anzahl der gefahrenen Kfz, etc.) ermöglichte. Um Kompatibilität mit den von EGLV benutzten Programmen zu gewährleisten und Vergleiche zu erleichtern, wurde die MSPM ebenfalls im *asc*-Format bereitgestellt.

Die Speicherung und Verarbeitung der Daten sollte auf einem gemeinsamen Datenserver (Backend) erfolgen. Hierfür musste zunächst ein geeigneter Serverstandort gefunden werden. Entscheidende Kriterien stellten in diesem Zusammenhang eine ständige Verfügbarkeit des Servers und eine hohe Sicherheit des Standortes gegen unbefugten Zugriff dar.

Konzept der IoT-Plattform

Eine weitere Anforderung an das Datenbanksystem beinhaltete die Visualisierung von aufbereiteten Daten. Zu diesem Zweck setzte die IAV eine Internet-of-Things-Plattform (IoT-Plattform) auf, auf der die erhobenen Daten dargestellt und mit verschiedenen Werkzeugen bearbeitet werden konnten. FiW und EGLV erarbeiteten dazu die notwendigen Anforderungen.

Zunächst wurden die auf der IoT-Plattform zu integrierenden Sensordaten definiert sowie die Tools und Visualisierungsmöglichkeiten, die eine Darstellung von Analyseergebnissen ermöglichen sollten. Weitere Anforderungen waren die erforderliche Bandbreite und die zeitliche und räumliche Auflösung der Sensordaten.

Die IoT-Plattform wurde so konzipiert (Abbildung 9), dass eine Lambda-Architektur (Abbildung 10) mit den erforderlichen Schnittstellen eingesetzt werden konnte. Die Lambda-Architektur ist in Abbildung 10 näher erläutert.

Abbildung 9: Konzept der IoT-Plattform. © IAV

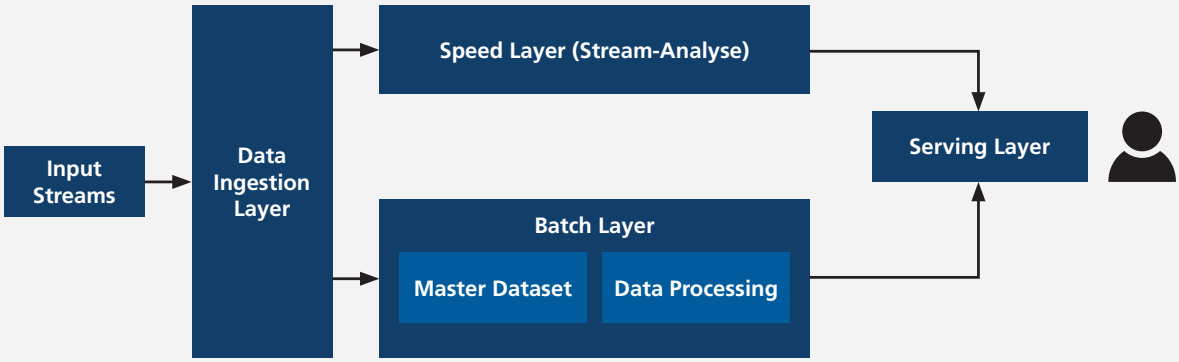


Abbildung 10: Schaubild einer Lambda-Architektur. © IAV

Um die Entwicklung der IoT-Plattform zu beschleunigen, wurde die Anwendung von Open-Source Lösungen der *Apache Software Foundation*[3] (vgl. Tabelle 2) bevorzugt. Die Apache Software implementierte den Stand der Technik bei gleichzeitiger Sicherstellung der Erfüllung aller Anforderungen, die im Projekt gestellt wurden. Für die unterschiedlichen Komponenten der Plattform wurde somit eine in der Praxis erprobte Lösung ausgewählt:

Tabelle 2: Komponenten der IoT-Plattform.

Software	Rolle
Apache Hadoop HDFS	Hochverfügbares Dateisystem zur Speicherung sehr großer Datenmengen auf den Dateisystemen mehreren Rechnern.
Apache Flink	Einheitliches Framework, um Datenströme und Stapeldaten zu verarbeiten.
Apache Kafka	Einheitliche Plattform, um Echtzeitdatenströme mit hohem Durchsatz und geringer Latenz zu verarbeiten.
Apache Zookeeper	Service für verteilte Systeme, der hierarchische Schlüssel-Werte-Paare bietet. Wird als verteilter Konfigurations- und Synchronisationsservice und Namensregister für große verteilte Systeme eingesetzt.
Apache Hadoop Yarn	Ermöglicht die Ressourcen eines Clusters für verschiedene Jobs dynamisch zu verwalten.

Als einzige kommerzielle Lösung wurde aufgrund der Möglichkeit zur Kafka-Integration der MQTT-Message-Broker HiveMQ verwendet, welcher von IAV bereitgestellt wurde. Für einen reibungslosen Betrieb war die schrittweise Konfiguration der Software erforderlich. Die Codes wurden in der Programmiersprache Java mit integrierter Entwicklungsumgebung Eclipse geschrieben. Für die Versionsverwaltung wurde GitLab verwendet.

Entgegen der ursprünglichen Planung konnte die IoT-Plattform jedoch nicht für eine Visualisierung öffentlich zugänglich gemacht werden. Die dort dargestellten und verarbeiteten Rohdaten entsprachen personenbezogenen Daten im Sinne der DSGVO und durften nicht ohne weitere Anonymisierung veröffentlicht werden.

Dieses Arbeitspaket lieferte somit folgende Ergebnisse: Der Datenserver zur Speicherung und Verarbeitung von Daten wurde definiert und über einen Unterauftrag gemäß den definierten Anforderungen eingerichtet. Für die Verarbeitung der Big-Data-Mengen wurde die entsprechende Software installiert, konfiguriert und in Betrieb genommen. Für die vereinfachte Veranschaulichung von Rohdaten wurde eine projektinterne IoT-Plattform aufgesetzt. Die automatisierte Datenübertragung zur IoT-Plattform konnte in Betrieb genommen werden.

TA 2.3 Konzeptionierung und Entwicklung der Informationsmodellierung

Um eine gemeinsame raum-zeitliche Grundlage zur Weiterverarbeitung von Niederschlagsdaten aus verschiedenen Quellen zu schaffen und die späteren Ergebnisse der verdichteten Niederschlagsinformationen veranschaulichen und öffentlich zugänglich machen zu können, entwickelte das Konsortium ein Konzept zur Informationsmodellierung. FiW und EGLV lieferten dafür Anforderungen an die Informationsmodellierung für die Ziele aus Sicht der Wasserwirtschaft. Dazu fand im Februar 2019 ein Workshop „Anforderungen und Genauigkeitskriterien“ in Essen statt.

Hierbei wurden die Daten in verschiedene Kategorien unterteilt, die jeweils unter den Gesichtspunkten „Bereitstellung“, „Analyse“ und „Visualisierung“ untersucht wurden. Der Verarbeitungsprozess sah demnach vor, die erhobenen Rohdaten zunächst zu aggregierten Produkten zusammenzufassen, die weiterhin lediglich die Kfz-Sensordaten enthielten. In einem dritten Schritt erfolgte dann die Verschneidung mit Drittdaten zu kombinierten Produkten. Für die Rohdatenanalyse und Evaluierung von Wechselwirkungen standen die Rohdaten und aggregierten Produkte allen Projektpartnern

3 Die Apache Software Foundation ist eine ehrenamtlich arbeitende Organisation zur Förderung der Apache-Softwareprojekte, zu denen unter anderem auch der Apache-Webserver gehört.

zur Verfügung. Kombinierte Produkte wurden nach ihrer Erstellung auf ihre Plausibilität untersucht. In der später angestrebten Visualisierung wurden primär die kombinierten Produkte und, falls möglich in Hinblick auf die datenschutzrechtlichen Vorgaben, aggregierten Produkte verwendet. Als Zielsysteme für die Anwendung der Projektergebnisse wurden das Hochwasserinformationssystem Delft-FEWS von EGLV und für die Visualisierung die IoT-Plattform definiert.

Besondere Bedeutung im Rahmen der Informationsmodellierung kam den konzeptionellen Überlegungen zur raumzeitlichen Aggregation von Niederschlagsdaten unterschiedlicher Auflösungen mit eingehenden Rohdaten zu einem finalen Produkt zuteil, welches gegenüber den Einzeldaten einen Mehrwert lieferte. Initial wurden einheitliche raumzeitliche Bezugsgrößen für die weitere Verarbeitung festgelegt – so z. B. ein einheitliches räumliches Koordinatensystem im Sinne von ETRS 89/UTM 32 N (EPSG: 25832) sowie ein einheitlicher Zeitbezug in Form von GMT bzw. UTC+0. Ebenso wurde beschlossen, übereinstimmend in den Einheiten „mm" für Niederschlagshöhen und „mm/h" bei Niederschlagsintensitäten zu operieren, um die Datenkonsistenz zu gewährleisten. Schließlich wurden im Zuge dessen die Formatspezifikationen von kompatiblen Datenformaten zur Anbindung des finalen Niederschlagsproduktes an Delfts-FEWS beleuchtet.

Meteorologische Vorüberlegungen

Im Rahmen der Konzeptionierung im Sinne einer Zusammenfassung von bekannten Wechselwirkungen von erhobenen Messdaten verfolgte EGLV einen induktiven Ansatz und präsentierte mit Unterstützung des Geographischen Instituts der RWTH Aachen den zeitlichen Verlauf diverser meteorologischer Parameter für zwei ausgewählte Niederschlagsereignisse (29.04.2018, 27.05.2020) auf Grundlage der Klimastation Aachen-Hörn. Der Fokus wurde hierbei auf die Entwicklung der relativen Feuchte, der Lufttemperatur sowie bodennahen Temperatur, des Luftdrucks, der Strahlung, der mittleren Windgeschwindigkeit sowie der maximalen Windgeschwindigkeit im Zeitraum hoher Niederschlagsintensitäten im Sinne von Starkregen gelegt. Im Ergebnis zeigte sich, dass je nach Wetterlage eine außergewöhnliche Temperaturabnahme, Reduktion der Globalstrahlungskomponente sowie Zunahme der Windgeschwindigkeit verzeichnet werden konnte, während beim Luftdruck eher der mittelfristige Trend von Bedeutung schien und die relative Feuchte keine eindeutigen Ergebnisse lieferte. Während der Einfluss von Veränderungen der Lichtverhältnisse sowie der Lufttemperatur auf Basis der Kfz-eigenen Sensorinformationen nachgebildet werden konnte, fiel die mögliche Integration von Veränderungen des Luftdrucks sowie der Luftfeuchtigkeit weg, da keine geeigneten Sensordaten in den Fahrzeugen zur Verfügung standen.

Überlegung zur Klassifikation von Niederschlagsklassen nach Starkregenindex (SRI)

Sollte die Qualität der grundlegenden Kfz-Sensordaten eine Intensitätsermittlung nicht ermöglichen, schlug das FiW alternativ eine Klassifikation von Niederschlagsintensitätsstufen vor. Kurzzeitig wurde in diesem Zusammenhang auch über eine Anlehnung an den Starkregenindex diskutiert. Dieser Ansatz wurde jedoch aus den folgenden Gründen nicht weiterverfolgt: Der Starkregenindex (Schmitt et al. 2018) baut auf den nach DWA-A 531 ermittelten Wiederkehrzeiten in Abhängigkeit von Niederschlagshöhe und Dauerstufe auf. Als Grundlage dienen jährliche bzw. partielle Serien möglichst langjähriger Aufzeichnungen. Weiterhin dient der SRI der retrospektiven Risikokommunikation von Starkniederschlagsereignissen. Für die Echtzeitanwendung ist dieser nicht geeignet, da das Ende eines Starkniederschlagsereignisses zum Zeitpunkt der Betrachtung nicht bekannt ist. Eine Klassifizierung von Radardaten auf Grundlage von langjährigen Stationsdaten wäre zwar grundlegend möglich, ist jedoch methodisch nicht korrekt, da der SRI auf einer ortsbezogenen extremwertstatistischen Auswertung von Niederschlagsstationsdaten beruht und somit aufgrund der lokalen Gegebenheiten nur für diesen einen Standort gültig ist. Eine einfache Übertragung in die Fläche ist an der Stelle fachlich nicht korrekt. Aus diesen Gründen wurde eine Klassifizierung von Sensordaten in Anlehnung an den SRI nicht weiterverfolgt.

Überlegungen zur Identifikation der relevanten Kfz-Sensoren

In Vorbereitung der Rohdatenanalyse brachten FiW und EGLV Überlegungen bezüglich der Wechselwirkungen von wasserwirtschaftlichen und meteorologischen Sensordaten ein. Die Menge an zu untersuchenden Kfz-Sensoren wurde auf diese Weise reduziert. Nachfolgend sind die in den Kfz gemessenen Parameter aufgelistet, die für einen Eingang in das Informationsmodell genauer untersucht wurden und somit das Potenzial für einen Mehrwert wasserwirtschaftlicher Anwendungen aufwiesen:

- Regensensor
- Scheibenwischerfrequenz, automatisch und manuell
- Temperatursensoren
- Lichtsensoren
- Luftfeuchte
- Fahrzeuggeschwindigkeit

Hinsichtlich der Berücksichtigung von Drittdaten legten FiW und EGLV beim Workshop im November 2019 eine finale Aufstellung der im Informationsmodell zu implementierenden Drittdaten vor. Im Zuge der grundsätzlichen Überlegungen zur Aggregation bzw. Verschneidung der drei verfügbaren Niederschlagsprodukte, bestehend aus dem EGLV-eigenen Niederschlagsradarprodukt „DX-Dämpfungskorrigiert", Aufzeichnungen von EGLV-Niederschlagsstationen sowie eingehenden Sensorinformationen, war vorab eine gemeinsame raumzeitliche Grundlage für die Weiterverarbeitung zu schaffen. So wurde festgelegt, die Sensorinformationen fünfminütlich, entsprechend der Auflösung des verwendeten Niederschlagsradarproduktes, zu aggregieren. Da eine zeitliche Aggregation bei nicht-stationären Messungen im Sinne sich stets bewegender Kfz stets auch eine räumliche Aggregierung bedeutet, wurde als Bezugsmatrix ebenfalls das räumliche Raster des Niederschlagsradarproduktes mit einer räumlichen Auflösung von 1 km x 1 km herangezogen. Hierbei wurden anhand von Fallbeispielen konkrete Varianten zur Aggregation von Kfz-Rohdaten beleuchtet. Betrachtet wurden beispielsweise das Verhalten mehrerer Kfz innerhalb einer Rasterzelle oder die Durchfahrt eines Fahrzeugs durch mehrere Rasterzellen innerhalb des fünfminütigen Aggregationsintervalls. Zur quantitativen Auswertung von Niederschlagsmessungen wurde ein mehrstufiges Vorgehen vorgeschlagen. Betrachtet wurde der Abgleich von Sensor-Messwerten von Fahrzeugen untereinander, von Sensor-Messwerten zwischen Regenschreibern und vorbeifahrenden Kfz in einem vordefinierten Radius sowie der Vergleich von fünfminütlich aggregierten Kfz-Messungen mit zugehörigen Niederschlagsradarpixeln. Schließlich wurden, da zu diesem Zeitpunkt noch keine gesicherten Erkenntnisse zur Qualität der Sensormessungen vorlagen, mehrere Szenarien ausgearbeitet, wie gewonnene Informationen im Forschungsvorhaben gewinnbringend verarbeitet werden können. Dies erfolgte in Abhängigkeit von dem Vorliegen valider Niederschlagsintensitäten oder dem Vorliegen von Niederschlagsinformationen in lediglich qualitativen Klassen.

Bei grundsätzlicher Übereinstimmung der Niederschlagsschwerpunkte zwischen terrestrischen Stationsdaten und Kfz-Messungen wurde die anschließende Kombination der drei Quellen im Sinne einer maximal möglichen Verdichtung vorhandener Messpunkte zur Erzeugung eines kombinierten Niederschlagsproduktes anvisiert. Im Rahmen einer Verschneidung der unterschiedlichen Quellen wurde seitens EGLV eine Priorität hinsichtlich der erwarteten Messgenauigkeit mitgeteilt. So waren prioritär terrestrische Messdaten zu verwenden, ergänzt um das Niederschlagsradarprodukt und aufgefüllt mit Kfz-Messungen, mittels derer aufgrund der hohen raum-zeitlichen Auflösung gewisse Unsicherheiten bei der radarbasierten Detektion von Niederschlag wie die Driftkorrektur oder bei Radarausfall hätten ausgeglichen werden können. Pluviometer gelten dabei als die genaueste Datenquelle am Erdboden, die jedoch lediglich für den Messstandort repräsentativ ist. Die Größenordnung zu erwartender Unsicherheiten im Rahmen von terrestrischen Messungen in Höhe von 10 % (Schellart et al. 2012) sowie vertretbare Gültigkeitsradien der punktuellen Aufzeichnungen wurden dem Konsortium mitgeteilt. Bei der Abschätzung der Unsicherheiten von radargestützten Messungen wurde auf die Entfernungs- und Höhenabhängigkeit der Messung vom Radarstandort verwiesen, sodass ein entfernungsabhängiges Konfidenzintervall in Höhe von 25 % bis in eine Entfernung von 40 km vom Radarstandort Essen und von 35 % darüber hinaus (Krämer et al. 2019) eingeführt wurde. In Bezug auf die Kfz-Messungen erfolgte eine Berücksichtigung des Anteils an Messungen eines Fahrzeugs in einer Zelle zur gewichteten Einbeziehung von Sensordaten im Umfeld von Stationen. Als Anwendungsziele wurden die Ableitung von Warninformationen sowie die Anwendung für die Vorhersage von Starkregenereignissen und für die Modellierung von Hochwasserereignissen festgehalten.

Für die Berücksichtigung im Informationsmodell und somit die Verschneidung mit den Kfz-Sensordaten wurden zwei Produkte ausgewählt:

- Das Dämpfungskorrigierte DX-Niederschlagsradarprodukt von EGLV wurde in Form eines Rasters zur Verfügung gestellt. Das EGLV-Gebiet ist auf dieser Karte in Zellen mit einer Größe von 1 km x 1 km unterteilt. Für jede Rasterzelle wurde ein repräsentativer Wert in der Einheit [mm/5min] zur Verfügung gestellt. Das zeitliche Intervall der Bereitstellung entsprach somit fünf Minuten.

- Die Pluviometer des terrestrischen Messstellennetzes von EGLV wurden analog zu den Kfz-Sensordaten als Punktwerte gemäß ihrer Lagekoordinaten zur Verfügung gestellt. Die Messwerte entsprachen Minutensummen und wurden daher in der Einheit [mm/min] zur Verfügung gestellt.

Rohdatenanalyse

Die Daten können grundsätzlich in Flächen- und Punktmessungen unterteilt werden. Die als Raster gespeicherten Niederschlagsradardaten entsprechen über dem gesamten EGLV-Gebiet erhobenen Volumenmessungen. Sowohl Kfz-Sensoren als auch die Pluviometer entsprechen Punktmessungen, d. h. sie erheben Daten jeweils an ihrer aktuellen Position.

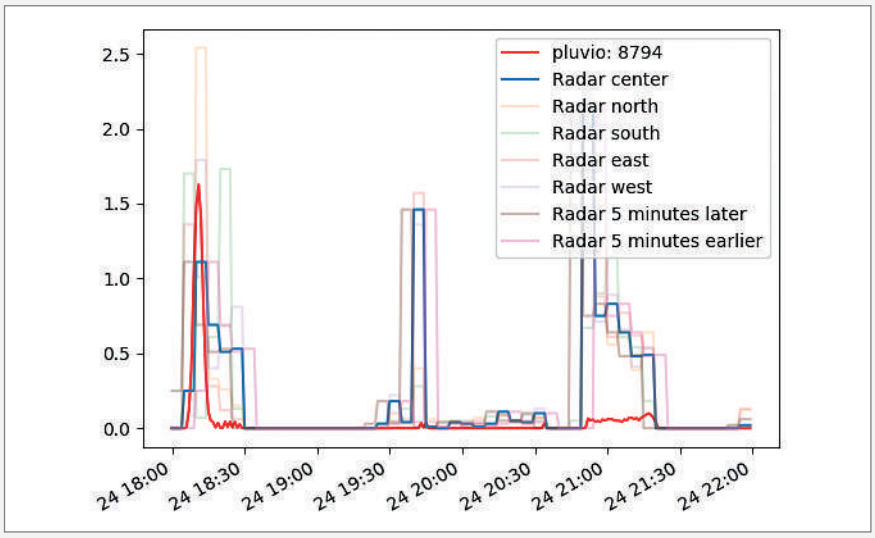


Abbildung 11: ***Gemessener Niederschlag an einem Pluviometer und in den umliegenden Radarzellen. © FiW e. V.***

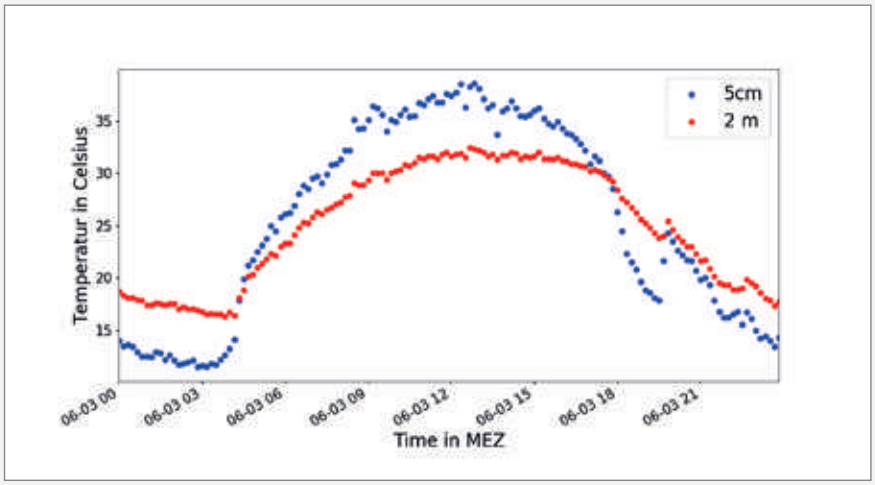


Abbildung 12: ***Temperaturaufzeichnung an einer DWD-Klimastation in einer Höhe von 5 cm (rot) und 2 m (blau). © FiW e. V.***

Da sowohl das Niederschlagsradar als auch die Niederschlagsschreiber etablierte Formen der Niederschlagserfassung sind, lag das Hauptaugenmerk der Rohdatenanalyse in mobile[VIEW] auf der Rohdatenanalyse der Fahrzeugsensoren. Jedoch wurden auch relevante Zusammenhänge zwischen Flächen- und Punktmessungen untersucht. Das Niederschlagsradar stellt als indirektes Messverfahren eine Messung der Reflektivität in der Höhe dar, von welcher auf den Niederschlag am Boden geschlossen wird. Allerdings können Fehlerquellen, wie beispielsweise Abschattungen, Vögelschwärme oder Winddrift zu einer Überschätzung des auf den Boden auftreffenden Niederschlag führen, auf der anderen Seite führen Dämpfungseffekte (z. B. Niederschlag über dem Radom, hintereinander liegende Ereignisse) zu einer Unterschätzung des Niederschlags. Abbildung 11 zeigt ein Niederschlagsereignis an einem Pluviometer und den durch das Niederschlagsradar aufgezeichneten Niederschlag direkt über der Station (Center) zum selben Zeitpunkt als auch um fünf Minuten versetzt, sowie den nördlich, östlich, südlich und westlich versetzten Niederschlag. Der Niederschlag um ca. 18:15 Uhr zeigt, dass die Niederschlagsspitzen durchaus deckungsgleich sein können, allerdings verzeichnet der Niederschlagsschreiber zu zwei späteren Zeitpunkten, im Gegensatz zum Radar, wenig oder keinen Regen. Dies kann einerseits mit den oben beschriebenen Fehlerquellen zusammenhängen, andererseits sind gerade konvektive Niederschlagsereignisse sehr kleinräumig verteilt. Da die Radarzelle eine Auflösung von 1 km x 1 km aufweist, kommt es vor, dass mittels Radar ein Niederschlag aufzeichnet wird, anhand des Pluviometers allerdings nicht.

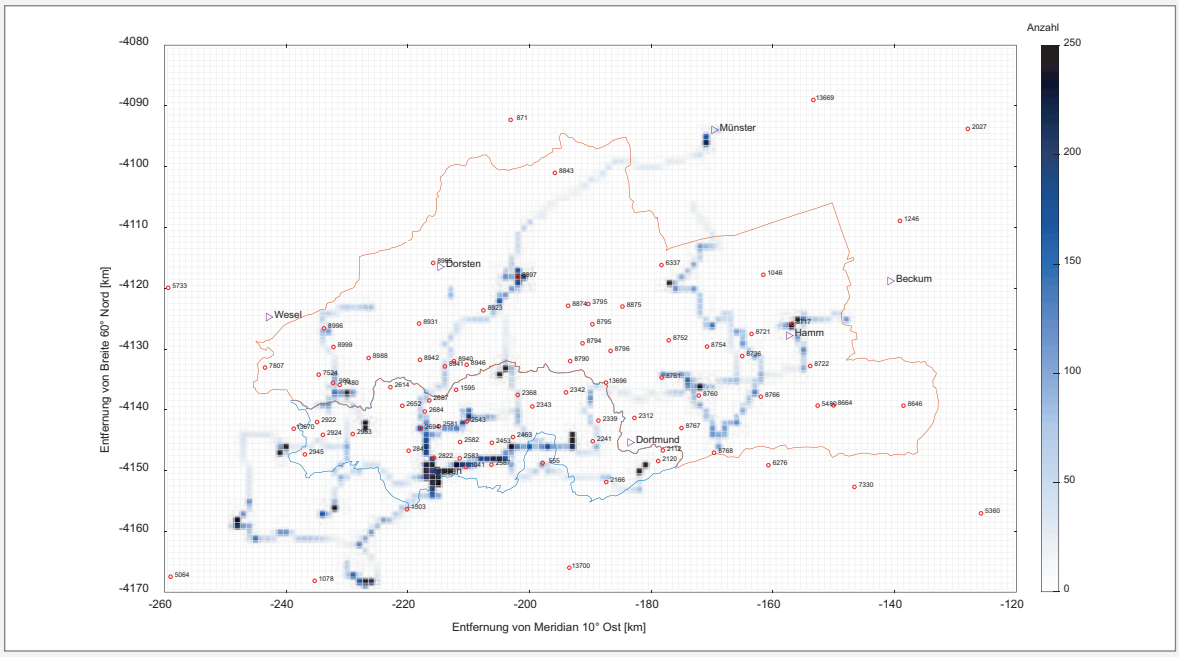


Abbildung 13: Heatmap von Fahrstrecken im EGLV-Gebiet zwischen 25.09.2018 und 05.10.2018. © IAV

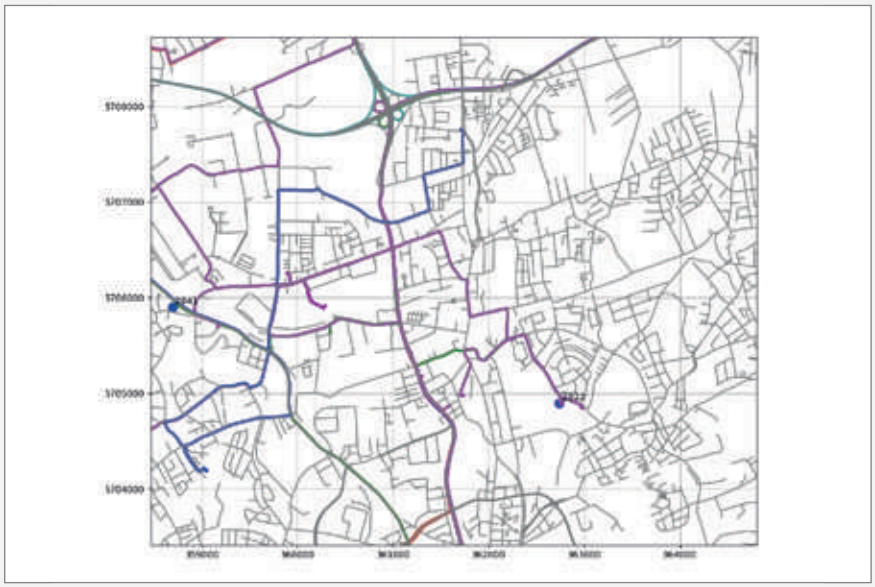


Abbildung 14: Fahrspuren verschiedener Autos (farbige Linien), Pluviometer (blaue Punkte) sowie Radarrasterzellen. © FiW e. V.

Abbildung 12 zeigt einen Temperaturverlauf über einen Tag im Sommer an einer Klimastation des Deutschen Wetterdienstes (DWD). Es ist eine klare Abhängigkeit der Temperaturverläufe von der Höhe erkennbar. Die Temperaturwerte, welche näher am Boden aufgezeichnet wurden, haben eine höhere Amplitude. Um 21 Uhr kann ein starker Temperaturabfall beobachtet werden. Ein Vergleich mit den für dieses Gebiet vorliegenden Niederschlagsdaten zeigte, dass es zu diesem Zeitpunkt zu einem Niederschlagsereignis kam. Somit konnte anhand dieses Ereignisses auf einen Zusammenhang zwischen dem Abfall der Temperatur und dem Niederschlagsgeschehen geschlossen werden. Auch wenn die Klimastationen des DWD aufgrund ihrer geringen Anzahl keinen Eingang in das spätere Informationsmodell fanden, waren diese Erkenntnisse bei der Erstellung des Modells hilfreich.

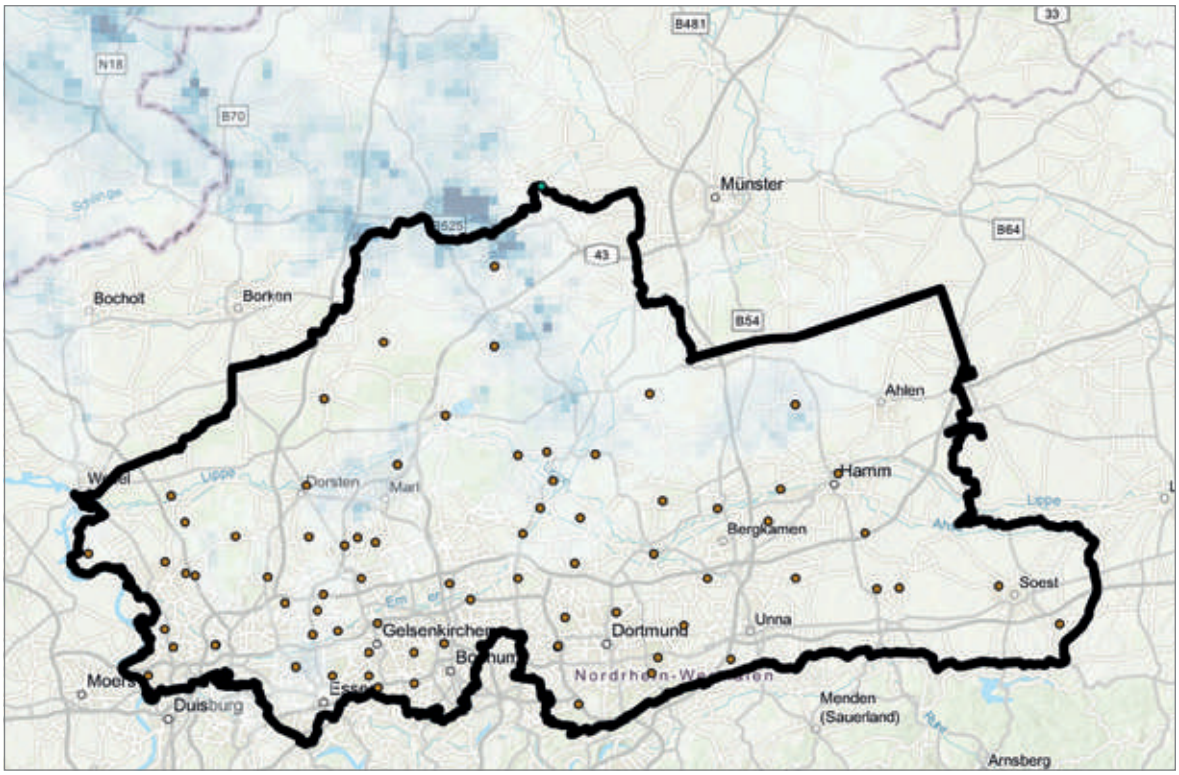


Abbildung 15: Überprüfung und Visualisierung von Ergebnissen in ArcGIS Pro. © FiW e. V.

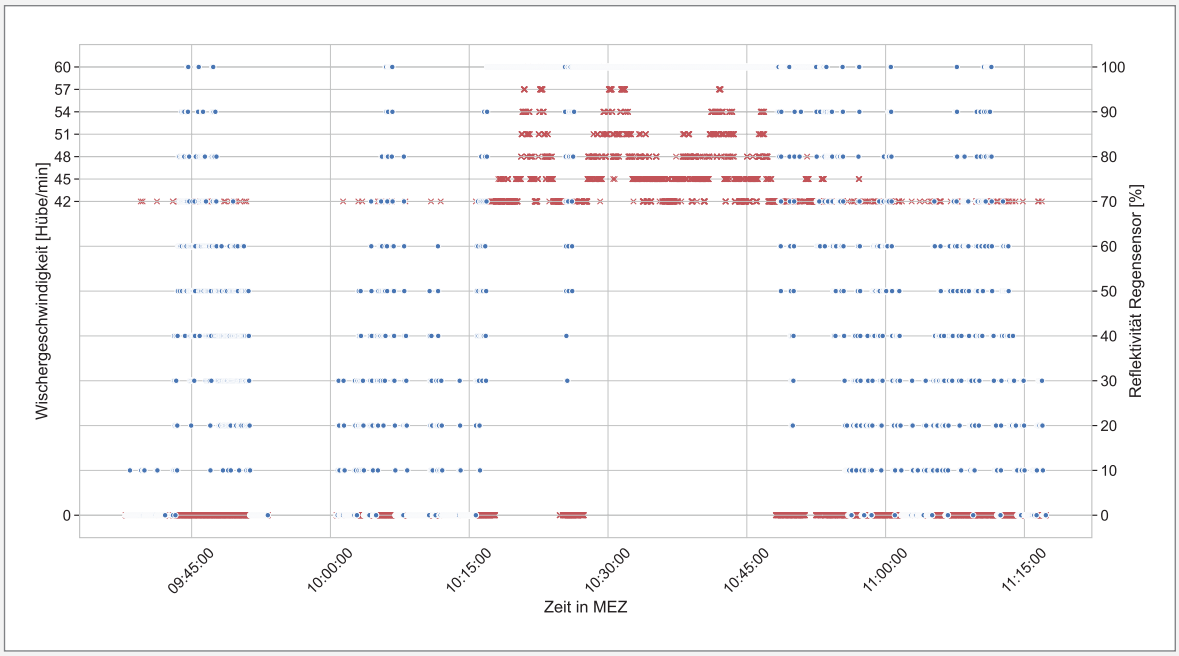


Abbildung 16: Zeitlicher Verlauf der Scheibenwischerstufen gegenüber Regensensorwerten. © FiW e. V.

Der nächste Schritt der Rohdatenanalyse der Fahrzeugsensoren war das Prüfen der Funktionsfähigkeit der Übertragungstechnik und Serverinfrastruktur. Dies geschah unter anderem durch Visualisierung von Fahrstrecken in Matlab, wie in Abbildung 13 für das gesamte Verbandsgebiet dargestellt als Heatmap oder durch straßengenaue Überprüfung der Fahrzeugfahrspuren, wie in Abbildung 14. Dort sind ebenso die Pluviometer (blaue Punkte) als auch die Auflösung der Radarzellen von je 1 km x 1 km (hellgraues Raster) verzeichnet. Ebenso wurden die Pluviometer und Testdatensätze zur schnellen Visualisierung und Überprüfung in Geoinformationssystem eingeladen (ArcGIS Pro) und die Ergebnisse (MSPM) hier laufend geprüft (Abbildung 15).

In jedem Fahrzeug konnte im Projekt technisch auf ca. 40 verschiedene Sensoren zugegriffen werden. Eine zentrale Herausforderung musste hier bewältigt werden: Einerseits sind Kfz-Sensoren generell auf die Fahrzeugführenden ausgerichtet, d. h. für Komfort und Sicherheit der Fahrzeugführenden bzw. für die Fahrzeugelektronik optimiert. Somit sind die meisten dieser Sensoren für die übergeordneten wasserwirtschaftlichen Fragestellungen ungeeignet. Andererseits ist die zugängliche Dokumentation der einzelnen Sensoren sehr eingeschränkt. Da für viele Sensoren, der genaue Verbauort oder Zweck nicht dokumentiert waren, mussten solche Zusammenhänge über eine ausführliche Rohdatenanalyse geklärt werden. Um einen Überblick über das Verhalten, die Wertebereiche und die Häufigkeitsverteilungen der Sensoren zu bekommen wurden zunächst zeitliche Verläufe und Histogramme erstellt.

Die Abbildung 16 zeigt die zeitlichen Verläufe der aktiven Wischerstufen im Auto sowie die durch den verbauten Niederschlagssensor aufgezeichneten Werte. Der Scheibenwischer wurde in diesem Fall automatisch durch das Fahrzeug gesteuert. Der Wert 0 entspricht Null Hüben pro Minute, der Wert 1 entspricht 42 Hüben pro Minute. Mit einer Abstufung von jeweils 3 Hüben pro Minute entspricht der Wert 7 somit 60 Hüben pro Minute. Der Regensensor gibt Werte zwischen 0 % und 100 % aus, mit Abstufungen von jeweils 10 %. Die Abbildung zeigt einerseits, dass die Werte einen stark schwankenden Verlauf aufwiesen, sich also nicht konstant einem Wert annäherten, sondern zwischen Werten hin und her oszillierten. Der einzige Wert, der durch den Niederschlagssensor im Fall eines Niederschlagsereignisses konstant ausgegeben wurde, war 100 %. Dies war auch der Wert, den die Scheibenwischer bei einem Wert größer als Stufe 1, also bei mehr als 42 Hüben/Minute, ansteuerten.

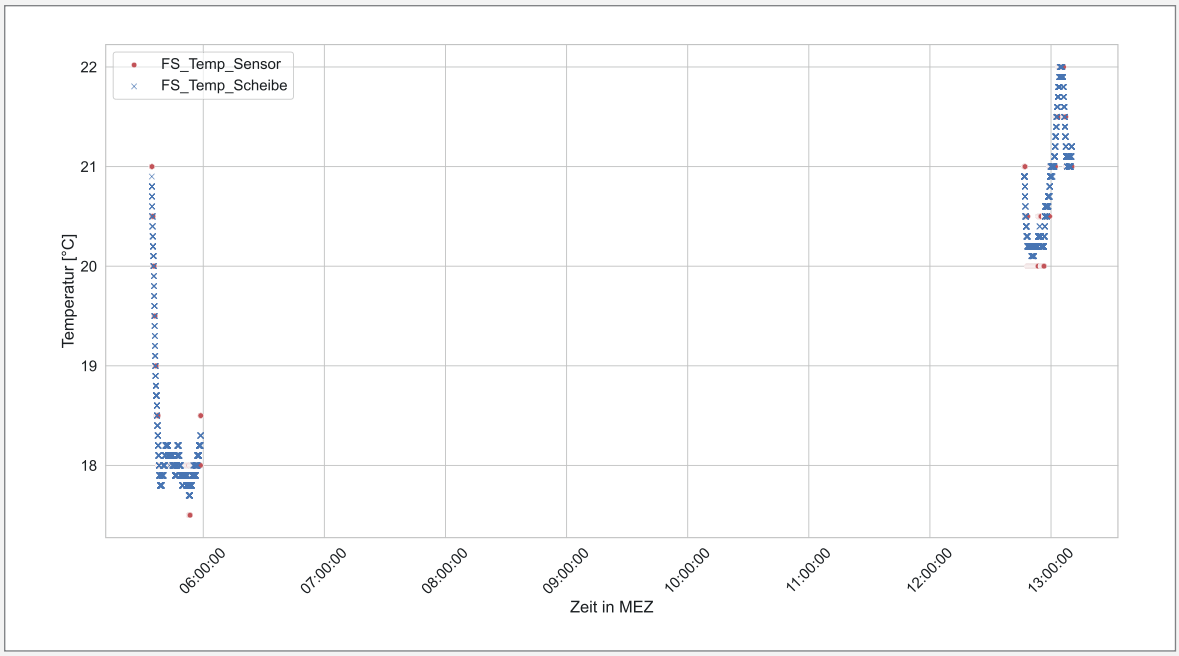


Abbildung 17: Zeitreihe der Fahrzeugtemperaturmesswerte an einem Morgen. © FiW e. V.

Die Volatilität der Messwerte wird ebenfalls an Abbildung 17 deutlich. Innerhalb von 15 Minuten fand um sieben Uhr morgens ein Temperaturabfall von 5 °C statt, obwohl der Trend um diese Uhrzeit in die entgegengesetzte Richtung ging. Bei der Datenübertragung wurden die ersten fünf Minuten jeder Fahrt nicht übertragen, was der Stabilisierung des GPS-Signals diente. Daher ist es unwahrscheinlich, dass in diesem Fall der Temperaturabfall durch eine Garagenausfahrt verursacht wurde. Dieses Verhalten wäre in einem möglichen Folgeprojekt auf Häufigkeit und Ursachen zu untersuchen.

Nachdem exemplarisch verschiedene Ereignisse geprüft wurden, wurden die Sensorwerte in ihrer Gesamtheit, also hinsichtlich ihrer Verteilungen und Wertebereiche, untersucht. Dafür wurde zunächst auf Histogramme zurückgegriffen. Die verfügbare Sensordokumentation zeigte, dass zwei verschiedene Lichtsensoren verbaut sind. Ein Sensor weist einen Wertebereich von 0 bis 6.126 lux, der andere von 0 bis 101.200 lux auf. Zum Vergleich, an einem bedeckten Sommertag auf 51° nördlicher Breite liegt, wenn die Sonne im Zenit steht, eine Beleuchtungsstärke von 19.000 lux (DIN 5034-2

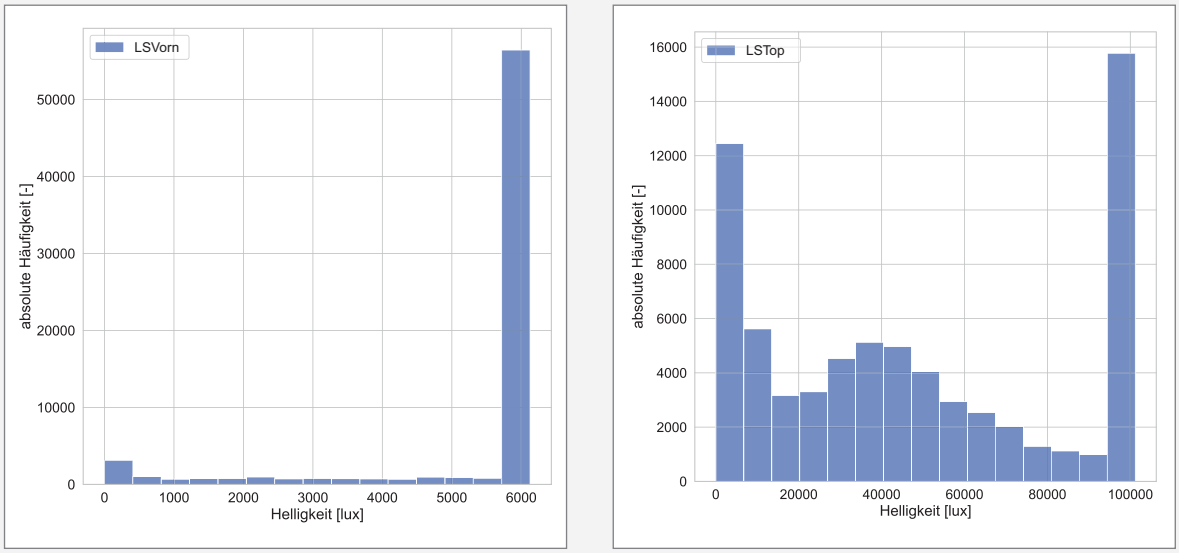


Abbildung 18: Histogramme der Lichtsensoren, 12.06.2019, x-Achse: Helligkeit in lux, y-Achse absolute Häufigkeit. © FiW e. V.

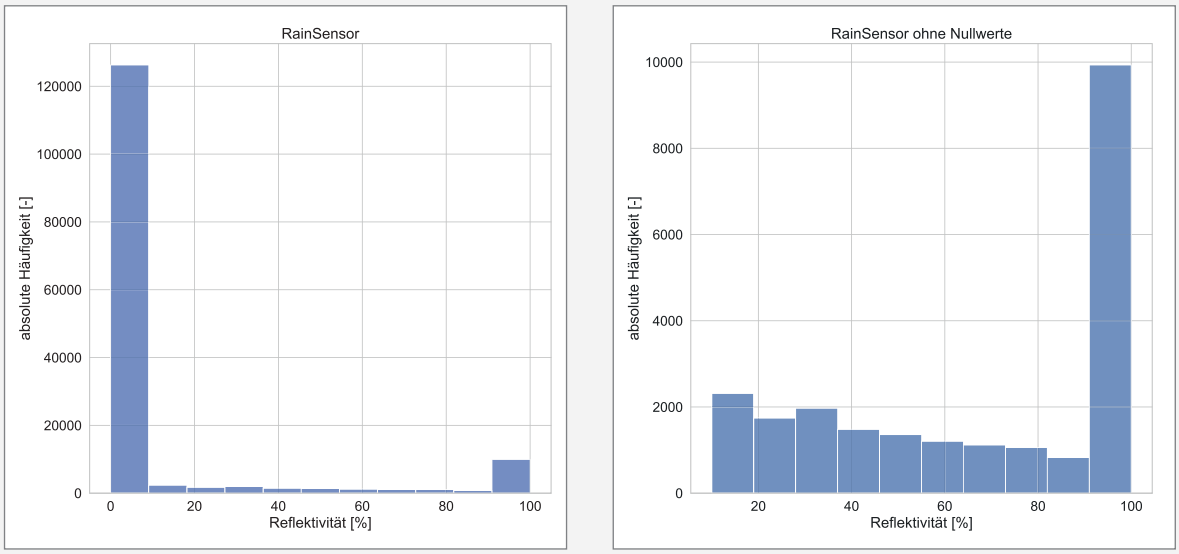


Abbildung 19: Histogramm der Regen-Licht-Sensoren aller Fahrzeuge am 12.06.2019. Auf der x-Achse die Werte des Regen-Licht-Sensors, auf der y-Achse die absolute Häufigkeit der Werte. © FiW e. V.

2019) vor. Am selben Ort zur selben Zeit bei direkter Sonneneinstrahlung herrschen ca. 90.000 lux (DIN 5034-2 2019). Abbildung 18 zeigt die Werteverteilung der beiden Sensoren für den 12.06.2019.

Da der Wert 0 hier nicht als Initialisierungs- oder Fehlerwert dokumentiert ist und der 12.06.2019 ein langer Sommertag war, sind in der späteren Analyse die Häufung der Nullwerte genauer zu prüfen. Der Sensor auf der linken Seite erreicht deutlich häufiger sein Maximum, welches allerdings auch deutlich geringer als das Maximum des rechten Sensor ist. Die Namen und die Wertebereiche der Sensoren legen nahe, dass der Sensor *LSVorn* so verbaut ist, dass er die Helligkeit vor dem Fahrzeug misst, beispielsweise Tunnelein- und -ausfahrten oder die Straßenbeleuchtung. Dieser Sensor dient vermutlich der Steuerung der Fahrzeugelektronik. Für meteorologische Zusammenhänge ist *LSTop*, welcher die Helligkeit über dem Fahrzeug misst, relevanter. Abbildung 19 zeigt das Histogramm für den Niederschlagssensor desselben Tages.

Der Regensensor gab Werte zwischen 0 % und 100 % aus, mit Abstufungen von jeweils 10 %. Der Wert 140 % bezeichnete einen Fehlerwert (Neu-Initialisierung des Sensors) und wurde später bereinigt. Das Histogramm zeigt deutlich, dass der häufigste aufgezeichnete Wert 0 % war, in diesen Fällen wurde kein Regen detektiert. Dies deckte sich häufig mit den meteorologischen Aufzeichnungen. Der zweithäufigste Wert ist 100 %, also eine völlige Sättigung des Sensors. Dies bedeutet, dass der Sensor ein sehr geringes unteres Detektionslimit aufweist, allerdings wird auch sehr schnell das obere Detektionslimit erreicht. Daraus ist zu schließen, dass bereits Ereignisse mit äußerst geringer Niederschlagsintensität detektiert wurden, allerdings aber auch, dass leichte, mittlere und schwere Ereignisse in der Klasse 100 % zusammengefasst wurden. Dieses Verhalten konnte bei späteren Messfahrten in der Nähe von Pluviometern bestätigt werden. Für eine automatische Steuerung der Scheibenwischeranlage ist diese Klassifizierung ausreichend und im Sinne der Fahrzeugsicherheit sinnvoll, aber diese „quasi-binäre" Verteilung der Sensorwerte

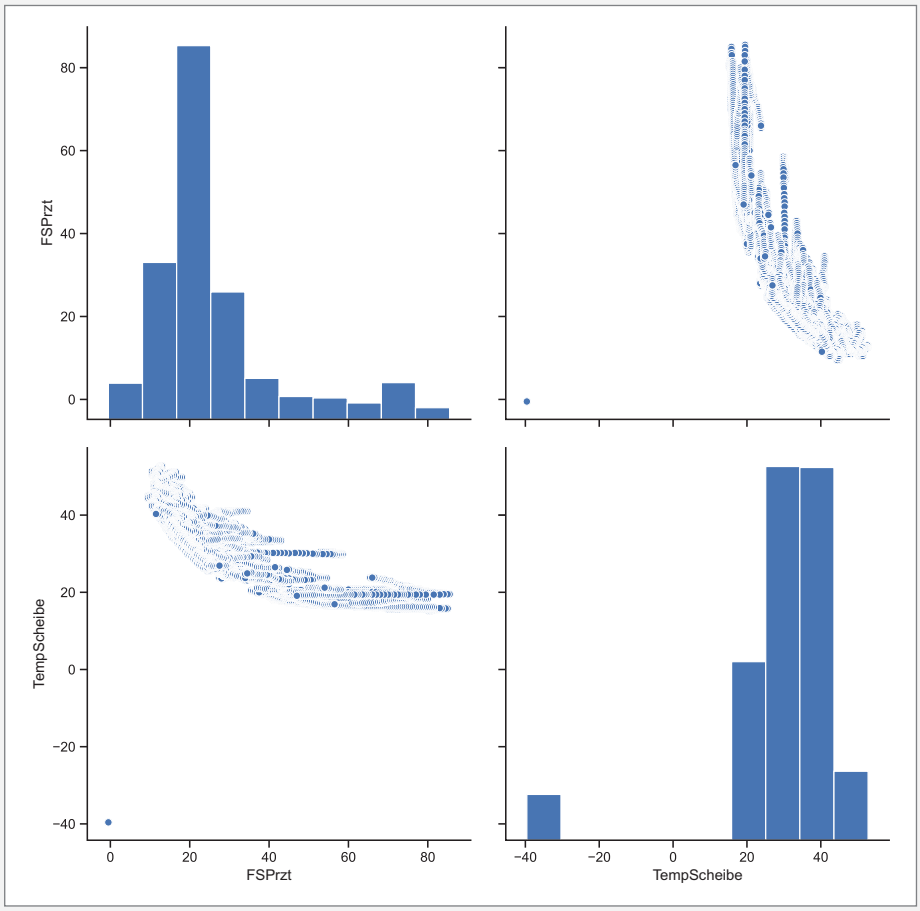


Abbildung 20: Streudiagramme vom 12.06.2019 mit dem Temperatursensor auf der Scheibe und dem Luftfeuchtesensor. © FiW e. V.

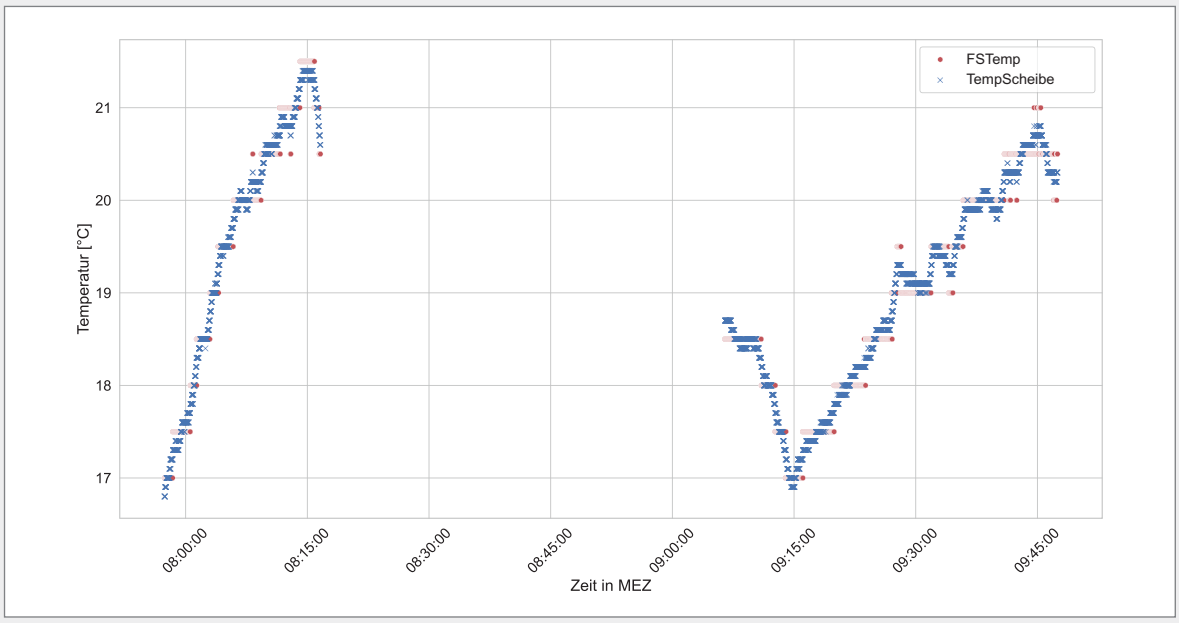


Abbildung 21: Zeitlicher Verlauf von FSTemp und TempScheibe. © FiW e. V.

machte eine Quantifizierung von Niederschlagsintensitäten ausschließlich aus den Regensensoren unmöglich. Dieser Zusammenhang stellte eine wichtige Erkenntnis des Projektes dar und war im Vorfeld des Projektes bereits als technisches Risiko definiert worden. Er beeinflusste die weitere Vorgehensweise zu Ableitung von Niederschlagsinformationen dahingehend, dass für diesen Prozess weitere Kfz-Sensoren hinzugezogen werden mussten. So wurde fortan ein Modell verfolgt, das Niederschlagsinformationen aus einer Kombination von Kfz-Sensordaten ableiten konnte.

Auf den Erkenntnissen der Histogramme aufbauend wurden Streudiagramme erstellt. Hier wurden Wertepaare zweier Sensoren in einem Diagramm und somit Abhängigkeitsstrukturen dargestellt. Abbildung 20 zeigt beispielsweise die Kombinationen der Wertepaare des in der Scheibe verbauten Temperatursensors *(TempScheibe)* mit dem Sensor für relative Luftfeuchte *(FSPrzt)*. Da die Werte hier unbereinigt dargestellt wurden, zeigten Temperaturen von knapp -40 °C sowie relative Luftfeuchte von 120 % jeweils Initialisierungs- bzw. Fehlerwerte an.

Um diese Zusammenhänge übersichtlich darzustellen und in Kennzahlen auszudrücken, wurden im nächsten Schritt Korrelationsmatrizen der Sensorwerte erstellt (exemplarisch für einige der Sensoren dargestellt in Abbildung 22). Ein Wert von „-1“ zeigt eine vollständige, negative lineare Korrelation an, ein Wert von „1“ eine vollständige positive lineare Korrelation. Der Wert „0“ hingegen bedeutet keine Korrelation. Zwei verschiedene Temperatursensoren (*TempScheibe* und *FSTemp*) zeigen beispielsweise eine sehr hohe positive Korrelation, d. h., dass beispielsweise eine Erhöhung um 1 °C eines Sensors mit einer fast identischen Erhöhung des anderen Sensors einhergeht. Abbildung 21 zeigt den Verlauf dieser beiden Sensoren, welche in demselben Fahrzeug verbaut sind. Tabelle 3 drückt die Ähnlichkeit in statistischen Kennzahlen aus. Aufgrund der Ähnlichkeit der Temperaturverläufe und der statistischen Kennzahlen wird deutlich, dass der Hauptunterschied die Auflösung der Messwerte ist und die Betrachtung eines Sensors im Modell ausreichend ist. Abbildung 21 ist auch ein Beispiel für die am FiW erstellten automatisierten Auswertungsroutinen in einem Dashboard, in welchem übersichtlich eine Vorauswahl an Sensoren dargestellt wird. Orientiert wurde sich dabei am Leitbild „Small multiple“. Hier wurde eine Reihe ähnlicher Graphen, welche dieselben Achsen und Skalen verwenden, nebeneinander dargestellt, um als Nutzer direkt einen visuellen Überblick zu bekommen. Von diesem Überblick aus, konnten dann einzelne Zusammenhänge betrachtet werden. Das Dashboard war interaktiv, sodass sowohl Sensoren als auch der Zeitausschnitt auch von technisch weniger affinen Nutzern eingestellt und so auch die Daten exploriert und Abbildungen erstellt werden konnten.

Tabelle 3: Statistische Auswertung der Temperatursensoren.

Parameter	FSTemp	TempScheibe
Anzahl	62337	62337
Arithmetisches Mittel	11,99	12,04
Standardabweichung	19,82	19,68
Minimum	-40,00	-39,60
unteres Quartil	16,50	16,30
mittleres Quartil	18,50	18,50
oberes Quartil	21,00	21,10
Maximum	33,00	33,20

Wichtig bei diesen Schritten war es, die Sensorwerte vor der Analyse zu bereinigen. Werden die Sensorwerte nicht vorher um Initialisierungs-, Fehler-, Null- sowie weitere Statuswerte bereinigt, können die Analysen zu falschen Interpretationen führen.

Ein Ergebnis der Korrelationsanalyse sind die Zusammenhänge der Sensoren für Luftfeuchte und den Taupunkt. Aus wasserwirtschaftlicher Sicht wurde zunächst angenommen, dass hier ein Zusammenhang zwischen meteorologischen Ereignissen und den Sensorwerten besteht. Die Korrelationsanalyse aus Abbildung 22 deutet allerdings auf das Gegenteil hin, dass diese Sensoren nicht mit den im Projekt erhobenen meteorologischen Daten der Umgebung korrelieren. In diesem Fall wurde die Korrelation noch eingehender untersucht. Hierzu wurden Fahrten einzelner Autos mit mindestens 100 Messungen, welche jeweils nicht mehr als fünf Minuten auseinanderliegen, sogenannte Trips, an mehreren aufeinanderfolgenden Tage betrachtet. Für die Sensorpaare Luftfeuchte und Regen-Licht-Sensor wurde der Pearson-Korrelationskoeffizient und der Durchschnitt pro Trip gebildet. Auch hier konnte keine Korrelation feststellt werden. Dies bestätigte die Vermutung, dass der Luftfeuchtesensor innenraumgerichtet, wahrscheinlich für die Steuerung der Klimaanlage, verbaut ist. Der Taupunkt stellt zudem lediglich einen aus den Parametern Luftfeuchte und -temperatur berechneten Wert dar.

Ebenso konnte hier der Zusammenhang zwischen dem Helligkeitssensor *(LSVorn)* vor dem Fahrzeug und dem Wert für die Hintergrundbeleuchtung des Displays *(AnzStr)* dargestellt werden. Aufgrund leichter negativer Korrelation führte ein Absinken der Umgebungshelligkeit zu einem erhöhten Wert der Helligkeit des Displays. *AnzStr* wurde somit aufgrund seines ausschließlichen Zwecks für die Fahrzeugführung im späteren Modell nicht verwendet. Ebenso wurde *LSVorn* dort nicht berücksichtigt, da dieser Sensor dem Erfassen des Lichts vor dem Fahrzeug dient als auch aufgrund seines geringen Maximalwertes, welcher die Erfassung von Sonneneinstrahlung nicht zuverlässig ermöglicht.

Aus den so gewonnen Erkenntnissen konnte unter anderem auf die Zusammenhänge der verschiedenen Signale der Scheibenwischeranlage geschlossen werden. In Abbildung 23 ist aufgezeigt, wie der Bordcomputer die Geschwindigkeit der Scheibenwischer einstellt. Die In-putparameter sind in den Parallelogrammen gezeigt, grün hinterlegt sind Einstellungen, die die Fahrerin einstellen kann, blau hinterlegte Parameter werden automatisch bestimmt. Orange hinterlegt ist der Outputparameter mit den entsprechenden Stufen, wobei hier die erste Zahl die ausgegebene Stufe meint, die zweite Zahl entspricht den tatsächlichen Hüben pro Minute. Es ist zu erkennen, dass manuell nur die unterste und die höchste Stufe angesteuert werden, das Fahrzeug jedoch über einen dem Konsortium unbekannten Algorithmus auch die Zwischenstufen ausgeben kann.

Das Verhalten der Sensorik wurde in Messfahrten überprüft. In Abbildung 24 ist die Strecke einer Messfahrt (rot) von Essen aus zu einer vom Deutschen Wetterdienst betriebenen Klimastation (blau) aufgezeichnet. Die Daten der Klimastation waren über das Webportal des DWD abrufbar. Während der Messfahrt lag in dem entsprechenden Gebiet ein Niederschlagsereignis vor und die Messfahrt bestand aus der Fahrt zur Klimastation als auch einem direkten Vergleich zwischen der Niederschlagsaufzeichnung des Pluviometers und des Niederschlagsensors. Letzterer wurde durch ein Stehen des Fahrzeugs in unmittelbarer Nähe des Pluviometers realisiert, wobei die Entfernung weniger als 100 m betrug. Die Werte dieses Vergleiches sind in Abbildung 25 dargestellt. Hier konnte belegt werden, dass der verbaute Regen-Licht-Sensor selbst sehr geringe Niederschlagsintensitäten (in diesem Fall 0.05 mm/min) detektiert.

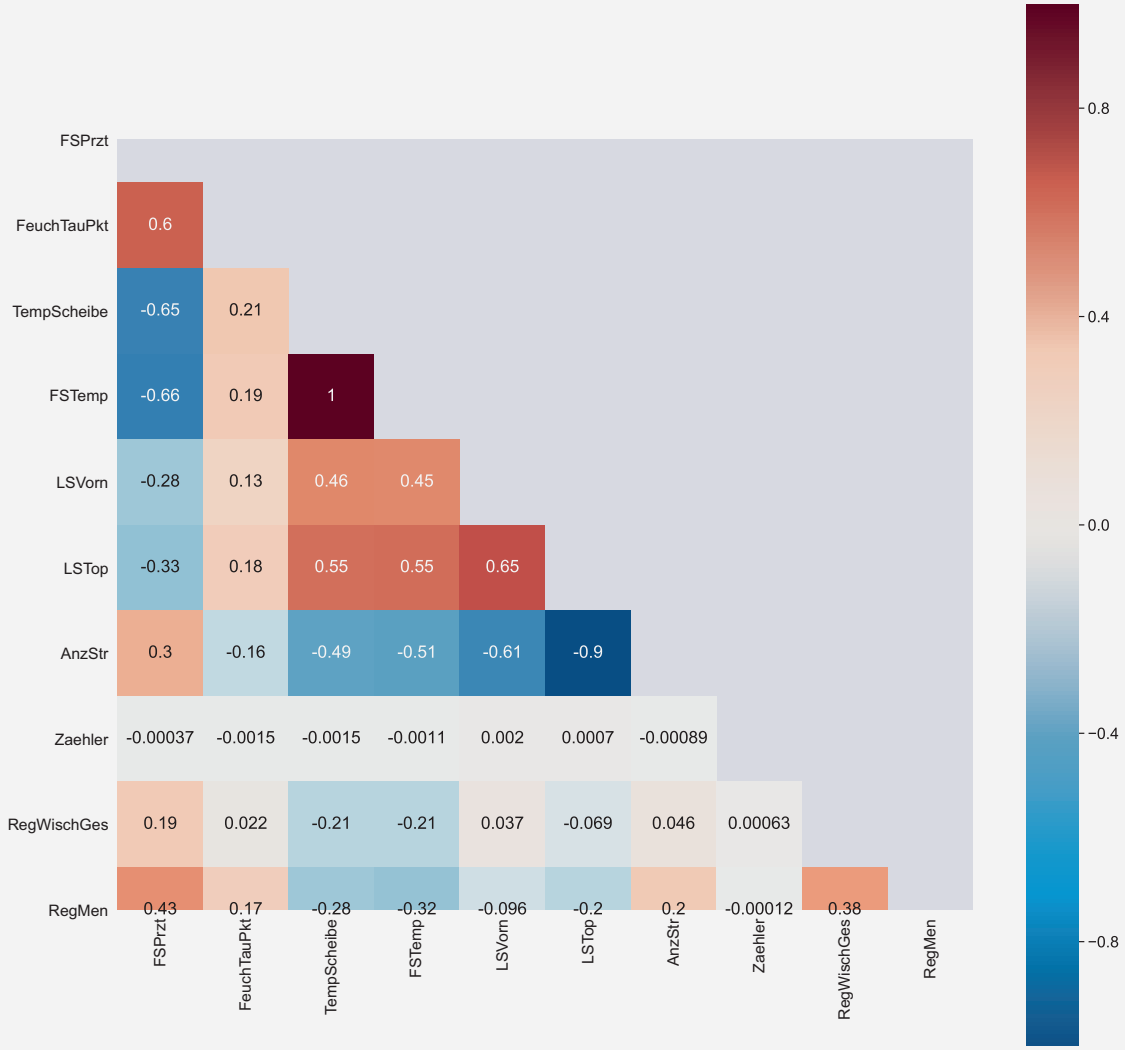


Abbildung 22: Korrelationsmatrix ausgewählter Kfz-Sensoren für den 28.05.2019. © FiW e. V.

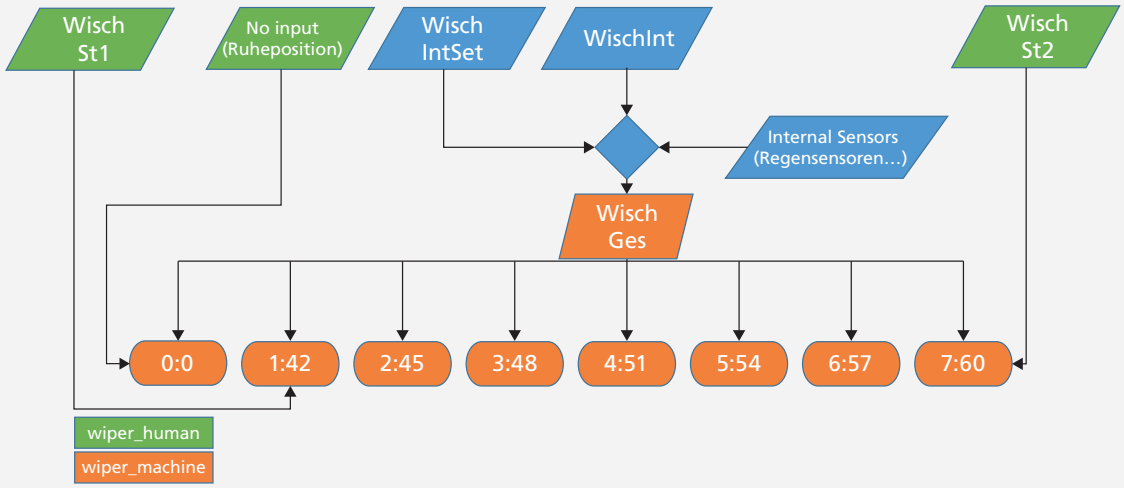


Abbildung 23: Flowchart der Signale der Scheibenwischeranlage. © FiW e. V.

Konzeptionierung des Informationsmodells

Zur Konzeptionierung des Informationsmodells wurden zuvor projektspezifische Anforderungen definiert (siehe TA 2.3). Essenzielle Voraussetzung hierfür war eine Aggregierung der unterschiedlichen Sensordaten (im Besonderen der Kfz-Sensordaten) zur Identifikation von Niederschlagsintensitäten. Neben der fachlichen Konsistenz wurde die Erklärbarkeit des Modells als entscheidende Voraussetzung definiert. Neben Interpretierbarkeit und Verständlichkeit stellt das Projekt auch Anforderungen an die Fehlerrate und die Wirksamkeit der Auswertealgorithmen. Des Weiteren ist ein Einsatz sowohl von Regressions- als auch von Klassifikationsmodellen geeignet, um sowohl diskrete als auch stetige Merkmale statistisch auswerten zu können.

In Absprache mit den Projektpartnern wurde ein Informationsmodell konzipiert, um die unterschiedlichen Einflussgrößen bei einem Niederschlagsereignis zu betrachten (Abbildung 26).

Die Regenintensität ist im Modell die Zielgröße. Eine direkte Volumenmessung des Niederschlages und eine Auswertung über ein bestimmtes Zeitintervall kann nur mittels eines Pluviometers durchgeführt werden. Radarmessungen und Fahrzeuge liefern nur indirekte Informationen, beziehungsweise abgeleitete Werte über Bestandteile der Regenintensität: Tropfengröße und Tropfenanzahl.

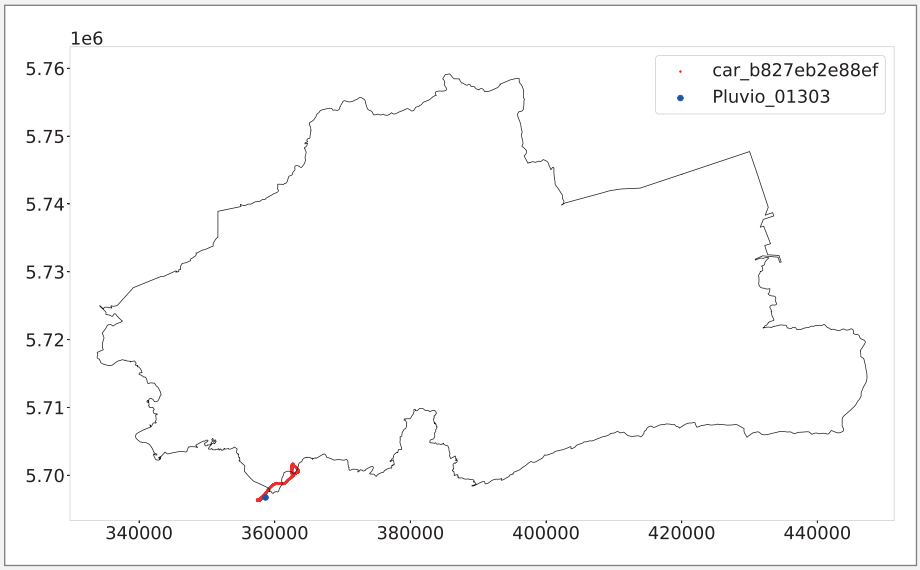


Abbildung 24: Messfahrt eines Fahrzeugs vom Standort Essen zu einer Klimastation des DWD. © FiW e. V.

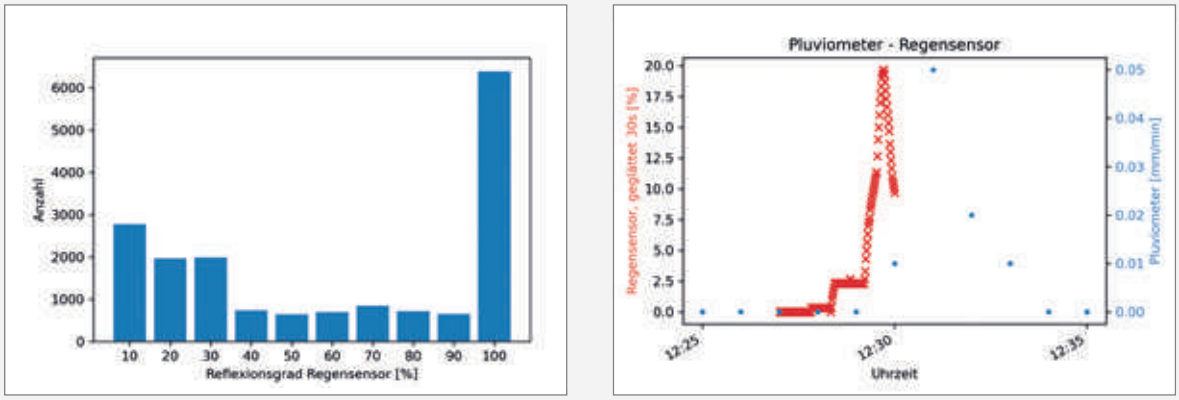


Abbildung 25: Vergleich der Messwerte Regen-Licht-Sensor (links) und Pluviometer (rechts) für eine Messfahrt von Essen zu einer Klimastation des DWD. © FiW e. V.

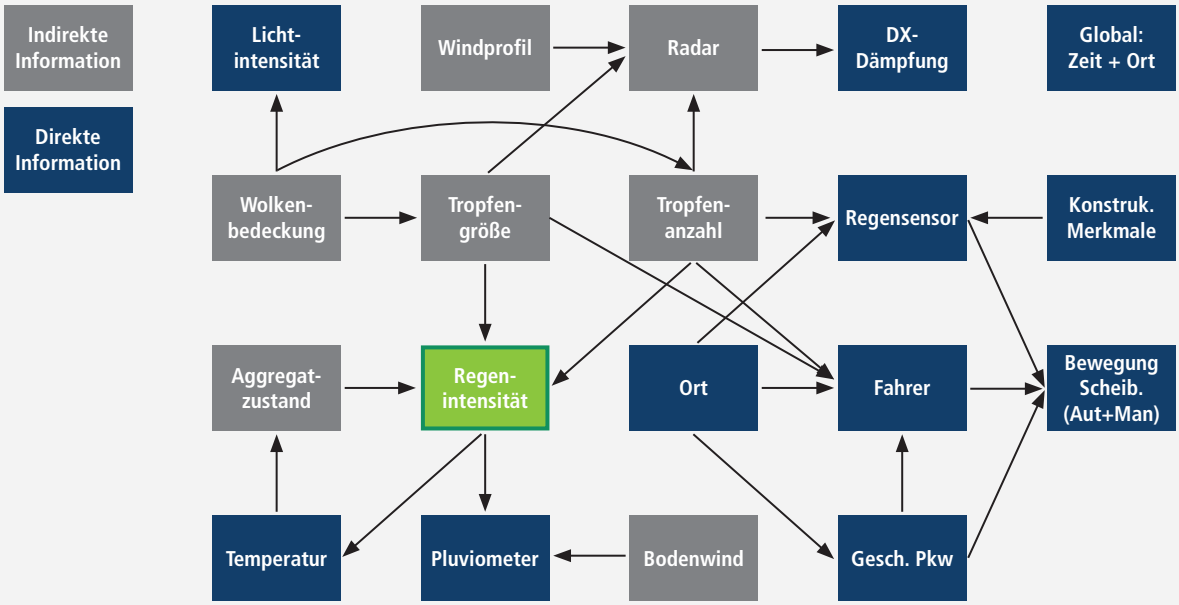


Abbildung 26: In das Informationsmodell eingehende Einflussgrößen. © IAV

Bei Fahrzeugregensensoren kann auf Grund ihrer technischen Konstruktion klar abgeleitet werden, ob es regnet (Kfz-Sensoren messen die Bedeckung der Sensoren durch Regentropfen). Die Intensität kann zudem aus der flächenanteiligen Bedeckung der Sensoren pro Zeiteinheit wie auch aus dem Scheibenwischerhub abgeschätzt werden. Insbesondere bei der Ableitung von Starkregenereignissen sind diese Informationen von erheblicher Bedeutung.

Als Anforderungen an die Rohdaten wurde erörtert, wie im Rahmen der strukturellen Plausibilisierung von eingehenden Rohdaten damit umzugehen ist, sollten Datenpakete nicht vollständig übertragen werden und z. B. GPS-Koordinaten, Zeitstempel oder inhaltliche Sensorinformationen fehlen. Ferner wurde festgelegt, in welchen Fällen Datensätze verworfen werden mussten bzw. noch anteilig genutzt werden konnten. Im Rahmen der Vorabplausibilisierung ist auch die serverseitige Verweigerung der Annahme von eingehenden Datenpaketen zu sehen, sollte bei dem betrachteten Fahrzeug der sog. „Privacy Switch“ aktiviert worden sein. Ebenfalls ist hier die Anwendung des Geofencing im Falle von Messdaten fernab des Untersuchungsgebietes anzusiedeln. Weiterhin wurde ein direkter Zugang für alle Partner zu verfügbaren Rohdaten zwecks fortlaufender Plausibilisierung bzw. zur Entwicklung von Algorithmen eingerichtet. Dieser Aspekt wurde durch die Bereitstellung eines REST-Services auf der IoT-Plattform ermöglicht, sodass dem Konsortium ein direkter Zugriff auf die Daten ermöglicht wurde. So konnten die eingehenden Datenpakete analysiert, auf Verwertbarkeit geprüft sowie in fachlicher Hinsicht auf Unplausibilitäten untersucht werden. Im Ergebnis konnten diverse Parameter aus der weiteren Betrachtung ausgeschlossen und weitere aggregiert werden, um eine homogene Verarbeitung zu gewährleisten.

Lastenheft für die Erstellung des Informationsmodells

Das Informationsmodell war die Basis für die Generierung von Niederschlagsinformationen mittels Fahrzeugsensoren. Auf Basis der verfügbaren Informationen und den definierten Projektzielen wurden Kriterien für den Modellaufbau festgelegt:

- Alle geeigneten Informationen wurden im Informationsmodell verwendet.
- Vertrauensintervalle mussten in den Berechnungen integriert werden.
- Fehlerhafte Fahrzeuginformationen mussten entfernt werden.
- Räumliche und zeitliche Auflösung wurden vom Radar übernommen (1 km x 1 km und 5 Minuten).
- MSPM mussten unmittelbar nach Eingang der Radarinformationen erstellt werden.
- MSPM wurden mit den verfügbaren Informationen erstellt (bei fehlenden Informationen wird trotzdem eine Karte erstellt).

Mit diesen Kriterien wurde eine Modellimplementierung konzipiert (vgl. Abbildung 27). Das Modell integriert dabei die Ergebnisse aus mehreren Teilarbeitspaketen (TA 2.4, TA 2.5 und TA 3.3).

Implementierung von Auswertealgorithmen für die Informationsmodellierung

Die Implementierung wurde in sechs Arbeitsschritte aufgeteilt.

1. State Definition
2. Radar-Modell
3. Pluvio-Modell
4. Car-Modell
5. Rain-Modell
6. Verschneidungs-Modell

State Definition

Hier werden die räumlich- und zeitlichen Parameter der MSPM, welche von den folgenden Modellen (Radar-, Pluvio-, Car-, Verschneidungs-) verwendet werden, definiert.

- Zuweisung der Charakteristika des Synchronisierungs-Frameworks (zeitlich-räumlich).

Umsetzung:

- State Definition wird von allen Modellen aufgerufen, die geographische und zeitliche Auswertungen ausführen.

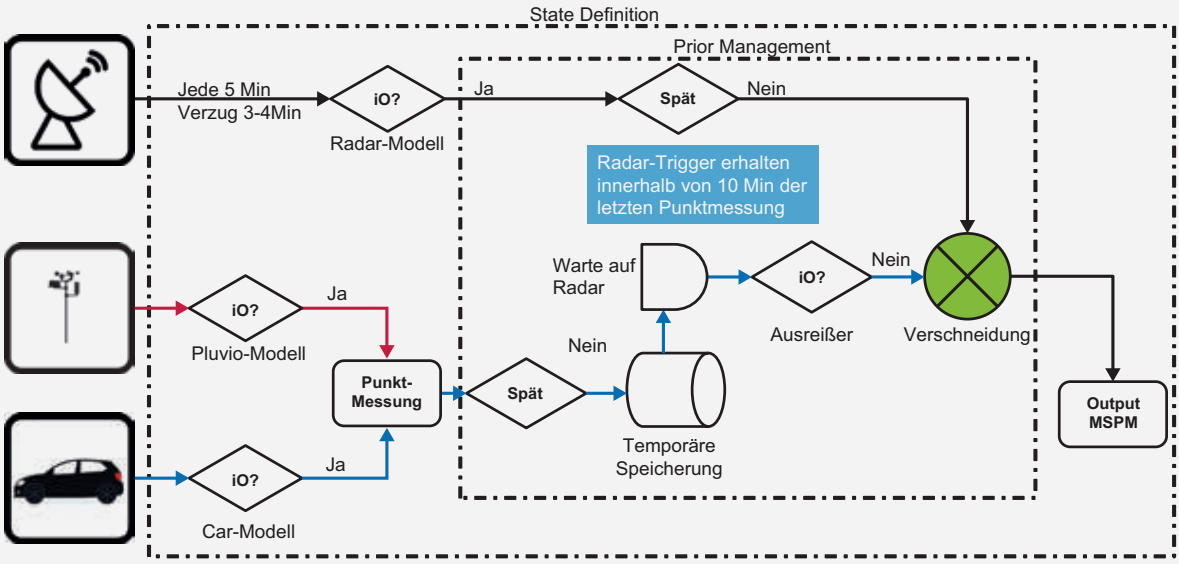


Abbildung 27: Gesamtkonzept des Informationsmodells. © IAV

Tabelle 4: State Definition.

Parameter	Wert
Anzahl von Stützstellen X-Richtung	256
Anzahl von Stützstellen Y-Richtung	256
Seitengröße einer Zelle [m]	1000
Referenzzeit [s]	300
Unterste linke Zelle X-Richtung [m UTM-32N]	230000
Unterste linke Zelle Y-Richtung [m UTM-32N]	5569000

Radar-Modell

Funktion:

- Preprocessing von Niederschlagsradar-Informationen (DX-Dämpfung), die von EGLV bereitgestellt wurden,
- Berechnung des Vertrauensintervalls für jede Zelle

Umsetzung:

- Loop über Messmatrix und Berechnung des Vertrauensintervalls,
- Implementierung der vorgegebenen Vertrauensintervalle.

Tabelle 5: Radar-Modell.

Parameter	Wert
Radarstandort Position X-Richtung [m UTM-32N]	358607
Radarstandort Position Y-Richtung [m UTM-32N]	5696897
Vertrauensintervall bei einer Entfernung geringer als 40 km vom Radarstandort Essen [%]	25
Vertrauensintervall bei einer Entfernung größer als 40 km vom Radarstandort Essen [%]	35
Minimale Unsicherheit [mm/5min]	0.1

Pluvio-Modell

Funktion:

- Preprocessing von Pluviometer-Informationen,
- Berechnung des Vertrauensintervalls für jede Messung

Umsetzung:

- Berechnung des Vertrauensintervalls,
- Implementation der vorgegebenen Vertrauensintervalle,
- Pluviometer werden als Punktmessungen interpretiert.

Tabelle 6: Pluvio-Modell.

Parameter	Wert
Vertrauensintervall [%]	10
Minimale Unsicherheit [mm]	0.01

Car-Modell

Das Car-Modell integrierte die entwickelten Plausibilitätsalgorithmen und Prüfkriterien.

Funktion:

- Preprocessing von Fahrzeug-Informationen:
  - Prüfung der Fahrzeugmessungsvollständigkeit:
    1. Fahrzeug-ID vorhanden
    2. GPS-Zeit und Positionen vorhanden
    3. Regensensorsignal vorhanden
    4. Temperatursignal vorhanden
    5. Lichtintensitätssignal vorhanden
    6. Wischersignal vorhanden
- Geofencing → Fahrzeug befindet sich innerhalb des Einzugsgebietes von EGLV
- Positionierung des Fahrzeugs nach der State-Definition
- Klassifizierung der Ausstattung des Fahrzeugs (Series/Prototype)
- Berechnung des kombinierten Wischerwertes
- Berechnung von abgeleiteten Werten
- Berechnung mit Hilfe des Rain-Modells der Fahrzeug-Regenmenge

Umsetzung:

- Implementierung von konzeptuellen Entscheidungen aus den Iterationsschritten,
- Fahrzeuge werden als Punktmessungen interpretiert.

Rain-Modell

Funktion:

- Berechnung eines Niederschlagwertes für jede Fahrzeugmessung.

Umsetzung:

- Statistische Modellierung von Radarwerten, ermittelt über der zeitlichen und räumlichen Position eines Fahrzeugs mit unterschiedlichen Fahrzeugsignalen.

Für die Ableitung von Niederschlagsinformationen aus den Kfz-Sensordaten war ein Niederschlagsmodell erforderlich, da die Regensensoren von Fahrzeugen selbst bei geringen Regenintensitäten sehr schnell einen Sättigungswert von 100 % erreichten. Zudem war eine direkte Relation zwischen dem Niederschlagsradarwert und den verschiedenen Fahrzeugsignalen nicht offensichtlich erkennbar. Deswegen wurde ein statistisches empirisches Regressionsmodell zur Ableitung von Niederschlagsinformationen ausgewählt.

Kriterien für die Modellierung waren:

- Das Modell wird mit Daten aus der Messdatenbank trainiert und validiert.
- Die Entscheidung des Modells muss nachvollziehbar und transparent sein.
- Das Modell muss ein Vertrauensintervall für jeden Output liefern können.

Hardware für Techniken des maschinellen Lernens

Im Rahmen von mobile[VIEW] wurden große Datenmengen aus drei Hauptquellen (Radarmessungen, Pluviometermessungen und Autosensorik) erhoben. Um aus diesen Daten einen Wissensgewinn zu erzielen, wurden verschiedene Techniken des maschinellen Lernens (ML) angewendet: Entscheidungsbäume, Entscheidungswälder sowie verschiedene Architekturen neuronaler Netzwerke. Das Trainieren verschiedener Architekturen neuronaler Netzwerke, um das beste Ergebnis zu erzielen, erforderte die Anschaffung einer leistungsstarken Workstation. Auf der vorhandenen Hardware hätten nur deutlich kleinere Netzwerke trainiert werden können. Im Bereich der künstlichen Intelligenz (KI) gab es in den letzten Jahren einen starken Innovationsschub in den Bereichen Soft- sowie Hardware, sodass sich ein ganz neues Segment an Grafikkarten-Modellen (GPU) etablieren konnte, die ausschließlich im Bereich KI eingesetzt werden. GPUs besitzen im Vergleich zu CPUs eine hoch parallelisierte Architektur, die ihnen ermöglicht, eine viel höhere Anzahl (kleinerer) Rechenoperationen parallel durchzuführen als CPUs (Lanaro 2017). Gerade im Bereich neuronaler Netze ist der Performancegewinn durch diese Parallelisierung enorm und lässt sich ebenfalls ökonomisch ausdrücken. Vergleiche verschiedener Architekturen neuronaler Netzwerke auf unterschiedlicher Hardware bei Microsoft Azure 2018 zeigten, dass ein GPU-Cluster ein CPU-Cluster in derselben Preisklasse um 415 % in der Leistung übertrifft (Boylu Uz et al. 2018). Die Anschaffung der Workstation machte ein zusätzliches „Housing" der Workstation notwendig. Bei der Nutzung des GPU-Clusters entstanden große Lärm- und Wärmeemissionen, welche vor Ort reduziert oder abgeführt werden mussten. Zusätzlich existierten Anforderungen aus dem Datenschutz, der Serversicherheit und der Redundanz der Systemtechnik. Die große Energieaufnahme, mit insgesamt acht Nvidia RTX 2080 Ti Grafikkarten, und das Sicherheitskonzept der Workstation machten hierbei ein externes Housing erforderlich. Durch den Innovationsgehalt und die nach wie vor geringe Verbreitung von Grafikkarten-Servern, stellte die Workstation neuartige Anforderungen an den Housing-Anbieter. Einige Anbieter konnten sich nicht auf diese Serverstrukturen umstellen. Gleichzeitig bedingte die projektbezogene Arbeit einen stetigen personellen Kontakt mit dem Grafikkartenserver sowohl zur einmaligen Inbetriebnahme als auch zur Wartung. Im Rahmen von mobile[VIEW] wurde mit „RelAix Networks" schließlich ein passender Anbieter im Raum Aachen gefunden, welcher im angemieteten Serverschrank sowohl die benötigte Energiemenge zur Verfügung stellt, Lärm- und Wärmeemission erfolgreich abführt, als auch hohe Anforderungen an die Sicherheitstechnik in redundanten Systemen lieferte.

Entscheidungsbaum

Der oben beschriebene Anforderungskatalog spannt den Rahmen für die Auswahl der Algorithmen innerhalb des Informationsmodells. Die dargelegten Anforderungen legen eine stochastische Auswertung mit sogenannten Entscheidungsbäumen nahe. Ein Entscheidungsbaum ist ein Entscheidungsunterstützungsinstrument, das ein baumartiges Modell von Entscheidungen und ihren möglichen Folgen verwendet, einschließlich der Ergebnisse von Zufallsereignissen, Ressourcenkosten und Nutzen. Die grafische Darstellung in Form eines Entscheidungsbaumes veranschaulicht hierarchisch aufeinanderfolgende Entscheidungen. Diese sind ein beliebtes Hilfsmittel beim maschinellen Lernen, zur automatischen Klassifikation oder zur Ableitung und Darstellung von formalen Regeln. Die flussdiagrammartige Struktur des Entscheidungsbaumes besteht aus einer Reihe von Blättern oder Knoten, welche jeweils einen Test eines bestimmten Attributes darstellen. Das Ergebnis wird in Form von Zweigen dargestellt, auf denen ein weiter Blattknoten mit einer neuen Entscheidung teilweise auf ein neues Attribut folgt. Die Wege von der Wurzel des Baumes zu den Blättern repräsentieren die jeweiligen Klassifikationsregeln. Dies erlaub es dem Nutzer, die Ergebnisse auszuwerten und Schlüsselattribute zu erkennen. Abbildung 28 stellt einen Abschnitt eines Entscheidungsbaums zur Klassifizierung dar.

Der erste Entwurf eines Entscheidungsbaumes auf Basis der im Projekt erhobenen Daten wurde am 25.09.2019 vom FiW in das Projekt eingebracht. Erste Entscheidungsbäume wurden in Form von Klassifikationsmodellen ins Informationsmodell eingebracht. Entscheidungsbäume haben sich seitdem als Schlüsselbaustein im Informationsmodel etabliert, um Korrelationen zwischen Kfz-Sensoren und Niederschlagsereignissen zu betrachten. Entscheidungsbäume können abhängig von der gewünschten Komplexität unterschiedliche Tiefen annehmen und eine stetige Korrelation der Kfz-Sensorwerte abbilden. Somit lassen sich genauere Rückschlüsse auf Niederschlagsintensitäten, welche im Informationsmodell mit Pluviometer- und Radardaten zusammengeführt werden können, ziehen. Abbildung 29 zeigt exemplarisch einen Entscheidungsbaum aus dem Kfz-Modell, welches die Kfz-Sensordaten den entsprechenden Niederschlagsintensitäten zuordnet.

Eine Vielzahl von unkorrelierten Entscheidungsbäumen ergeben einen Zufallswald (Random Forest). Der Begriff Random Forest wurde durch Leo Breiman geprägt (Breiman 2001). Der Aufbau von Entscheidungsbäumen ist zum Teil durch Zufallsvariablen geprägt, wodurch diese unterschiedliche Gewichtungen ausprägen können. Bei der Regression oder Klassifikation durch Zufallswälder fließt die Entscheidung jedes Baumes ein und definiert über die Summe das Endergebnis des

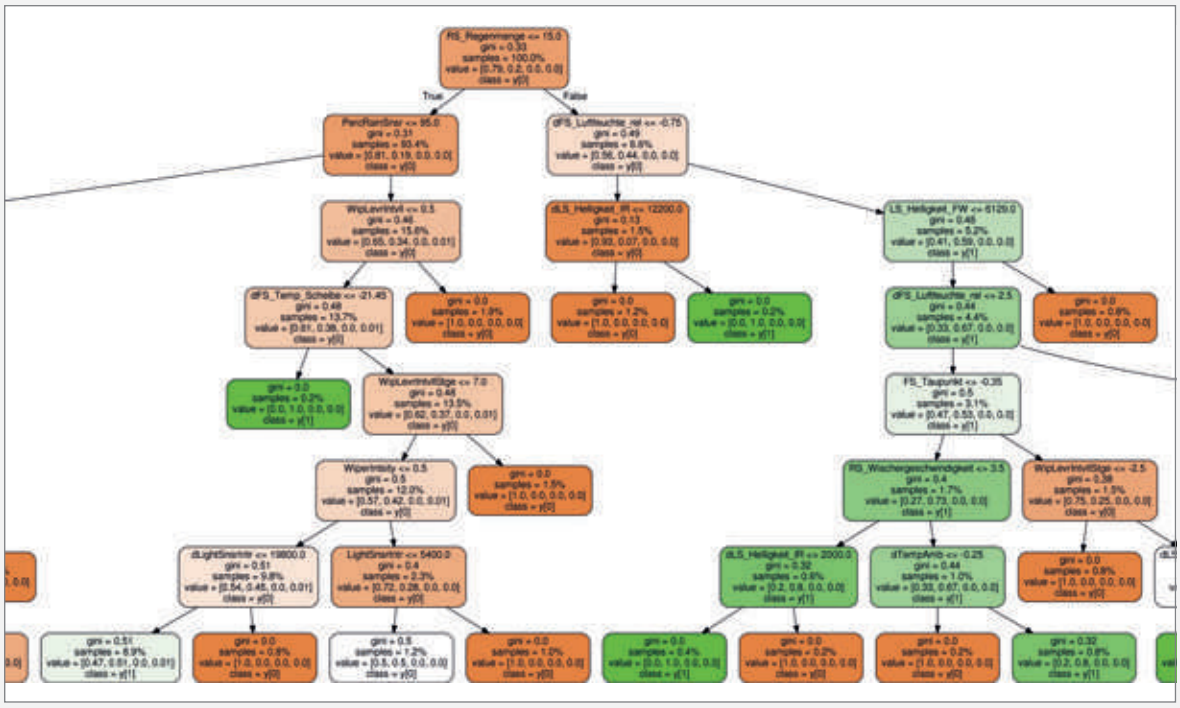


Abbildung 28: Abschnitt eines Entscheidungsbaums zur Klassifizierung von Niederschlagsintensitäten. © FiW e. V.

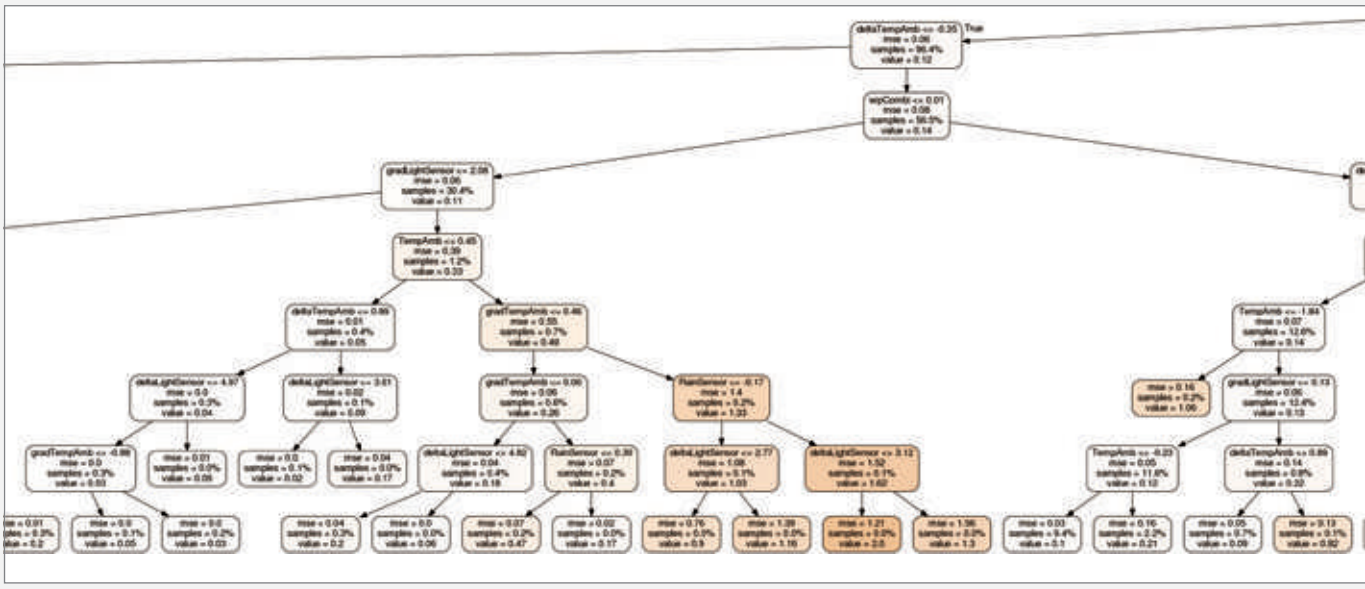


Abbildung 29: Teil eines Entscheidungsbaums zur Regression innerhalb des Kfz-Modells. © FiW e. V.

gesamten Waldes. Zufallswälder als Komposition vieler Entscheidungsbäume erhöhen statistisch gesehen die Genauigkeit des Modells und können eine bessere Generalisierung ermöglichen. Zufallswälder helfen dabei, die Ergebnisse des Kfz-Modells zu plausibilisieren, indem sie die Abhängigkeit zu einzelnen Entscheidungsbäumen reduzieren und grobe Simplifizierungen der Entscheidungsbäume methodisch ausschließen. Durch einheitliche Zielgrößen und den Vergleich zu den tatsächlichen Werten von Messfahrten, Pluviometern und Niederschlagsradaraufzeichnungen können die unterschiedlichen Modelle ausgewertet und miteinander verglichen werden. Ein weiterer Vorteil von Entscheidungsbäumen und Zufallswäldern liegt in der inhärenten Integration von Konfidenzintervallen in Form von Gini-Koeffizienten für jeden Knoten eines Entscheidungsbaumes. Hierdurch lassen sich die Ergebnisse des Kfz-Modells mit Niederschlagswerten der Radar- und Pluviometerstationen verschneiden.

Neuronale Netze

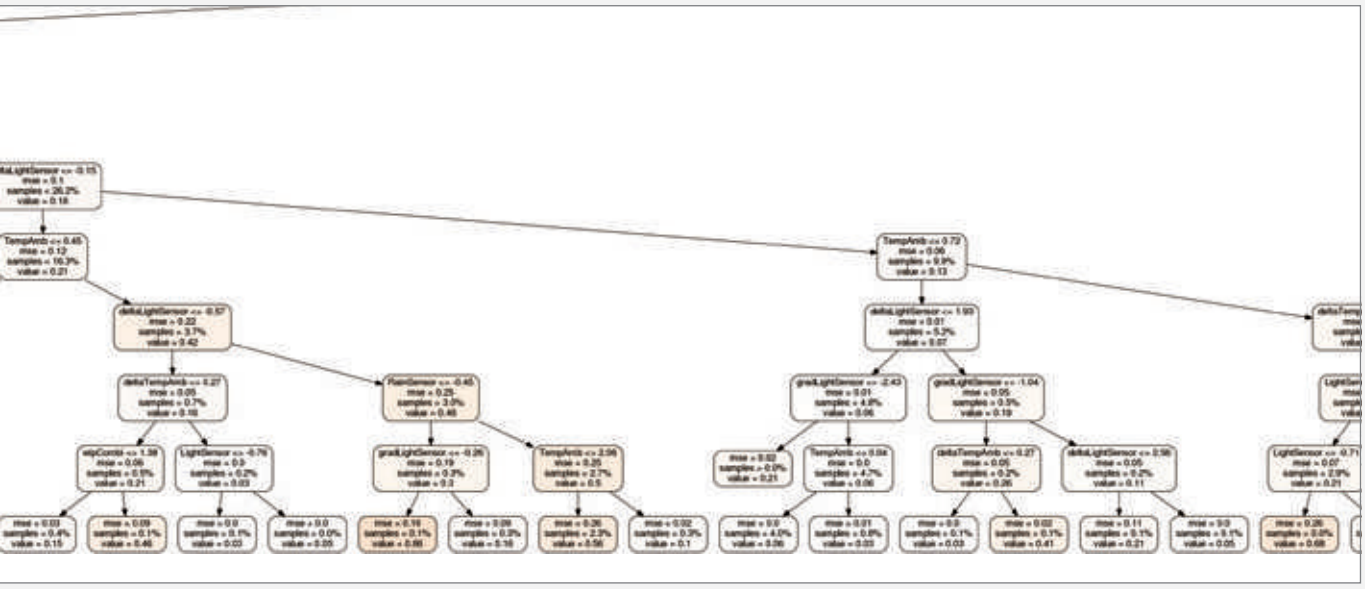
Eine weitere Methode zur Erstellung von Vorhersagemodellen sind künstliche neuronale Netzwerke – KNNs (auch neuronale Netzwerke – NN, artificial neural networks – ANNs). Diese sind ein Teil der künstlichen Intelligenz, insbesondere ein Teil des maschinellen Lernens. Die genauen Anwendungen von neuronalen Netzwerken können unterschiedlich sein. Ein häufiges Beispiel ist ein Klassifizierungsproblem, in dem das neuronale Netzwerk anhand von Eingabedatensätzen lernen soll, ähnlichen ungesehenen Datensätzen eine Kategorie zuzuweisen. Ein Beispiel ist das automatisierte Kategorisieren von Bildern, die verschiedene Tiere darstellen.

Neuronale Netzwerke haben verschiedene Architekturen. Eine bekannte und einfache Architektur ist das mehrlagige Perzeptron (Multi-Layer Perceptron, MLP), das in Abbildung 30 zu sehen ist. Das MLP ist ein Beispiel für ein Feedforward Neural Network (FFNN), dessen Ablauf aus iterativen Berechnungen durch die Schichten von Eingabe zur Ausgabe besteht.

In der Abbildung ist von links nach rechts abgebildet: Eingabeschicht (blau), Hidden-Schichten (grün), Ausgabeschicht (lila). Die Kreise in den Hidden-Schichten des Netzwerks sind Neuronen (auch neurons, nodes). Die Kreise repräsentieren hier jeweils eine Zahl, die durch Eingabe oder Berechnungen zustande kommt.

Die Eingabedatensätze sind meist eine Reihe von Merkmalen (auch Features), die einen rohen Datensatz erweitern oder davon abstrahieren. Zum Beispiel, wenn der Preis eines Hauses aus verschiedenen Merkmalen wie Zimmeranzahl, Quadratmeter, Lage, Baujahr, usw. ermittelt werden soll, müssen die Zahlen aufgrund der Größenordnungen skaliert werden, weil größere Zahlen wie Baujahr das Lernen des Netzwerks schwerer machen.

Die Hidden-Schichten transformieren schrittweise die Eingabe durch lineare Transformationen und Nichtlinearitäten bis zur Ausgabeschicht. Die Transformationen sind Matrizenmultiplikationen zwischen den Ausgaben der vorherigen Schicht und den durch das Trainieren ge-



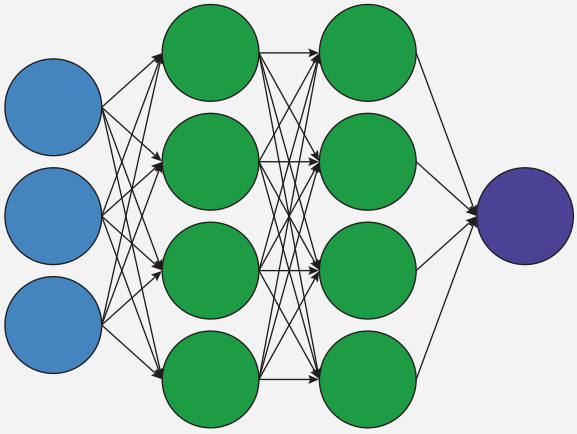


Abbildung 30: Beispielhafte Architektur eines mehrlagigen Perzeptrons. © FiW e. V.

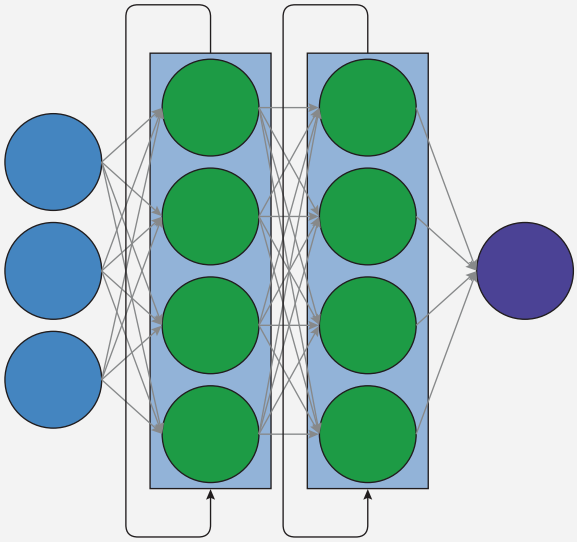


Abbildung 31: Architektur eines rekurrierenden neuronalen Netzwerks. © FiW e. V.

lernten Gewichten (Pfeile in Abbildung 30). Die Nichtlinearitäten folgen der Transformation und sind einfache nichtlineare Aktivierungsfunktionen, die auf die Ausgabe der Transformation verwendet werden. Diese Nichtlinearitäten sind nötig, damit das Netzwerk auch nichtlineare Funktionen lernen kann.

Eine weitere relevante Art von neuronalen Netzwerken sind rekurrierende neuronale Netzwerke (recurrent neural networks, RNN). Durch ihre Rückkopplungen über mehrere Eingaben sind sie besonders geeignet für Eingabesequenzen. Ein Beispiel ist die Spracherkennung, wo vorherige erkannte Wörter nützlich sein können, um das aktuelle Wort zu erkennen.

In Abbildung 31 ist grob dargestellt, wie diese Rückkopplung funktionieren kann. Dabei wird die Ausgabe der vorherigen Berechnung der Hidden-Schicht als zusätzliche Kontextinformation für die aktuelle verwendet. Es gibt eine besondere Architektur von rekurrierenden neuronalen Netzwerken, die als langes Kurzzeitgedächtnis (long short-term memory, LSTM) bezeichnet wird. Durch weit entfernte Abhängigkeiten zwischen verschiedenen Eingaben, kann es bei normalen RNN dazu führen, dass das Netzwerk nicht mehr lernt (vanishing gradient problem). Das lange Kurzzeitgedächtnis hilft, indem es lernt, Werte von früheren Abhängigkeiten zu speichern, die dann für spätere Ausgaben wichtig sind. Dieses Speichern ermöglicht arbiträre Entfernungen zu früheren Abhängigkeiten.

Die größten Vorteile von neuronalen Netzwerken sind ihre Schätzfähigkeit, Einsatzflexibilität, und Allgegenwart in der heutigen Welt. Mit einer Hidden-Schicht bestehend aus ausreichend vielen Neuronen kann ein NN theoretisch jede Funktion ausreichend gut approximieren. Dadurch können neuronale Netzwerke für viele verschiedene Zwecke sehr gute Ergebnisse liefern. Heutzutage sind neuronale Netzwerke sehr beliebt, sodass das Entwerfen von neuen Modellen ohne sehr großen Aufwand denkbar ist. Leider sind neuronale Netzwerke nicht ohne Nachteile. Der größte Nachteil von neuronalen Netzwerken, im Vergleich mit z. B. Entscheidungsbäumen, ist die mangelnde Transparenz und Erklärbarkeit der Ergebnisse.

Die größte Herausforderung beim Trainieren des neuronalen Netzwerkes war der Datensatz. Der Testdatensatz wurde vor dem Trainieren aus dem Gesamtdatensatz blockweise extrahiert, um zu vermeiden, dass viele der Daten im Testdatensatz beim Trainieren schon gesehen werden. Die restlichen Daten wurden fünf Mal jeweils in 80 % Trainingsdatensatz und 20 % Validierungsdatensatz aufgeteilt (5-fache Kreuzvalidierung). Standardisierung, Min-Max Normierung und Einheitsvektor-Normierung wurden alle als Featureskalierungsmethode ausprobiert.

Im ersten Schritt wurden MLP-Modelle angewendet. Die Eingabeschicht bestand aus einem Feature-Vektor, der einen bestimmten Zeitpunkt repräsentierte. Danach war eine verschiedene Anzahl von Hidden-Schichten und Neuronen pro Schicht implementiert. In der Ausgabeschicht wurden zwei Arten von Ausgaben ausprobiert: Eine, die verschiedene Wertebereiche als Klassen darstellte, und eine mit einer kontinuierlichen Ausgabe. Als Aktivierungsfunktion wurden RELU und Sigmoid verwendet.

Einige MLP-Modelle schienen vielversprechend zu sein, allerdings wurden später einige mögliche Fehlerquellen in der Datenvorbereitung gefunden. Die Genauigkeit der Netzwerkarchitekturen konnte iterativ verbessert werden. Das Netzwerk konnte zwischen Regen und Nicht-Regen entscheiden, aber eine genauere Einstufung wurde durch die teils quasi-binären Sättigungswerte der Regen-Licht-Sensoren erschwert. Der nächste logische Schritt war zu untersuchen, ob eine Sequenz von Messvektoren eine bessere Regenvorhersage für den aktuellen Zeitpunkt lieferte. Dabei wurden verschiedene Sequenzlängen mit bis zu 100 Messvektoren mit einem LSTM RNN ausprobiert. Das LSTM besitzt eine ähnliche Architektur wie das MLP mit eingebauter Rückkopplung und einer Sequenz als Eingabe. Leider führte die Sequenzierung nicht zu den gewünschten Ergebnissen und eine hohe Genauigkeit beschränkt sich auf die binäre Unterscheidung zwischen Regen und kein Regen.

Für eine uniforme Bewertung wurde die mittlere quadratische Abweichung (mean squared error – MSE), das mittlere Abweichungsquadrat (root-mean-square error – RMSE), die kumulative Abweichung und der Vergleich zu historischen Daten herangezogen. Hierdurch konnten die unterschiedlichen Kfz-Modelle unabhängig voneinander ausgewertet werden. Random Forests und künstliche neuronale Netze helfen somit dabei, die Ergebnisse der Kfz-Modelle zu plausibilisieren und auf Ausreißer und andere Auffälligkeiten zu untersuchen. Für die Integration ins Informationsmodell und die Anwendung im Hochwasservorhersagesystem (Delft-FEWS) von EGLV sind Verständlichkeit und Erklärbarkeit der Ergebnisse von entscheidender Bedeutung, weshalb das Konsortium sich für die Verwendung von Entscheidungsbäumen als Grundlage für das Kfz-Modell entschied. Die Modelle der neuronalen Netze und Zufallswälder ermöglichten statistischen Plausibilisierungen und trugen zur Verbesserung des finalen Informationsmodells bei. So konnten die Ergebnisse nachhaltig verbessert werden.

TA 2.4 Erweiterung 1 des Datenbanksystems, Plausibilisierung der Rohdaten

Die Plausibilisierung der Rohdaten ist ein integraler Bestandteil jeder Datenanalyse und zur Validierung der Modelle mittels Referenzmessungen erforderlich. Die unterschiedlichen Rohdaten weisen in Abhängigkeit der Datenquellen verschiedene räumliche und zeitliche Auflösungen auf. Radarwerte werden im 5-minütlichen Intervall erhoben, Pluviometerstationen übertragen die Rohdaten in minütlich, wohingegen die Kfz-Werte sekündlich erfasst und übertragen werden. Zudem verfügt jede Datenquelle über einen eigenen Zeitstempel mit teilweise unterschiedlichen Zeitformaten. Bei den Rohdaten der Pluviometer musste zusätzlich die Umrechnung von Central European Time (CET) in Coordinated Universal Time (UTC) berücksichtigt werden. Zur Plausibilisierung der Rohdaten wurden weitere Datenquellen hinzugezogen, welche zwar zur Validierung genutzt wurden aber keinen Eingang in das finale Informationsmodell fanden. Zudem unterscheidet sich die räumliche Auflösung der Daten stark. Während die rasterbasierten Niederschlagsradardaten einen Niederschlagsintensitätswert pro Quadratkilometer liefern, handelt es sich bei den Pluviometer- und den Kfz-Daten um Punktmessungen in der jeweiligen Zelle. Bei der Qualitätssicherung spielt die Rohdatenqualität eine entscheidende Rolle. Pluviometer und Radarstationen werden regelmäßig geeicht und auf Fehlwerte der Niederschlagsintensität überprüft, wohingegen Sensorwerte der Kfz primär für den Fahrkomfort und Sicherheit genutzt werden und Niederschlagsintensitäten durch das Kfz-Modell abgeleitet werden müssen. Während zur Abschätzungen der Unsicherheitsbereiche der Radar- und Pluviometerdaten jeweils entsprechende Literaturwerte zum Einsatz kommen, musste das Validierungskonzept für Kfz-Werte im Projekt erarbeitet werden. Eine Herausforderung hierbei war die Definition einer geeigneten „ground truth“, an der die Kfz-Sensoren entsprechend geeicht werden konnten.

Die Verwendung der GPS-Signale der Datenboxen stellten die räumliche Genauigkeit der Fahrzeug-Lokalisierung sicher (unter 10 Meter). Die geographischen Angaben für Pluviometer und Radardaten wurden als exakt angenommen. Weitere Unsicherheiten ergaben sich allerdings aufgrund der verschiedenen räumlichen Auflösungen der unterschiedlichen Produkte und der Zusammenführung in ein gemeinsames räumliches Bezugssystem.

Abbildung 32 zeigt einen Abgleich der Lage von bereitgestellten Radardaten zwischen EGLV und IAV zur Sicherstellung der korrekten Interpretation der Daten durch IAV.

Bei der Plausibilisierung der Kfz-Werte wurde eine vollständige Datenreihe gefordert, da beispielsweise Kfz-Werte ohne Georeferenzierung (während der Initialisierung des GPS) nicht mit anderen Datenquellen verschnitten werden konnten. Zudem wurden einzelne Sensorwerte anhand von Literaturwerten auf ihre Fehleranfälligkeit geprüft. Ausreißer beim Erfassen der Lichtintensität oder der Umgebungstemperatur konnten so direkt erkannt und im weiteren Verlauf automatisch behandelt werden.

Die Latenzzeit der Fahrzeugsensoren betrug immer weniger als eine Sekunde. Problematisch waren die schnellen Signalfluktuationen. Um die Signale zu stabilisieren, wurden einige Signale über 30 Sekunden gemittelt (Abbildung 33). Die Stabilität war für die statistische Modellierung wichtig, denn sprunghafte Messungen führen zu Modellunsicherheiten.

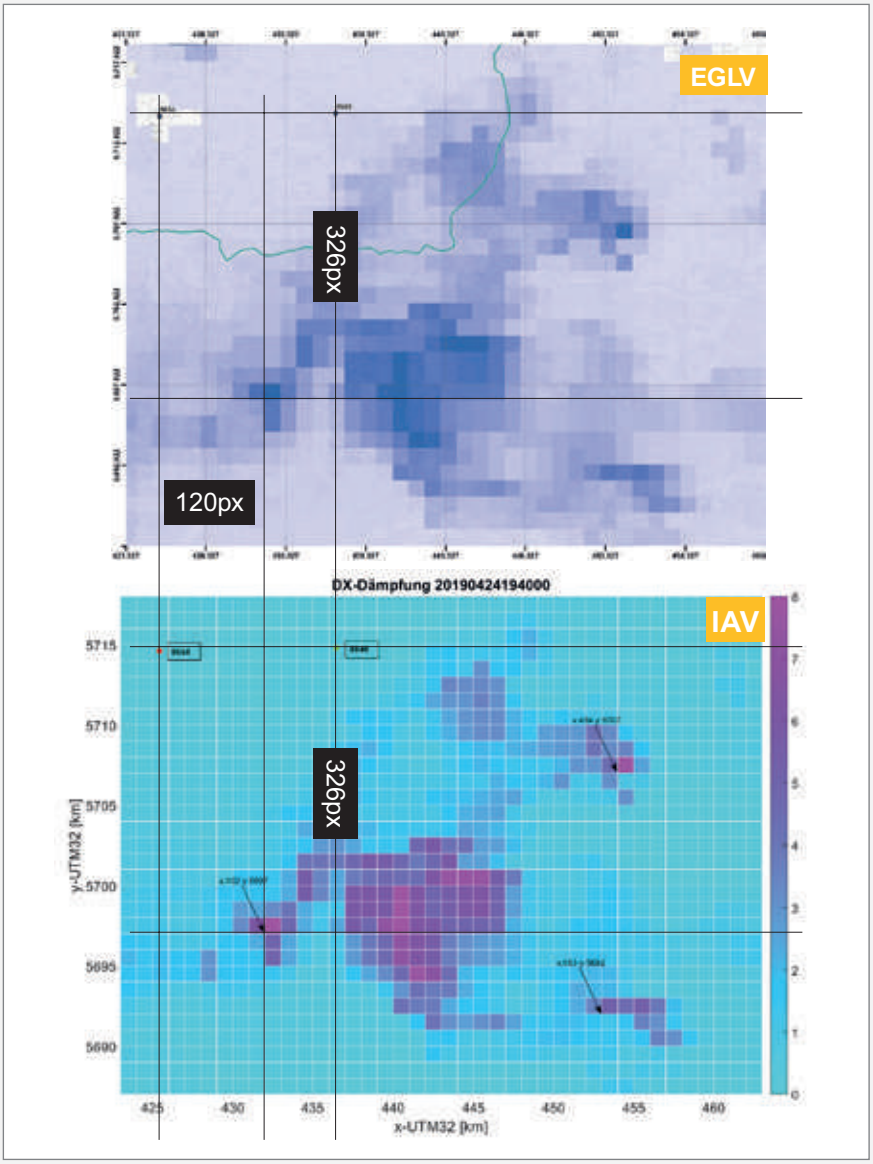


Abbildung 32:
Vergleich einer Informations-positionierung von FEWS und dem Informationsmodell.
© FiW e. V.

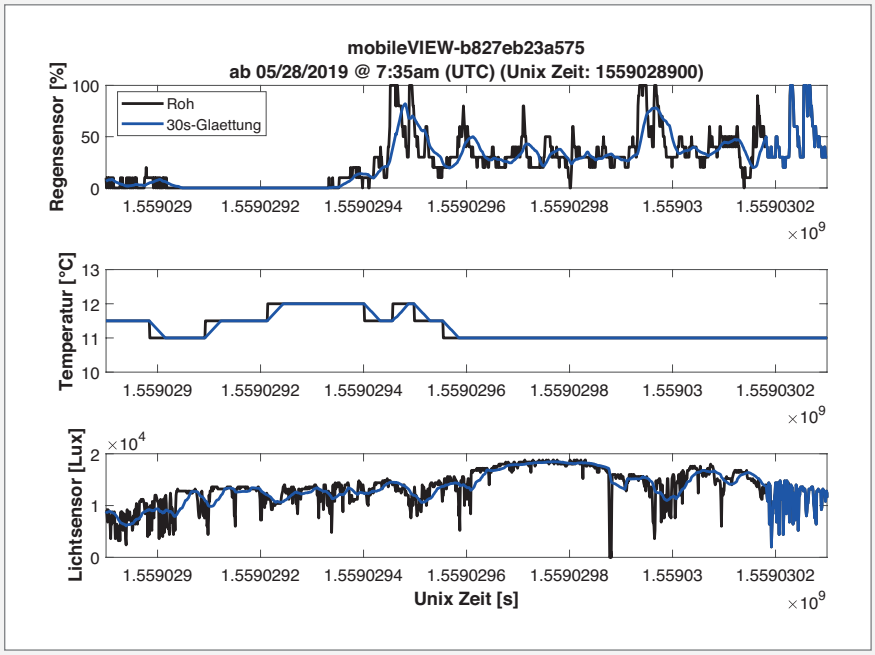


Abbildung 33:
Gemittelte Rohsignale am 28.05.2019. © FiW e. V.

TA 2.5 Qualitätsmanagement der erhobenen und bearbeiteten Daten

Innerhalb des Projekts wurden Prüfkriterien für Rohdaten, prozessierte Daten und Zusatzin-formationen entworfen, um Auffälligkeiten und mögliche Fehler frühzeitig zu identifizieren und ein verlässliches Endprodukt von aggregierten Niederschlagsintensitäten zu erstellen. Teil der Prüfung und Qualitätskontrolle ist z. B. die Definition von Ausreißer-Kriterien und das Eliminieren von Ausreißern (physikalisch oder logisch) nicht valider Werte. Ausreißer in den Eingangsdaten wurden durch Glättung behandelt. Das Chauvenet-Kriterium wurde angewandt, um Ausreißer in den durch das Informationsmodell erzeugten Ergebnissen zu eliminieren. Dabei wird ein zu erwartender Wertebereich um den Mittelwert der Ergebnisse bestimmt.

Unter der Annahme, dass die Daten normalverteilt sind, kann die maximale Abweichung vom Mittelwert mittels (1) bestimmt werden

Formel 1: Das Chauvenet-Kriterium.

$$\tau = \frac{|X_i - \bar{x}|}{\sigma}$$

Zur Plausibilisierung des Informationsmodells wurden verschiedene Modell-Versionen mit teils unterschiedlichen Modellarchitekturen iteriert und in stetiger Rekursion den Anforderungen im Betrieb angepasst. Es konnten somit Prüfkriterien für das Qualitätsmanagement definiert werden. Ein weiterer Meilenstein wurde mit einer automatisierten Datenverarbeitung und einer Datenschnittstelle zwischen IAV, EGLV und dem Hochwasservorhersagesystem Delft-FEWS erreicht.

Die Sensorqualität der Fahrzeuge konnte nicht beeinflusst werden. Der Fokus der Qualitätskontrolle lag deshalb auf dem sogenannten Quality-of-Service (QoS). Diese Auswertung befasste sich damit, die folgenden Fragen zu klären:

1. Wie oft sind die Fahrzeug-Einträge vollständig?
2. Welche Fahrzeugsignale werden nicht richtig erfasst?
3. Wie oft war der GPS-Tracker ohne Funktion?
4. Wann hat sich ein Fahrzeug das letzte Mal gemeldet?

Diese Analysen wurden regelmäßig ausgeführt. Abbildung 34 zeigt beispielhaft das Ergebnis einer Analyse für Daten aller in diesem Zeitraum operierender Fahrzeuge zwischen dem 12.02.2019 und dem 19.02.2019. Im Ergebnis der ausgeführten und bewerteten QoS-Analyse stellte sich heraus, dass eine andere Konfiguration für einige Fahrzeuge erforderlich war.

TRW/BCS und Valeo sind verschiedene Hersteller bzw. Zulieferer der Kfz-Sensoren.

	Total	Serien	TRW/BCS	Valeo	Unklar
Count	1070117	747749	200231	30812	91325
Complete	648233	472338	145446	2271	28178
QoS Ratio	0.61	0.63	0.73	0.07	0.31

Gewichtete Maßnahmen, angenommen ist eine 90% Zuverlässigkeit nach der Änderung:

- neue Konfiguration Seat Alhambra, VW Sharan, Audi A6: +14,8 %
- geänderte Konfigurationen für Testzweck zurücksetzen: +6.3 %
- VW-Passat Sensorstatus prüfen: +3.3 %
- Serien Seat Leon (nicht ST) Sensorstatus prüfen: +2.5 %
- Unklare Ursache. Neue Konfiguration ermal probieren. Potential +2.1 %

Abbildung 34: QoS-Auswertung für den Zeitraum 12.02.2019 – 19.02.2019. © IAV

TA 2.6 Erweiterung 2 des Datenbanksystems, Datenbereitstellung in Modellen

Die Datenbereitstellung im Web-Interface wurde in kontinuierlichen Iterationsschritten an die Anforderungen einer projektinternen Nutzung angepasst. Unter Verwendung von OpenStreetMap-Inhalten wurde das EGLV-Verbandsgebiet dargestellt. Die Karte erhielt außerdem eine Zoom-Funktion. Die zeitliche Betrachtungsebene wurde an die Erhebungsintervalle der Niederschlagsradardaten angepasst [5 min]. Über eine Auswahlmöglichkeit konnten Live-Daten oder historische Daten betrachtet werden. Die zu dem Zeitpunkt aktiven Fahrzeuge konnten als Punkte auf ihrem Weg verfolgt werden. Durch eine MouseOver-Funktion konnten diverse aktuelle Fahrzeugparameter für jeden erzeugten Datensatz angezeigt werden.

Der Webservice stellte für alle Projektpartner die Schnittstelle zur Verfügung, an der die Fahrzeugmessungen und die MSPM visualisiert und heruntergeladen werden konnten. Der Webservice war unter https://iot-board.iavtech.net für berechtigte Nutzer erreichbar (Abbildung 35).

Die IoT-Plattform wurde von Anfang an als erweiterbar konzipiert. Als zum Beispiel neue Speicherkapazität erforderlich war, konnte diese innerhalb kurzer Zeit integriert werden. Nach dem Login wurden unterschiedliche Optionen angeboten: z. B. Kartenvisualisierung, eine Statistik über Regentage oder eine User-definierte Konfiguration. Für die Kartenvisualisierung gab es dabei mehrere Konfigurationsoptionen (Abbildung 36).

- Live-Daten oder Daten aus der Vergangenheit,
- Karte mit Fahrzeugen, Radar (DX Daten), Pluviometer und MSPM,
- Exportfunktionen.

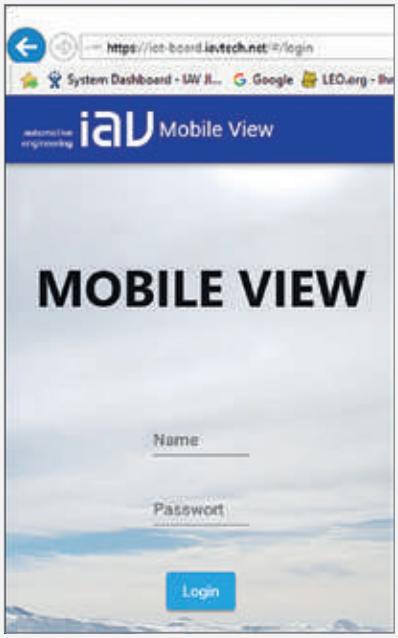


Abbildung 35: Login-Maske Webserver für interne Projektmitarbeiter. © IAV

Abbildung 36: Visualisierungsoptionen für die Kartendarstellung im Webservice. © IAV

Die Karten (Abbildung 37) wurden als Niederschlagsfelder über einer topografischen Karte dargestellt. Dabei waren verschiedene Verknüpfungen und Variationen in der Darstellung möglich (alleinige oder gleichzeitige Darstellung von MSPM und Niederschlagsradardaten, Einstellung unterschiedlicher Farbskalen, Zuschaltung von Ergebnissen aus Pluviometermessungen, …).

Der Zweck der Darstellung war abzubilden, wo die abgeleiteten Informationen aus Fahrzeugen und Pluviometern eine resultierende MSPM verändern würden. Eine solche Veränderung ist zum Beispiel die räumlichzeitliche Erweiterung einer Niederschlagzone als Grundlage für ein Meldesystem.

Weiterhin wurde ein Geofencing implementiert, welches die Datensätze ausblendet, die in einer Entfernung von über 30 km zum EGLV-Verbandsgebiet aufgenommen wurden. Als Drittdaten wurden die dämpfungskorrigierten DX-Niederschlagsradardaten von EGLV sowie die Pluviometerstandorte von EGLV integriert. Die Niederschlagsintensität des Radarproduktes wurde durch eine entsprechende Färbung angezeigt.

Der Webservice ermöglichte das Herunterladen von Daten über die Webseite. Das Herunterladen konnte auch über eine programmierbare Schnittstelle erfolgen, die sogenannte REST-API. Diese Schnittstelle war für den Transfer von großen Datenmenge oder einen automatischen Transfer geeignet. Zum Beispiel würde der folgende Befehl (Link 1) alle Daten zwischen den 9. Nov 2020 06:00 bis 9. Nov 2020 18:00 liefern.

Die WebGIS-Oberfläche konnte jedoch nicht zur Visualisierung von Projektergebnissen für die Öffentlichkeit genutzt werden. Da es sich bei der Darstellung der Fahrzeugrouten um personenbezogene Daten handelte, war eine Veröffentlichung außerhalb der Forschungsfragen des Konsortiums aus datenschutzrechtlichen Gründen nicht möglich. Das Konsortium fokussierte stattdessen die Veröffentlichung der Projektergebnisse über die mCLOUD, die vom BMVI im Rahmen der mFUND-Förderrichtlinie betrieben wird. Dementsprechend wurde eine Darstellung von Vorhersagemodellergebnissen über die IoT-Plattform nicht weiterverfolgt.

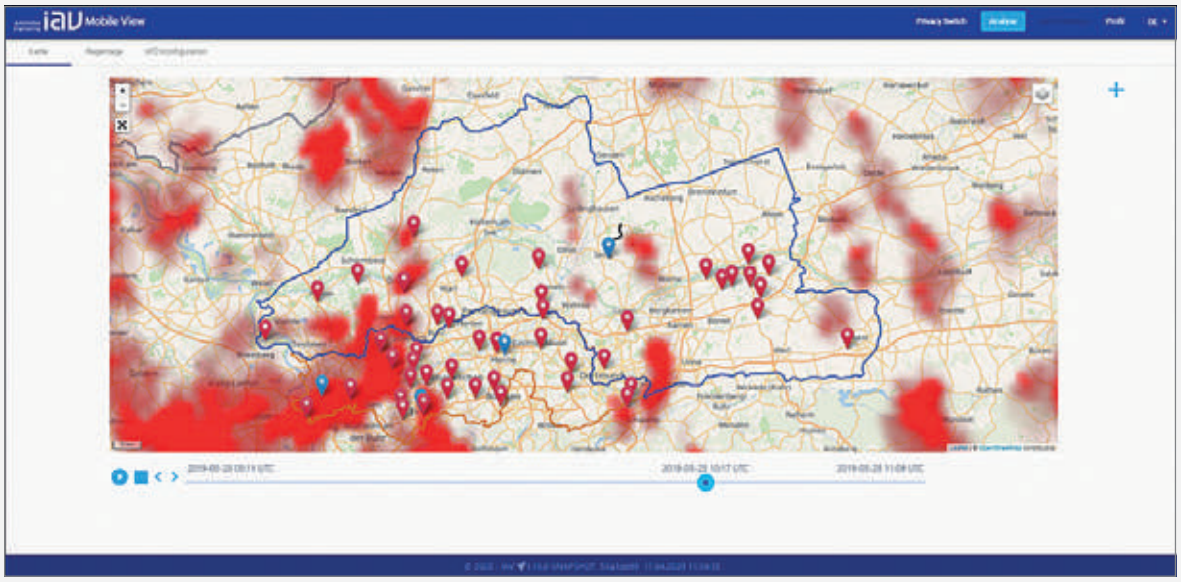


Abbildung 37: Kartendarstellung von Daten der Kfz, Pluviometer und des Niederschlagsradars im Webservice. © IAV

https://iot-board.iavtech.net/api/measurement/carlids/2020-09-11:06:00:00/2020-09-11:18:00:00

Link 1: Beispielbefehl über die REST API (inaktiv).

Für die Veröffentlichung von Daten in der mCLOUD entwickelte das Konsortium ein eigenes Produkt, da sich auch das Endprodukt des Informationsmodells aufgrund der Verschneidung mit Niederschlagsradar- und Pluviometerdaten nicht für eine Darstellung der erhobenen Kfz-Daten eignete.

Das Produkt für die mCLOUD wurde zwecks Anonymisierung auf die Kacheln des verwendeten Niederschlagsradarproduktes übertragen und befand sich somit in einem kartesischen Koordinatensystem mit einer Kantenlänge von 1 km. Als zeitliche Auflösung wurde die Auflösung des Niederschlagsradarproduktes gewählt, dementsprechend fünf Minuten. Im finalen Produkt wurde so für jede Rasterzelle eine Aggregierung aller während der fünf Minuten in einem Rasterfeld aufgenommenen Kfz-Datensätze vorgenommen. Eine Einsicht in alle für die Aggregierung relevanten Rohdatensätze wurde ebenfalls ermöglicht. Als Georeferenzierung wurde aus datenschutzrechtlichen Gründen nur jeweils die entsprechende Rasterkachel und nicht die exakte GPS-Position des Fahrzeuges angegeben.

6.1.3 Teilprojekt TP 3: Datenverwertung

TA 3.2 Integration von Drittdaten in die Datenbank – Erweiterung 3 des Datenbanksystems

Im Teilprojekt 3, „Datenverwertung", führte das Konsortium die erhobenen Daten mit weiteren Drittdaten zusammen, um einen Mehrwert für wasserwirtschaftliche Anwendungen zu generieren. Im ersten Schritt fand eine Recherche zu verfügbaren Drittdaten statt. Die Drittdaten wurden unterteilt in Daten, die aus dem Kontext dem BMVI stammen, weitere frei verfügbare Drittdaten aus öffentlichen und privaten Quellen sowie nicht frei verfügbare Drittdaten. Im Folgenden werden die Ergebnisse der Datenrecherche dargestellt:

Berücksichtigung fanden im Zuge der Recherche insgesamt 19 verschiedene Satelliten aus den Programmen Copernicus (ESA), EPS und MTP (EUMETSAT), GOES, Landsat und EOS (NASA). Festgestellt wurde hier jedoch, dass die raum-zeitliche Skala von Satellitenprodukten grundsätzlich nur schwerlich vereinbar war mit der hochfrequenten und räumlich genauen Datenübertragung von Kfz-Sensordaten. Generell kann in Bezug auf die Umlaufbahn zwischen sonnensynchronen und geostationären Satelliten unterschieden werden. Während sich erstere unter Berücksichtigung von Bahnhöhe und Fluggeschwindigkeit Aufnahmen in relativ hoher räumlicher, je nach Fragestellung jedoch verhältnismäßig schlechter zeitlicher Auflösung von bis zu 16 Tagen auszeichnen, befinden sich geostationäre Wettersatteliten in einer Bahnhöhe von ca. 36.000 km. Dies erlaubt ihnen zwar ca. alle fünf bis 15 Minuten eine neue Aufnahme zu tätigen, angesichts der Entfernung zwischen Aufnahmeinstrument und oberflächennaher Atmosphäre sind je nach verwendetem Kanal des elektromagnetischen Spektrums jedoch nur räumliche Auflösungen von teils ca. 4,5 km x 4,5 km zu erzielen. Beide Optionen schienen sich daher für das Forschungsvorhaben nur wenig zu eignen und wurden daher nach erfolgter Bestandsaufnahme nicht weiterverfolgt.

An dieser Stelle fand eine Unterscheidung zwischen den vom BMVI bereitgestellten Daten und weiteren frei verfügbaren Daten (z. B. DWD) statt. Die frei verfügbaren öffentlichen Daten im Kontext des BMVI waren nur sehr bedingt für eine Verwendung geeignet. Zu Projektbeginn stellte die mCLOUD lediglich eine Metadatenplattform dar, über die keine Inhalte bezogen werden konnten. Die in 2018 verfügbaren enthaltenen Daten eigneten sich meist aufgrund ihres jeweiligen Betrachtungsgebietes und der fehlenden Konstanz der Verfügbarkeit in einer hohen zeitlichen Auflösung nicht für eine Integration in mobile[VIEW].

Unter die darüber hinaus frei verfügbaren Daten aus öffentlichen Quellen fielen die Niederschlagsradarprodukte des DWD, die seit 2017 öffentlich verfügbar sind. EGLV-intern findet eine Weiterverarbeitung des DX-Niederschlagsradarproduktes des DWD statt. So wird u. a. das vom DWD verwendete polare Koordinatensystem in ein kartesisches Koordinatensystem überführt. Weiterhin werden verschiedene Methoden angewendet, um bspw. Dämpfungskorrekturen vorzunehmen. Neben einer Störechobehandlung findet hier auch eine grundsätzliche und Radomdämpfungskorrektur sowie eine dynamische R-Z-Beziehung in Abhängig der Wetterlage Anwendung (vgl. Abbildung 38). Das Konsortium entschied sich für die Nutzung des dämpfungskorrigierten DX-Niederschlagsradarproduktes von EGLV. Hierbei handelte es sich schließlich um ein nicht frei verfügbares Produkt, das auf dem frei verfügbaren DX-Produkt des DWD basiert.

Im EGLV-Verbandsgebiet lagen neun über das Climate Data Center (CDC) bereitgestellte Niederschlagsstationen des DWD, die in .txt-Dateien Niederschlagsinformationen in Intervallen von einer Minute erhoben. Aufgrund der verzögerten Bereitstellung nach 15 Minuten

erfolgte jedoch keine Integration in das Informationsmodell, das eine höhere zeitliche Auflösung vorwies. In der Nachbetrachtung und Validierung kamen die CDC jedoch zum Einsatz. Ebenso wurden neben den über das CDC frei verfügbaren Minutensummen des Niederschlages auch die Aufzeichnungen der eigenen Niederschlagsstationen von EGLV in das Forschungsvorhaben integriert. Somit konnte die Anzahl an im Untersuchungsgebiet verfügbaren Stationen nicht nur deutlich erhöht werden, sondern auch eine Verfügbarkeit von Niederschlagsinformationen von 54 Stationen in Echtzeit garantiert werden. Die Anbindung der acht eigenen Klimastationen von EGLV, an denen weitergehende meteorologische Parameter erhoben werden, konnte aufgrund der technischen Rahmenbedingungen nicht in Echtzeit erfolgen, sondern lediglich im Nachgang, wodurch die Daten jedoch für retrospektive Auswertungen genutzt werden konnten. Weitere frei verfügbare Produkte des DWD boten im Hinblick auf die Fragestellungen des Projektes keinen Mehrwert.

Das räumlich höher aufgelöste (BUFR-)Sweep-Produkt (250 m x 1°) machte zwar einen vielversprechenden Eindruck, befand sich jedoch zum Projektstart nicht auf demselben Qualitätsniveau wie das herangezogene Niederschlagsradarprodukt „DX-Dämpfungskorrigiert“, was die Nachbildung sämtlicher dort bereits implementierten Prozessierungsschritte erfordert hätte. Zudem ist das Produkt noch nicht operationell in die Hochwasservorhersage eingebunden, sodass keine Anwendungserfahrungen vorlagen.

Bei der Integration von Drittdaten wurden gemeinsame Einheiten für die unterschiedlichen Drittdaten definiert. Alle Daten müssen mit der State Definition des Informationsmodells synchronisierbar sein. Tabelle 7 zeigt, dass die ausgewählte Auflösung (1 km x 1 km und 5 min) für die Berechnung von neuen ErgebnisKarten für viele Drittdaten geeignet war. Dennoch hatten alle Drittdatenquellen eine eigene Struktur, die in ein JSON-Format konvertiert werden mussten.

Dabei war die Konvertierung der Binär-Codierung des DX-Produktes besonders aufwendig. Diese Dateien beinhalteten einen ASCII-Teil zur Beschreibung der Inhalte und einen binären Teil, welcher einen 360°-Umlauf mit einem Radius von 128 km in einer Auflösung von 16-Bit nachzeichnete.

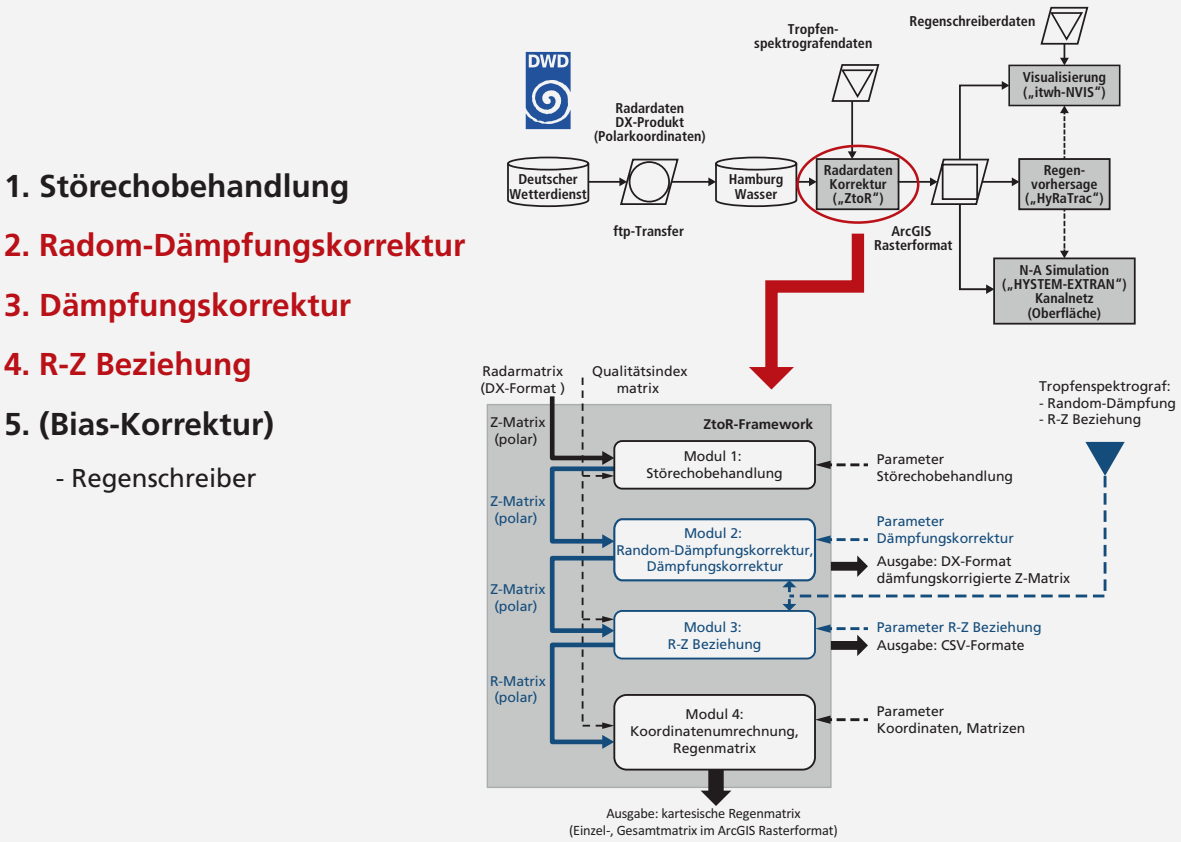


Abbildung 38: Übersicht über durchgeführte physikalische Korrekturen im Rahmen der Prozessierung des Niederschlagsradarproduktes „DX-Dämpfungskorrigiert“. © EGLV

Tabelle 7: Für eine mögliche Integration in das Informationsmodell untersuchte Drittdaten.

		Produkt	Inhalt	Auflösung		Erstellungszeit	Verfügbarkeit	Format	Quelle
				räuml.	zeitl.				
RADARPRODUKTE	lokal	DX	Reflektivitäten (dBZ), hochaufgelöst	1 km x 1°	5 min		sofort	binär	DWD
		(BUFR-)Sweep	Niederschlags-Scan	250 m x 250 m	5 min		nach ca. 30 s	HDF5	DWD
	EG/LV	DX-Dämpfung	Niederschlagsintensitäten (mm), Radarkomposits, hochaufgelöst	1 km x 1 km	5 min		nach ca. 2 min	*.asc	EG/LV
		DX-Online	Kompositdaten aus drei Radaren, hochaufgelöst	1 km x 1 km	5 min		nach ca. 2 h	HDF5	EG/LV
	RADOLAN	RY	qualitätsgeprüfte Radardaten, nach Abschattungskorrektur und nach Anwendung der verfeinerten Z-R-Beziehungen	1 km x 1 km	5 min		nach ca. 2 min	binär	DWD
		RW	RADOLAN-Produkt nach Aneichung an Regenschreiber	1 km x 1 km	60 min		nach ca. 25 min	binär	DWD
NUMERISCHE WETTER-VORHERSAGE	deter-ministisch	ICON-EU	Kurzfrist-Europamodell	6,5 km x 6,5 km	1 h	4 Vorhersagen am Tag; 00/06/12/18 Uhr	nach ca. 3 h	GRIB2	DWD
		COSMO-DE/-D2	Kürzestfrist-Regionalmodell	2,8 km x 2,8 km	1 h	8 Vorhersagen am Tag; 00/03/06/09/ 12/15/18/21 Uhr	nach ca. 1 h	GRIB2	DWD
		RQ	Nowcasting-Vorhersage (0–2h)	1 km x 1 km	15 min		nach ca. 3 min	binär	DWD
	Ensemble	COSMO-DE EPS	Kürzestfrist-Ensemblevorhersage	2,8 km x 2,8 km	1 h	8 Vorhersagen am Tag; 00/03/06/09/ 12/15/18/21 Uhr	04:30/07:35/10: 30/13:30/16:30/ 19:30/22:15/01: 30 Uhr	GRIB2	DWD
		COSMO-LEPS	Kurzfrist-Mittelfrist-Ensemblevorhersage	7 km x 7 km	1 h	2 Vorhersagen am Tag;	01:40/13:40 Uhr	GRIB	DWD

Nach der Konvertierung des binären Formats auf JSON wurde die räumliche Transformation der polaren Koordinaten zu kartesischen Koordinaten durchgeführt. Erst nach dieser Bearbeitung waren die DX-Dateien für die Informationsmodellierung verfügbar und in der Datenbank vorhanden.

Als potenziell verfügbare Geodaten wurden das Amtliche Topographisch-Kartographische Informationssystem (ATKIS), das Amtliche Liegenschaftskatasterinformationssystem (ALKIS) sowie ein Baumkataster identifiziert. Der durch diese Daten generierbare Mehrwert wurde als nicht primär für die in mobile^{VIEW} verfolgten Projektziele bewertet und aufgrund der zeitlich aufwendigen Integration in das Informationsmodell nicht weiterverfolgt.

Als weitere nicht frei verfügbare Datenquelle wurde das terrestrische Messstellennetz von EGLV ausgewählt. Die stationären Pluviometermessungen lieferten Niederschlagsintensitäten in Minutensummen. Zunächst ermöglichten historische Daten nur eine retrospektive Betrachtung von Ereignissen, da sie anfangs in Intervallen von einer Stunde zur Verfügung gestellt wurden. Während der Projektlaufzeit entwickelte EGLV das eigene Messstellennetz allerdings dahingehend weiter, dass für 54 Pluviometer Echtzeit-Minutensummen zur Verfügung gestellt werden konnten. Diese Daten kamen somit auch in der Echtzeit-Kombination mehrerer Datenquellen zum Einsatz.

Aufgrund der hohen Qualität und Verfügbarkeit der beiden gewählten Drittdaten-Produkte, den EGLV-Pluviometern und dem Niederschlagsradarprodukt „DX-Dämpfungskorrigiert", verzichtete das mobile^{VIEW}-Konsortium auf die Einbindung von frei verfügbaren Daten aus privaten Quellen. Die Prüfung dieser weiteren Drittdaten im Hinblick auf zeitliche und räumliche Auflösung und Verfügbarkeit inkl. Ausfallsicherheiten lieferte nicht die gewünschte Qualität. Eine Vermischung der qualitativ hochwertigen Produkte mit Daten, deren Qualität und Verfügbarkeit nicht zu jedem Zeitpunkt gewährleistet werden konnte, wurde dementsprechend abgelehnt.

Im Projekt wurden ausschließlich Drittdaten mit hochwertiger Qualität als Daten- und Informationsquellen verwendet. Diese Daten, u. a. von DWD und EGLV bereitgestellt, wurden im Projekt nicht mehr plausibilisiert. Es wurde davon ausgegangen, dass sie in der vorliegenden Form vertrauenswürdig sind und generell verwendet werden können.

Eine vergleichende Bewertung von Drittdaten ergab jedoch, dass die Messempfindlichkeit bei geringer Niederschlagsintensität unterschiedlich war. In Tabelle 8 und Abbildung 39 wurden hierzu DXD-Radarwerte und Pluviometerwerte verglichen.

Die Radarwerte liegen deutlich häufiger zwischen 0 und 0.1 mm/5 min als die Ergebnisse der Pluviometer-Messungen, was u. a. auch in der Ansprechschwelle der Pluviometer von 0,1 mm/min begründet liegt. Im Ergebnis musste also festgestellt werden, dass sich die Drittdaten bei niedrigen Regenintensitäten als nicht übereinstimmend darstellten und somit als mit Unsicherheiten behaftet einzuordnen waren.

Tabelle 8: Vergleich von Pluviometer- und Radarwerten bei verschiedenen Niederschlagsintensitäten.

RadarTop	0–0.1 mm / 5 min	0.1–0.2 mm / 5 min	0.2–0.3 mm / 5 min	0.3–0.4 mm / 5 min	0.4–0.5 mm / 5 min	>0.5 mm / 5 min
Anzahl der Werte	2.151.490	343.901	239.847	92.900	48.297	111.109
Pluvio >0 mm / 5 min	10 %	34 %	51 %	57 %	61 %	70 %
Pluvio = 0 mm / 5 min	90 %	66 %	49 %	43 %	39 %	30 %

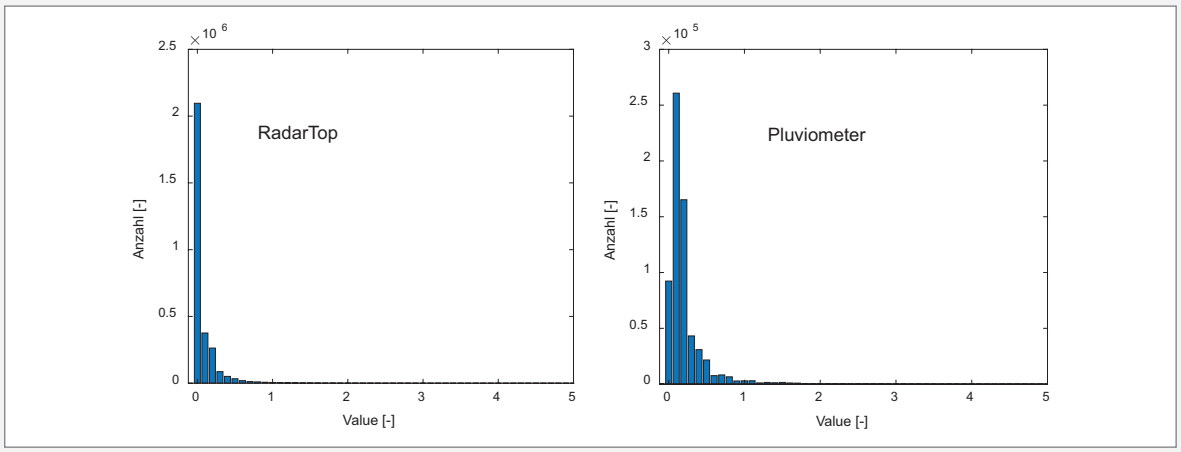


Abbildung 39: Verteilung von Pluviometerwerten und Radarwerten. © IAV

TA 3.3 Informationsverdichtung durch Datenverknüpfung

Das Informationsmodell war die Grundlage für die Auswahl von Drittdaten. Einige der verwendeten Informationen konnten nur von den Fahrzeugen geliefert werden (Regen-Licht-Sensor, GPS, Wischerfrequenz und Geschwindigkeit). Andere Informationen mussten aus Drittdaten für die Modellierung abgeleitet werden. Die gesuchten Informationen sind in Tabelle 9 gelistet.

Das Fahrzeug selbst war die Quelle für die Lichtintensität und die Außentemperatur. Niederschlagsinformationen lieferte EGLV (Radardaten und Pluviometer). Für die anderen Informationen wurde keine Quelle gefunden, die in Echtzeit verfügbar war und für die Erstellung von MSPM geeignet war. Sie wurden daher nicht weiter berücksichtigt.

Tabelle 9: Informationen für die Datenassimilation.

Information	Quelle
Niederschlagintensität	EGLV Pluviometer
Windprofil	Nicht vorhanden
Niederschlagradar	EGLV DXD
Lichtintensität	Fahrzeuge
Wolkenbedeckung	Nicht vorhanden
Tropfengröße	Nicht vorhanden
Tropfenanzahl	Nicht vorhanden
Temperatur	Fahrzeuge
Bodenwind	Nicht vorhanden
Aggregatzustand (Schnee, Hagel, Regen)	Nicht vorhanden

In der Rohdatenanalyse fand bereits eine Prüfung der Relevanz von Mustern der erhobenen Daten im Hinblick auf Niederschlagsereignisse statt. Eine Rekursion der Mustererkennung und eine Evaluierung der Dateneignung im Betrieb wurden kontinuierlich durchgeführt. Darauf aufbauend wurde im Anschluss eine Prüfung der Relevanz der verschiedenen Kfz-Sensordaten für die qualifizierte Erfassung von Niederschlag und für die Warnung vor Starkregenereignissen durchgeführt.

Abgesehen von der fachlichen Relevanz eines Sensors konnte dabei auch die Qualität der übermittelten Daten ausschlaggebend sein. Einige Sensoren gaben überwiegend Fehlerwerte aus oder konnten aufgrund der gewählten Übertragungstechnik über eine kontaktlose CAN-Schnittstelle nicht ausgelesen werden. Der Datenexport wurde in mehreren Iterationen vom Konsortium verbessert, sodass Fehlerwerte reduziert und die Zuverlässigkeit von Signalen erhöht werden konnte. Dennoch wurden für einige Sensoren konstante Werte (z. B. „0“) unabhängig von der jeweiligen Wetterlage ausgegeben. In diesen Fällen erfolgte eine Eliminierung des Parameters aufgrund fehlender fachlicher Relevanz und technischer Verfügbarkeit.

Die folgende Tabelle 10 führt die Begründungen für die im Informationsmodell nicht verwendeten Kfz-Sensordaten auf.

Tabelle 10: Im Informationsmodell nicht verwendete Kfz-Sensordaten.

Sensorbezeichnung	Kommentar	Erläuterung der Nichtverwendung
WischTipp	Wischerhebel: Tippwischen	Tippwischen nicht relevant für die Anzeige von Niederschlagsereignissen. Da es ein manuelles Signal ist, würde kein Fahrer bei einem Niederschlagsereignis permanent manuell die Scheibe wischen. Es wird eher für Wasser auf der Scheibe außerhalb von Niederschlagsereignissen eingesetzt.
WischVorn	Waschwasser Frontscheibe aktiviert	Binärer Sensor zur Anzeige, ob Waschwasser auf die Frontscheibe gegeben wird. Rohdatenanalyse zeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt (= Waschwasser). Ein direkter Zusammenhang zu Niederschlagsereignissen ist nicht gegeben.
WischHeck	Waschwasser Heckscheibe aktiviert	Binärer Sensor zur Anzeige, ob Waschwasser auf die Heckscheibe gegeben wird. Rohdatenanalyse zeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt (= Waschwasser). Ein direkter Zusammenhang zu Niederschlagsereignissen ist nicht gegeben.
WischVornEin	Status: Wischer vorne	Parameter kann aufgrund der gewählten Übertragungstechnik nicht übermittelt werden
WischVornGes	aktuelle, durch Sensorik gemessene Wischgeschwindigkeit	Parameter kann aufgrund der gewählten Übertragungstechnik nicht übermittelt werden
WischVornAus	Wischer Frontscheibe inaktiv	Parameter kann aufgrund der gewählten Übertragungstechnik nicht übermittelt werden
WischVornNorm	Wischer Frontscheibe normale Geschwindigkeit	Parameter kann aufgrund der gewählten Übertragungstechnik nicht übermittelt werden
WischVornSchnell	Wischer Frontscheibe hohe Geschwindigkeit	Parameter kann aufgrund der gewählten Übertragungstechnik nicht übermittelt werden
WischVornFeh	Fehlerstatus Wischer Frontscheibe (Anzeige im Kombi)	Parameter kann aufgrund der gewählten Übertragungstechnik nicht übermittelt werden
RegSnrStFeh	Fehlererkennung Regensensor	Binärer Sensor, der lediglich anzeigt, ob der Regensensor einen Wert übermittelt. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt (= kein Fehler vorhanden).
RegSnrSt	Verbauinfo Regensensor	Binärer Sensor, der lediglich anzeigt, ob im Kfz ein Regensensor verbaut ist. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 1 anzeigt (= Regensensor verbaut).
LichtSnrStFeh	Fehlererkennung Lichtsensor	Binärer Sensor, der lediglich anzeigt, ob der Lichtsensor einen Wert übermittelt. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt (= kein Fehler vorhanden).
LichtSnrStFeh	Verbauinfo Lichtsensor	Binärer Sensor, der lediglich anzeigt, ob im Kfz ein Lichtsensor verbaut ist. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 1 anzeigt (= Lichtsensor verbaut).

TempSt	Fehlererkennung Temperatursensor	Binärer Sensor, der lediglich anzeigt, ob der Temperatursensor einen Wert übermittelt. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt (= kein Fehler vorhanden).
SonIntRe	Sonnenintensität rechts (2D Sensor)	Rohdatenanalyse hat gezeigt, dass entweder ein Fehlerwert (NULL-Wert) oder eine 0 übertragen wird.
SonIntLin	Sonnenintensität links (2D Sensor)	Rohdatenanalyse hat gezeigt, dass entweder ein Fehlerwert (NULL-Wert) oder eine 0 übertragen wird.
FeuchFeh	Fehlererkennung Feuchtesensor	Binärer Sensor, der lediglich anzeigt, ob der Feuchtesensor einen Wert übermittelt. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt (= kein Fehler vorhanden).
FeuchTauPkt	Taupunktberechnung in Abhängigkeit der Luftfeuchtigkeit	Berechnete Größe, liefert daher keinen zusätzlichen Mehrwert zur relativen Luftfeuchtigkeit.
AnzStr	erweiterte Helligkeitssensierung für Head-Up Display	Parameter steht nicht in unmittelbarem Zusammenhang mit meteorologischer Situation
AnzFeh	Fehlererkennung RLS	Binärer Sensor, der lediglich anzeigt, ob der RLS einen Wert übermittelt. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt (= kein Fehler vorhanden).
LichtFeh	Fehlererkennung LS	Binärer Sensor, der lediglich anzeigt, ob der Lichtsensor einen Wert übermittelt. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt (= kein Fehler vorhanden).
RegFeh	Fehlererkennung RS	Binärer Sensor, der lediglich anzeigt, ob der Regensensor einen Wert übermittelt. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt (= kein Fehler vorhanden).
RegAdpFeh	Fehlerkennung RS-Adaptierung	Binärer Sensor, der lediglich anzeigt, ob der betreffende Sensor einen Wert übermittelt. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt (= kein Fehler vorhanden).
RegLicht	Regenlicht Anforderung	Binärer Sensor, der lediglich anzeigt, ob beim Fahrzeug Regenlicht aktiviert ist. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt (= kein Regenfahrlicht aktiviert).
LichtAB	Autobahnfahrlicht Anforderung	Binärer Sensor, der lediglich anzeigt, ob beim Fahrzeug Autobahnfahrlicht aktiviert ist. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt (= kein Autobahnfahrlicht aktiviert).
LichtDun	Regen- / Lichtsensor	Binäre Information über Helligkeit/Dunkelheit. Aus der bloßen Information von Tag oder Nacht kann kein direkter Zusammenhang zu einem Niederschlagsereignis hergestellt werden.
LichtInv	Inverses Signal zu eingeschaltetem Licht	Binäre Information, ob Licht eingeschaltet ist. Aus der bloßen Information von Tag oder Nacht kann kein direkter Zusammenhang zu einem Niederschlagsereignis hergestellt werden.

LichtEin	Licht eingeschaltet	Binäre Information, ob Licht eingeschaltet ist. Aus der bloßen Information von Tag oder Nacht kann kein direkter Zusammenhang zu einem Niederschlagsereignis hergestellt werden.
RegPark	Scheibenwischer aktiviert	Binäre Information, ob der Scheibenwischer durch das Signal des Regensensors aktiviert wird. Für eine genauere Quantifizierung ist ein anderer Parameterwert zuständig.
RegVer	–	Binäre Information. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt. Ein Zusammenhang zu meteorologischen Bedingungen konnte nicht hergestellt werden.
Quali	–	Binäre Information. Ein Zusammenhang zu meteorologischen Bedingungen konnte nicht hergestellt werden.
RSSch	Anzeige, ob plötzlich ein Schwall Wasser auf die Frontscheibe trifft (z. B. Spritzwasser LKW)	Binäre Information. Die Rohdatenanalyse hat gezeigt, dass dieser Parameter entweder nicht übertragen wird (NULL-Werte) oder 0 anzeigt. Die Anzeige würde Niederschlag anzeigen, obwohl es sich um Spritzwasser handelt und wir daher nicht betrachtet.
Zaehler	–	Funktion nicht eindeutig. Ein Zusammenhang zu meteorologischen Bedingungen konnte nicht hergestellt werden.
Fehler -997	Label ist vorhanden, Wert nicht	Fehlerwert für Rohdatenanalyse IAV
Fehler -998	Label ist nicht vorhanden	Fehlerwert für Rohdatenanalyse IAV
Fehler -999		Fehlerwert für Rohdatenanalyse IAV
ESPLatBesc	Beschleunigungswert in y-Achse	Kein erkennbarer Zusammenhang für Fragestellung
ESPLonBesch	Beschleunigungswert in x-Achse	Kein erkennbarer Zusammenhang für Fragestellung
ESPWinkelGes	Winkelgeschwindigkeit um die z-Achse, ESP_VZ_Gierrate & ESP_QBit_Gierrate beachten	Kein erkennbarer Zusammenhang für Fragestellung
ESPEin	Statuswert für ABS	Binäre Information, ob ABS eingeschaltet ist. Es kann kein direkter Zusammenhang zu einem Niederschlagsereignis hergestellt werden.
ESPEin	Statuswert für Esp	Binäre Information, ob Esp eingeschaltet ist. Es kann kein direkter Zusammenhang zu einem Niederschlagsereignis hergestellt werden.
RadRPS	Radumdrehungen pro Sekunde sind Eingangsgröße für die Geschwindigkeitsberechnung im Kombi. Aus dieser Raddrehzahl und dem Radumfang wird die Fahrzeuggeschwindigkeit errechnet.	Informationen werden mit „EspVehV" geliefert.
NaviLat	Breitengrad des Ortungsergebnisses nach WGS84 als Kombination von mehreren GNSSen	Zusätzlicher Ortsstempel, kein Mehrwert für Fragestellung gegenüber bestehenden Sensoren.
NaviLatRich	Richtung des Breitengrads des Ortungsergebnisses nach WGS84 als Kombination von mehreren GNSSen	Zusätzlicher Ortsstempel, kein Mehrwert für Fragestellung gegenüber bestehenden Sensoren.
NaviLon	Längengrad des Ortungsergebnisses nach WGS84 als Kombination von mehreren GNSSen	Zusätzlicher Ortsstempel, kein Mehrwert für Fragestellung gegenüber bestehenden Sensoren.

NaviLonRich	Richtung des Längengrads des Ortungsergebnisses nach WGS84 als Kombination von mehreren GNSSen	Zusätzlicher Ortsstempel, kein Mehrwert für Fragestellung gegenüber bestehenden Sensoren.
NaviQuaGesch	Gesamtgenauigkeit bezüglich Geschwindigkeit am Erdboden	Qualitätsstempel für nicht betrachteten Sensor (EspLongitudinalAcc)
NaviQuaRich	Gesamtgenauigkeit bezüglich der Ausrichtung (Kurswinkel)	Qualitätsstempel für nicht betrachteten Sensor (EspYawRate)
NaviQuaWR	Gesamtgenauigkeit bezüglich der 2D-Position (horizontal)	Kein Mehrwert für die Fragestellung
NaviQuaSR	Gesamtgenauigkeit bezüglich der 1D-Position (vertikal)	Kein Mehrwert für die Fragestellung
NaviZeit	UTC-Zeit, Sekunden seit dem 01.01.1970 00:00	Zusätzlicher Zeitstempel, kein Mehrwert für Fragestellung gegenüber bestehenden Sensoren.
NaviHoh	Höhe über dem Meeresspiegel	Physikalische Zusammenhänge möglich, allerdings existieren im Untersuchungsgebiet nur geringfügige Unterschiede. Mehrwert für Niederschlagsfragestellung nicht erkennbar.
AnzGesch	Geschwindigkeit auf Armaturenbrett	Ergebnis von EspVehV zzgl. eines Aufschlags, Verwendung von EspVehV daher zielführender

Somit konnte der Datentransfer um die nicht benötigten Sensordaten reduziert werden. Als Ergebnis wurde eine Datenvorauswahl aus den erhobenen Daten getroffen, die in Tabelle 11 aufgeführt werden. Hierbei ist zu beachten, dass ein Teil der Fahrzeugflotte bereits ab Werk mit Niederschlags- bzw. Helligkeitssensoren ausgestattet war und ein Teil nachträglich mit Sensoren ausgerüstet wurde. Aus diesem Grund sind diese Sensoren in der folgenden Tabelle zweimal genannt und sind jeweils nur in einem Teil der Fahrzeuge verbaut. Die Sensoren funktionierten analog und die ausgegeben Werte somit miteinander vergleichbar.

Tabelle 11: Im Informationsmodell verwendete Sensordaten mit einer Kurzbeschreibung.

Sensorbezeichnung	Kommentar	Wertebereich; Einheit
longitude	GPS-Längengrad im Koordinatenreferenzsystem EPSG:4326	-180.0° – 180.0°
latitude	GPS-Breitengrad im Koordinatenreferenzsystem EPSG:4326	-90.0° – 90.0°
timestamp	Zeitstempel in koordinierter Weltzeit (UTC)	–
RegSnr	Niederschlagssensor der Windschutzscheibe	Reflektivität in 0 bis 100 %
RegMen	Niederschlagssensor der Windschutzscheibe	Reflektivität in 0 bis 100 %
WischSt1	Manueller Scheibenwischstufe	0 oder 1
WischSt2	Manueller Scheibenwischstufe	0 oder 1
WischInt	Automatische Scheibenwischergeschwindigkeit	0 bis 7
LSTop	Lichtsensor über dem Fahrzeug	0 bis 101.200 lux
LichtTop	Lichtsensor über dem Fahrzeug	0 bis 101.200 lux
Temp	Umgebungstemperatur in ° C	-50 bis 75 °C

Bei der Definition von Konfidenzintervallen und Wertebereichen zur Abschätzung der innewohnenden Unsicherheit der jeweiligen Niederschlagsprodukte in Abhängigkeit der Erhebungsmethode wurde auf Werte aus der Literatur zurückgegriffen. Aufgrund der Bauform nach Hellmann ergeben sich bei der stationären Niederschlagsmessung diverse Fehlerquellen, von denen Windeinfluss, Verdunstung sowie Spritz- und Benetzungsverluste die wichtigsten sind. Während der Verdunstungsverlust bei Geräten nach dem Wägeprinzip, wie sie im Projekt zum Einsatz kamen, rechnerisch durch die damit verbundene Gewichtsabnahme eliminiert werden kann, bleiben v.a. Fehler durch Wind bestehen und führen hierbei zu Abweichungen von bis zu 10 % gegenüber der tatsächlichen Niederschlagshöhe bei flüssigem Niederschlag (Schellart et al. 2012). Diese Größe fand auch im Rahmen des Forschungsvorhabens als Konfidenzintervall Berücksichtigung. Bei windexponierten Standorten kann sogar von einem Fehler von bis zu 20 % ausgegangen werden, während der Messfehler bei Schnee auf bis zu 40 – 50 % ansteigen kann. Beim verwendeten Niederschlagsradarprodukt „DX-Dämpfungskorrigiert" wurde die Entfernungsabhängigkeit der Messung vom Radarstandort Essen berücksichtigt. Bis in eine Entfernung von 40 km wurde eine Unsicherheit von 25 % angenommen, während darüber hinaus, aufgrund der zunehmenden Höhe der Messung, von einem 35 %-igen Unsicherheitsbereich ausgegangen wurde (Krämer et al. 2019).

Die Unsicherheit der Niederschlagsintensität, welche mittels KI aus den Fahrzeugsensoren abgeleitet wird, wird durch das KI-Modell, einem Entscheidungsbaum, dynamisch berechnet. Hier gilt also kein vordefinierter Wert, sondern dieser ist jeweils abhängig von der zugrundeliegenden Sensorkombination. Die Werte sind aber in der Regel höher als die des Radars bzw. der Pluviometer. Diese Unsicherheit kann über die mittlere quadratische Abweichung durch einen Entscheidungsbaum für jeden vorhergesagtem Wert quantifiziert werden.

Für den Livebetrieb des Informationsmodells wurde lediglich ein KI-Modell angewendet: ein Regressionsentscheidungsbaum. Um dessen Ergebnisse zu überprüfen und den Entscheidungsbaum mittels eines iterativen Prozesses zu verbessern, wurden die Ergebnisse aus drei weiteren Modellen berücksichtigt: die Kombination aus mehreren Regressionsentscheidungsbäumen (Random Forest) sowie zwei verschiedene Architekturen neuronaler Netzwerke. Die erste Architektur war ein Multilayer Perceptron (MLP) mit zwei versteckten Schichten mit jeweils 64 Knoten und ReLu Aktivierungsfunktion, als Verlustfunktion wurde die mittlere quadratische Abweichung (MSE) und als Optimierungsfunktion die Adam-Methode angewendet. Die zweite Architektur war ein rekurrentes neuronales Netzwerk (RNN) mit long short term memory (LSTM)-Technik. Dieses hat ebenfalls zwei versteckte Schichten mit jeweils 64 Knoten sowie die gleichen Verlust- und Optimierungsfunktionen, aber eine tanh-Aktivierungsfunktion.

Das direkte Verwenden der Sensorwerte in dem nachgeschalteten Informationsmodell war allerdings nicht möglich und umfangreiches Preprocessing war notwendig. So mussten beispielsweise die Daten der beiden verschiedenen Sensorbaureihen zusammengeführt und als eine Quelle betrachtet werden. Zusätzlich wurden im Laufe des Projektes die Bezeichnungen der Sensoren mehrmals angepasst, dies musste über entsprechende Updates ebenso berücksichtigt werden wie das „Flackern" der Sensorik, Ausreißer, Fehler-, Initialisierungs- und Nullwerte.

Nachdem in der Rohdatenanalyse die Funktionalität und Wertebereiche der beiden Sensoren für Niederschlag und Helligkeit sichergestellt wurden, konnten diese Sensoren zusammengeführt werden. Dies bedeutet, dass das KI-Modell diesen Parameter später als gleichwertig und nicht als zwei verschiedene Datenquellen betrachtete. Selbiges wurde mit den Helligkeitssensoren durchgeführt. Die Rohdatenströme aus *RegSnr* und *RegMen* wurden unter dem Parameter *RainSensor* und analog *LSTop* und *LichtTop* zu *LightSensor* zusammengefasst.

Ein wichtiger Schritt des Preprocessing war das Glätten der Kfz-Sensorwerte. Einige der verbauten Sensoren zeigten unstetige Werte. Dies konnte mehrere Gründe haben: messmethodische, ursachenbezogene oder aufgrund der verbauten Auflösung. Ein Beispiel für messmethodische Ursachen war der Niederschlagssensor. Dieser misst die Reflektivität eines kleinen Bereiches (2 cm^2) der Windschutzscheibe oberhalb des Rückspiegels. Je mehr Tropfen sich auf der Windschutzscheibe innerhalb des vom Sensor erfassten Bereiches befinden, desto höher ist die Reflektivität. Gerade bei niedrigen Niederschlagsintensitäten kam es vor, dass die Anzahl der Tropfen stark voneinander abwich und somit schwankende Werte ausgegeben wurden.

Eine ursachenbezogene Quelle für Schwankungen des Niederschlagssensors war auch Gischt. Eine ursachenbezogene Schwankung der Sensorwerte lag weiterhin beim Helligkeitssensor vor. Äußere Einflüsse, wie z. B. Alleendurchfahrten, künstliche Beleuchtung oder ein sich abwechselndes Wolkenbild führten zu unstetigen Werten. Ein Beispiel für unstetige Werte aufgrund der verwendeten Auflösung war der Temperatursensor *Temp*. Die geringste Abstufung betrug 0,5 °C. Lag die aktuelle Temperatur (z. B. 20,2 und 20,3) zwischen dieser maximalen Auflösung des Sensors, schwankten die Werte hier aufgrund der Auflösung zwischen 20,0 °C und 20,5 °C. Eine Mittelung lag hier also näher an der realen Temperatur als die ausgegebenen Werte. Ebenso konnte bei einigen Sensoren ein Abfall der Werte auf null für die Dauer einer Messung beobachtet werden. Diese Messfehler wurden durch die Glättung ausgeglichen. Durch Mittelung bzw. Glättung konnten also einerseits nachteilige Eigenschaften der Sensoren ausgeglichen werden, andererseits diente die Glättung bei Zeitreihen der Reduzierung der Dimensionalität (Fu 2011). Als Glättungsmethode hat das Konsortium einen gleitenden arithmetischen Mittelwert mit Fensterbreiten von 30 s und 300 s gewählt. Die Fensterbreite entsprach der Anzahl der Messungen, da eine Messung pro Sekunde stattfand, sofern keine einzelnen Messpunkte fehlten.

Fehler-, Initialisierungs- und Nullwerte waren der Dokumentation der Sensoren zu entnehmen und mit entsprechenden Routinen aus den Datensätzen zu entfernen. Fehlende Werte wurden nicht durch Interpolation ergänzt, sondern die entsprechenden Messungen gelöscht.

Feature Engineering

Bevor die Rohdaten der Sensorik also in einem KI-Modell verwendet werden konnten, mussten sie zu Features aufbereitet werden. Ein Feature bezeichnet hier die numerische Darstellung eines Teilaspekts der Rohdaten. Für die sogenannte Featureentwicklung ist ein gutes Verständnis der Fragestellung als auch Expertise über die wasserwirtschaftlichen Zusammenhänge erforderlich. Die Leistungsfähigkeit des nachgeschalteten KI-Modells hing wesentlich von der Güte der Features ab. Die Featureentwicklung war somit ein entscheidender Schritt, um ein KI-Modell anwenden zu können und stellte zusammen mit der Datenbereinigung (Überprüfen auf Validität, Vollständigkeit, Konsistenz) einen zentralen Punkt der Datenaufbereitung dar (Zheng et al. 2018).

Im Prozess wurden iterativ einzelne Features entwickelt, in verschiedenen KI-Modellen verwendet und anhand der Güte der Ergebnisse verworfen oder weiter angepasst. Einige der getesteten Features sind in Tabelle 12 aufgeführt sowie im Folgenden beschrieben.

Normierung

Temperatur und Sonneneinstrahlung weisen sowohl einen täglichen als auch saisonalen Verlauf auf. Um Messungen, die zu unterschiedlichen Tages- oder Jahreszeiten aufgezeichnet wurden, miteinander vergleichbar zu machen, implementierte das Konsortium eine Normierung in Abhängigkeit von der Tages- und Jahreszeit. Vereinfacht wurde angenommen, dass bei der Berechnung einer Niederschlagsintensität auch ein bedeckter Himmel vorliegt. Der Himmel gilt als bedeckt, wenn im Tagesmittel mehr als 80 % der sichtbaren Himmelsfläche bewölkt sind (DIN 4710 2003). Die Normierung der Temperatur erfolgte anhand des in *DIN 4710* gegebenen langjährigen Mittels der Lufttemperatur am Standort Essen (DIN 4710 2003). Die Normierung der Sonneneinstrahlung E_a erfolgte anhand der in *DIN 5034-2* gegebenen Formel 2 für die Horizontalbeleuchtungsstärke an bedeckten Tagen (DIN 5034-2 2019). Abbildung 40 zeigt die Verläufe der Horizontalen Beleuchtungsstärke E_a in *klx* und der horizontalen Bestrahlungsstärke E_e in *W/m²*. Die Sonneneinstrahlung ist abhängig von der Sonnenhöhe γ_s, diese ist wiederrum abhängig von Längen- und Breitengrad sowie der Zeit (DIN 17037 2019).

Formel 2: Die Horizontalbeleuchtungsstärke.

$$E_a = (300 + 21000 \sin \gamma_s) * lx$$

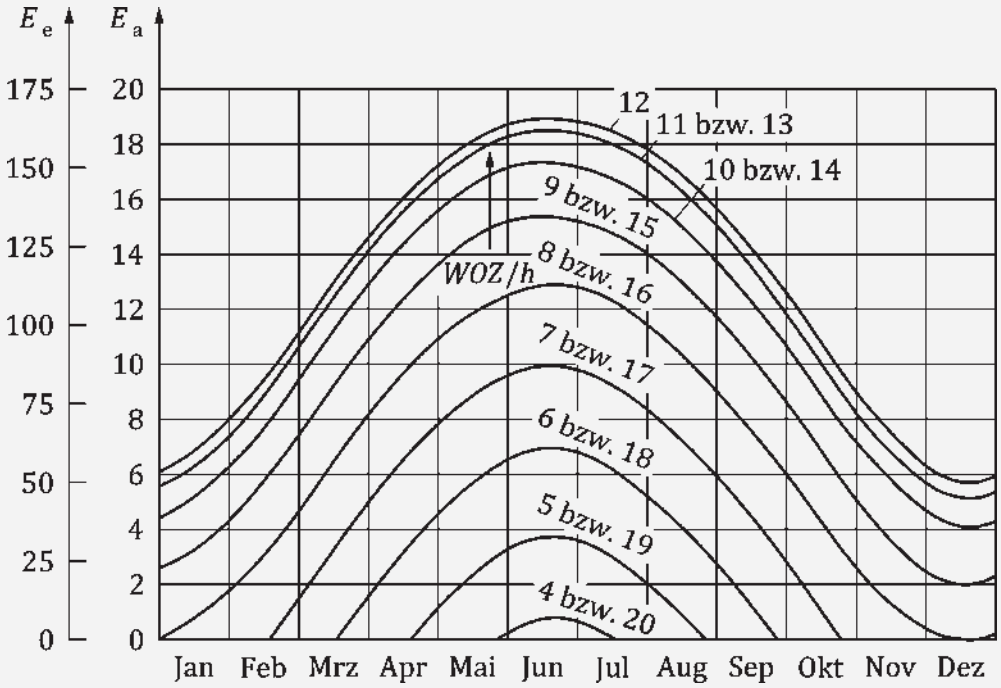


Abbildung 40: Horizontale Beleuchtungsstärke E_a [klx] und horizontale Bestrahlungsstärke E_e [W/m²] bei bedecktem Himmel für 51° nördlicher Breite in Abhängigkeit von Jahres- und Tageszeit. Quelle: Deutsche Norm DIN 5034-2

Gradienten

Da eine Messung lediglich Informationen zum Zeitpunkt der Messung beinhaltet, wurde ein weiteres Feature entwickelt. Gradienten (∇), also die Veränderungen der gemessenen Temperatur bzw. Helligkeit, wurden anhand der gemittelten Werte gebildet. Dabei wurde als Zeitraum stets die höchste verfügbare Zeit zwischen 30 s und 300 s (t_1) gewählt. Dieser dynamische Übergang wurde gewählt, um den Verlust an Messungen möglichst gering zu halten. Die Untergrenze von 30 s ist die Untergrenze, ab der im Projekt relevante Veränderungen der gemittelten Temperatur (T) bzw. Helligkeit (H) frühestens zu erwarten waren und 300 s entsprechen einem Radarumlauf bzw. der maximalen Zeitspanne der Mittelung.

$$\nabla_{Temp} = \frac{T(t_0) - T(t_1)}{t_0 - t_1} = \frac{\partial_T}{\partial_t}$$

$$\nabla_{Hell} = \frac{H(t_0) - H(t_1)}{t_0 - t_1} = \frac{\partial_H}{\partial_t}$$

Tabelle 12: Entwickelte Features für das KI-Modell.

Feature	Beschreibung	Einheit
MeanT30	Gemittelte gemessene Temperatur der letzten 30 s	°C
MeanT300	Gemittelte gemessene Temperatur der letzten 300 s	°C
Tnorm	Gemittelte gemessene Temperatur, normiert mit langjährigem Mittel der stündlichen und täglichen Lufttemperatur bei bedecktem Himmel in Essen	–
dTdt	Gradient der gemittelten Temperatur in der max. verfügbaren Zeitspanne zwischen 30 s und 300 s	K/s
MeanLight30	Gemittelte gemessene Beleuchtungsstärke der letzten 30 s	lux
MeanLight300	Gemittelte gemessene Beleuchtungsstärke der letzten 300 s	lux
Lightnorm	Gemittelte gemessene Beleuchtungsstärke, normiert mit tages- und jahreszeitlicher Horizontalbeleuchtungsstärke bei bedecktem Himmel in Essen	–
dLightdt	Gradient der gemittelten Beleuchtungsstärke in der max. verfügbaren Zeitspanne zwischen 30 s und 300 s	lux/s
RainSensor	Wert des Regensensors (Reflektivität)	%
MeanRainSensor30	Gemittelter Wert des Regensensors der letzten 30 s	%
WipCombi	Tatsächliche Wischergeschwindigkeit (entspricht, falls vorhanden, der manuell eingestellten Geschwindigkeit sonst der durch das Auto vorgeschlagenen Geschwindigkeit	Hübe / min
MeanWipCombi30	Gemittelter Wert von WipCombi der letzten 30 s	Hübe / min
WipCar	Durch das Auto vorgeschlagene Wischergeschwindigkeit	Hübe / min
MeanWipCar30	Gemittelter Wert von WipCar der letzten 30 s	Hübe / min
dxd_rounded	Zentrierter, gleitender Durchschnitt der über dem Auto gemessenen Radarwerte (je die nächsten und letzten 90 s)	mm / 5min

In Abbildung 27 wurde dargestellt, wie die Daten in einem Multi-Source-Information-Cluster (MSIC) integriert wurden. Dabei wurden Fahrzeug- und Pluviometer-Messungen als Punktmessungen betrachtet. Die Punktmessungen wurden mit einer Referenzzeit ergänzt, um die Verbindung mit den Radarmessungen zu bewerkstelligen. Das Radar wurde zur Definition der Referenzzeit verwendet (Tabelle 13).

Tabelle 13: Zeitliche Bedeutung des Zeitstempels bei verschiedenen Datenquellen.

Quelle	Zeitliche Bedeutung des Zeitstempels
Fahrzeug	Verfügbarere Werte auf dem CAN-Bus bei der sekündlichen Abfrage
Pluviometer	Niederschlagmenge kumuliert in der Minute bis zum Zeitstempel
Niederschlagradar	Aufnahme in den letzten fünf Minuten vor dem Zeitstempel

Bis zum Eingang der Radarmessung wurden die Punktmessungen in der Liste je nach Referenzzeit gruppiert. Der Empfang der Radarmessung triggerte den nächsten Clustering-Schritt.

Multi-Source-Precipitation-Map als Eingangsgröße in die Hochwasservorhersage

Der Multi-Source-Information-Cluster und die Multi-Source-Precipitation-Map wurden in drei Schritten erstellt:

1. Raum-zeitliche Aggregierung der Punktmessungen unter Berücksichtigung von Gewichtungsfaktoren (Pluviometer, Kfz-Messung)
2. Verschneidung mit Niederschlagsradarwerten innerhalb einer Zelle, Definition eines Wirkungsradius
3. Verbreitung von den verschnittenen Werten auf Nachbarzellen

Die Punktmessungen wurden zu den entsprechenden Zellen gruppiert und aggregiert. Die Aggregierung von Punktmessungen in einer Zelle folgte Formel 3 und wurde je nach Datenquelle gewichtet. Fahrzeugmessungen hatten ein Gewicht von 1 und die Pluviometer 450. Grund für den Unterschied war einerseits der zeitliche Unterschied (1s vs. 60s) und eine qualitative Abschätzung, dass eine Pluviometermessung deutlich genauer als eine Fahrzeugmessung ist (7,5-mal). 450 ist das Ergebnis der Multiplikation beiden Zahlen. Sie stellt keine valide ermittelte Kennzahl dar, sondern bedient sich der beschriebenen Annahmen.

Formel 3: Aggregierung der Kfz-Werte.

$$X = \frac{1}{\sum_{i=1}^{i=N} g_i} \sum_{i=1}^{i=N} x_i g_i,$$

$$wo\ x = Messung\ und\ g = Gewichtung$$

Nach der Aggregierung wurden die Werte in der Zelle verschnitten. Die Verschneidung wurde nach der Ableitung von Tipaldi (2015) gemacht (Formel 4).

Formel 4: Verschneidung der Datenquellen nach Tipaldi.

$$\mu = \frac{\sigma_2^2 \mu_1 + \sigma_1^2 \mu_2}{\sigma_1^2 + \sigma_2^2},\ \sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2},$$

$$wo\ \mu = Mittelwert\ und\ \sigma = Standard\ Abweichung$$

Niederschlagswerte in Zellen, in denen keine Punktmessung vorhanden war, wurden nach der Verschneidung nicht geändert. Diese Werte konnten aber bei dem Verbreitungsschritt angepasst werden, falls sie im Anpassungsradius lagen. Diese Anpassungen wurden erforderlich, um plötzliche Sprünge in der MSPM zu verhindern. Die Verbreitung wurde mithilfe einer kubischen Formel durchgeführt (Formel 5).

Formel 5: Verbreitung eines veränderten Wertes in der MSPM.

$$\mu_{Zelle-neu} = \mu_{Zelle-alt} + (1-f)\Delta\mu_{Ref-Zelle}$$

$$wo\ f = K_2 r^2 - K_1 r^3\ \ und\ K_1 = \frac{2}{r_{ref}^3} und\ K_2 = 3K_1 \frac{r_{ref}}{2}$$

Abbildung 41 stellt die Verbreitung dar. An der Stelle r=0, also dort, wo die Kfz-Daten erhoben wurden, ist die Gewichtung (1-f) gleich 1. Das bedeutet, dass die Differenz $\Delta\mu_{Ref\text{-}Zelle}$ bei r=0 ohne Abzug auf dem Niederschlagswert berechnet wird. Wenn die Entfernung von der Referenzzelle steigt, wird die Anpassung entsprechend geringer ausfallen. Ab einer Entfernung von r=3 wird keine Anpassung mehr vorgenommen. In Abbildung 41 wird auch die zellenbasierte Anpassung gezeigt. Dies stellt die tatsächliche Anpassung auf der Karte (MSPM) dar. Der Referenzradius (r_{ref}) entsprach 3 km, dementsprechend also drei Zellen (da die Länge der quadratischen Zelle 1 km entspricht).

Nachdem alle Verschneidungs- und Verbreitungsschritte berechnet wurden, wurden alle Daten zusammengefasst, um die MSPM zu erstellen. Abbildung 42 zeigt oben die Radarkarte, in der Mitte die MSPM-Karte und unten die Differenzenkarte (MPSM minus DXD). Die Position der Punktmessungen sind in der MSPM und der Differenzkarte gekennzeichnet.

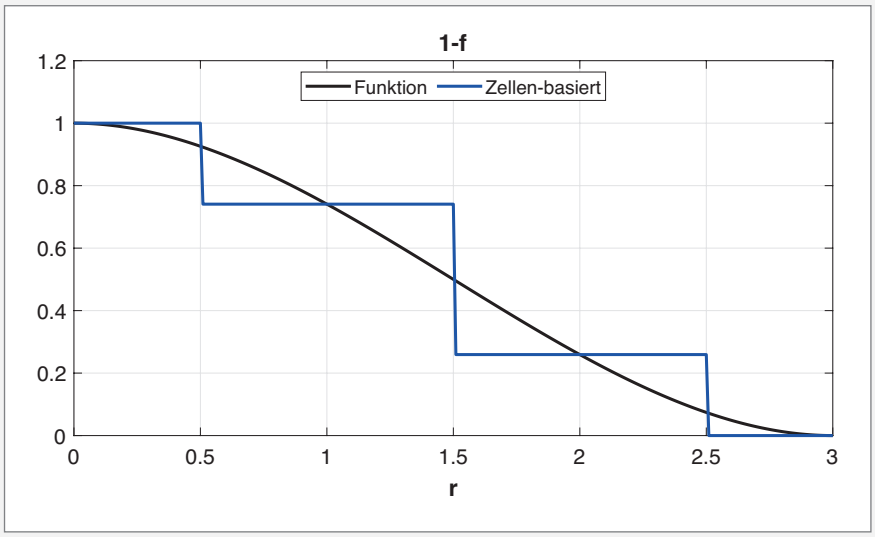


Abbildung 41: Verbreitungsbeispiel eines veränderten Niederschlagswertes auf Nachbarzellen. © IAV

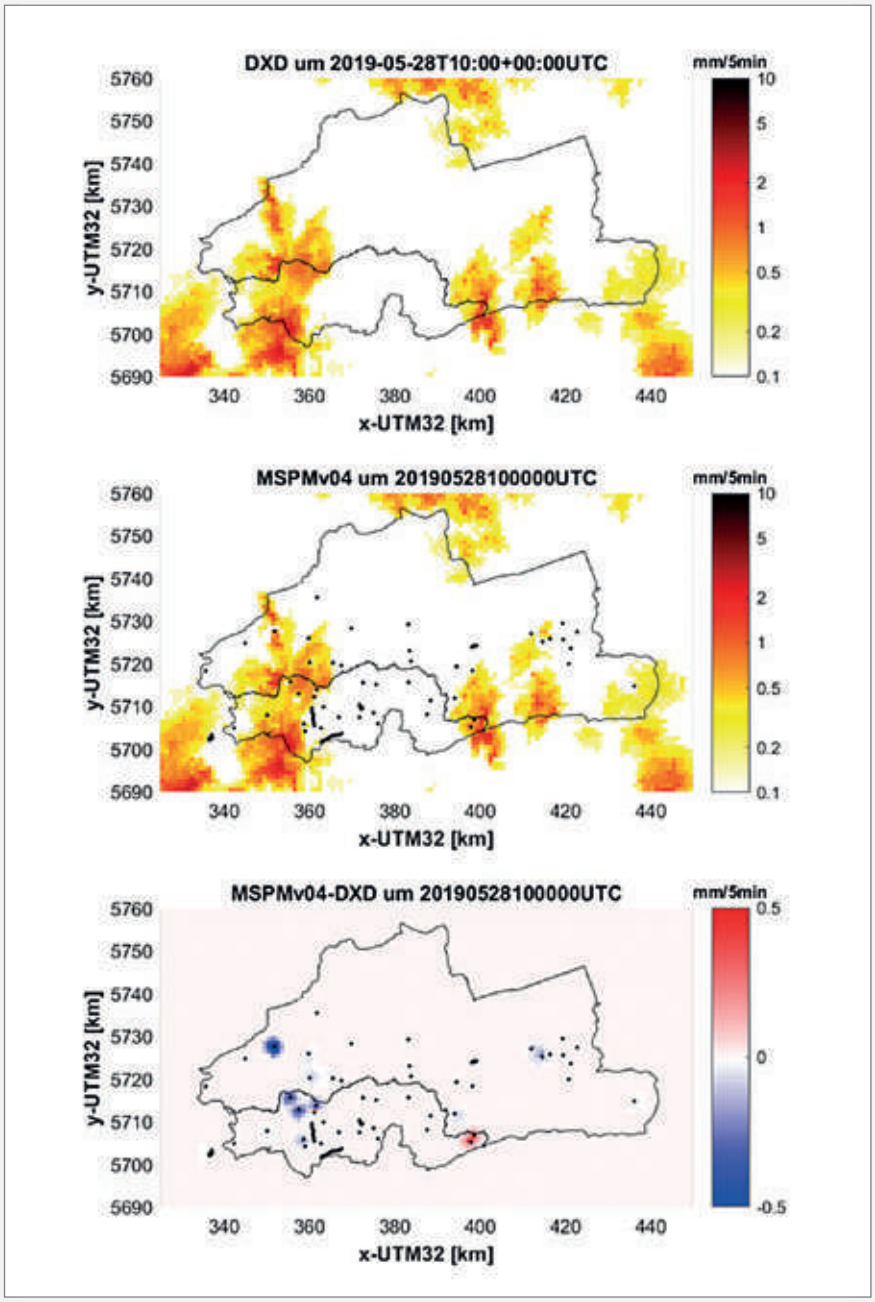


Abbildung 42: Visualisierung des DX-Produktes (oben), der um die veränderten Werte angepassten MSPM (Mitte) sowie der Differenzenkarte der beiden Produkte (unten). © IAV

TA 3.4 Verdichtung bestehender Radarprodukte des DWD und Anwendung auf die Hochwasservorhersage

Gemäß den Ausführungen in TA 2.6 wurde als Zielformat der MSPM zur nahtlosen Integration in das Hochwasservorhersagesystem Delft-FEWS mit dem ESRI-Grid-Format, auch unter dem Namen ArcInfo ASCII verbreitet, dasselbe Format umgesetzt, in dem auch bereits die Niederschlagsradardaten im Forschungsvorhaben bereitgestellt wurden. Das Format an sich ist menschenlesbar und ermöglicht eine barrierefreie Einsichtnahme in die Inhalte, bietet jedoch wenige Optionen für die Definition von Metadaten, welche daher in den Dateinamen codiert werden müssen – z. B. in der nachfolgenden Form: „MSPM_2020-05-18_18-45-00.asc". Der Inhalt besteht aus einem sechszeiligem Header mit Lageinformationen zur korrekten Verortung im Raum in Form der Ursprungskoordinate, der Anzahl an Zeilen und Spalten sowie der Zellgröße im jeweiligen Koordinatenbezugssystem – wobei dieses bekannt sein muss, da es sich in dem Format nicht explizit codieren lässt – und einer Matrix aus n*m inhaltlichen Zellen, in denen sich die Niederschlagsintensität hinterlegt findet:

- ncols 256
- nrows 256
- xllcorner 230000
- yllcorner 5569000
- cellsize 1000
- NODATA_value -9999

Im Vorfeld einer ereignisbezogenen Betrachtung ist festzuhalten, dass sich der Zeitraum der Projektdurchführung und Messdatenerhebung, insbesondere in der starkregenträchtigen Saison April bis September in allen drei Jahren 2018-2020 als überaus trocken darstellte. Bedingt durch weitere Einschränkungen in Form der betrieblichen Dienstzeiten (Montag bis Freitag, 6:00-19:00 Uhr) und dem grundsätzlich kleinräumigen Auftreten von Starkregen, muss an dieser Stelle ebenfalls festgehalten werden, dass trotz eines Kollektivs von 100 Fahrzeugen nicht die erhoffte Zahl an Messungen mit hohen Niederschlagsintensitäten erreicht werden konnte. Zwar ergaben sich in den knapp 18 Monaten an Aufzeichnungsdauer insgesamt 67 Zeitpunkte, in denen sich ein Fahrzeug bis auf 3 km Entfernung an einen Niederschlagsradarpixel mit einer Intensität von 4 mm/5 min näherte, allerdings führt dies im Umkehrschluss auch dazu, dass der überwiegende Teil der verwendeten Datenbasis keine Starkregensituationen abbildet und somit auch nicht für die Erkennung von Starkregen zum Einsatz kommen konnte.

Vielversprechende konvektive Niederschlagssituationen, die zumindest von einigen Fahrzeugen partiell erfasst wurden und auf welche das Konsortium seine Auswertungen im Verlauf des Forschungsvorhabens stütze, trugen sich z. B. an den nachfolgenden Tagen zu:

Tabelle 14: Übersicht über exemplarische Starkregensituationen mit vorhandenen Kfz-Daten.

lfd. Nr.	Datum	Zeitraum [UTC]	hN_{max} in D = 24 h auf Basis von Stationsdaten
1	28.05.2019	09:00 – 19:00	40
2	12.06.2019	09:00 – 14:00	21
3	27.07.2019	13:00 – 16:00	25
4	13.08.2019	13:00 – 18:00	8

Zur Abschätzung der Eignung des finalen Niederschlagsproduktes, der Multi-Source-Precipitation-Map (MSPM), für die hydrologische Modellierung und Hochwasservorhersage vor dem Hintergrund eingebrachter inhaltlicher Unterschiede gegenüber dem verwendeten Niederschlagsradarprodukt „DX-Dämpfungskorrigiert" ist die nachfolgende exemplarische Bewertung zu sehen. Für das erweiterte Untersuchungsgebiet (Verbandsgebiete von Emschergenossenschaft und Lippeverband zuzüglich eines Radius von 30 km), lassen sich die mittels Pluviometer und Kfz-Messungen eingebrachten absoluten Differenzen in die MSPM für den 28.05.2019 auf im Mittel 0,04 mm für alle 288 Zeitschritte bei einer maximalen mittleren Abweichung in Höhe von 0,22 mm quantifizieren. Zur räumlichen Auswirkung der Veränderungen lässt sich festhalten, dass sich die von Differenzen betroffener Pixel größer als 0,1 mm über den Betrachtungszeitraum im Mittel auf eine Fläche von 36 km² belaufen, die jedoch wiederum nicht zusammenhängend ist und sich räumlich über das erweiterte Untersuchungsgebiet, welches sich auf eine Größe von 18.256 km² beziffern lässt, verteilt.

Gemessen an dieser Gesamtfläche ergeben sich im Hinblick auf die über 24 Stunden kumulierten Differenzen zwischen beiden Produkten räumliche Unterschiede auf 43,3 % der betrachteten Fläche über insgesamt 7.896 km². Der Median der täglichen Differenz in den Niederschlagsabbildungen beträgt hierbei 0,01 mm, das 25 %- bzw. 75 %-Quantil belaufen sich auf 0,01 mm bzw. 0,02 mm. In Summe sind bei 49 nicht zusammenhängenden Pixeln Differenzen von über 10 mm über 24 Stunden festzustellen, bei 709 Pixeln beläuft sich die Abweichung auf über 1 mm. Im Umkehrschluss belaufen sich bei einer Fläche von 7.187 km², was einem Anteil von ca. 91 % entspricht, die verursachten Differenzen zwischen Radardaten und MSPM auf unter 1 mm.

Betrachtet man für den gewählten Tag lediglich den hauptsächlichen Zeitraum des Niederschlagsereignisses (ca. 10–13 Uhr UTC) und nicht den gesamten Tag, so ergeben sich mittlere Differenzen in Höhe von 0,07 mm auf durchschnittlich 106 größtenteils dispers verteilten Pixeln. Zwar tritt somit die Größenordnung eingebrachter Veränderungen etwas deutlicher zum Vorschein, bewegt sich hierbei jedoch nach wie vor noch in einem verhältnismäßig geringen Bereich. Diese grundsätzliche Tendenz wird durch analoge Auswertung der übrigen in Tabelle 14 gelisteten Ereignisse bestätigt.

Im Falle des 28.05.2019 konzentrierten sich die kumulierten Differenzen auf den Raum Hamm als auch auf den Raum Gelsenkirchen/Bochum (vgl. Abbildung 43).

Basierend auf den als Trainingsdatensatz verfügbaren Niederschlagsradardaten ist festzustellen, dass eine

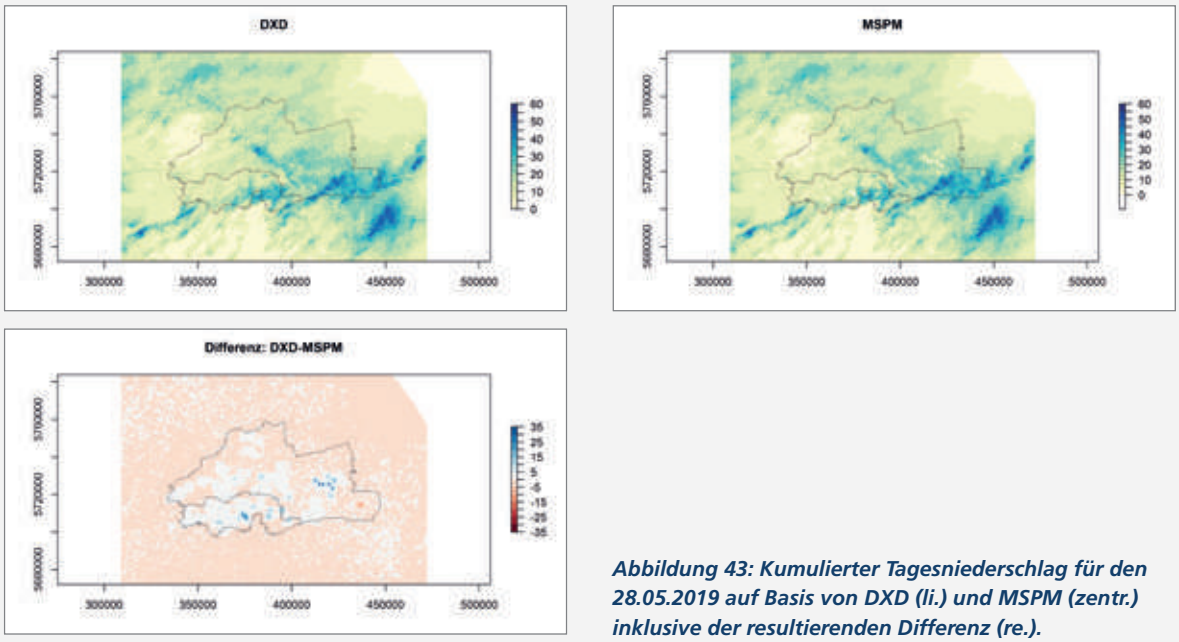


Abbildung 43: Kumulierter Tagesniederschlag für den 28.05.2019 auf Basis von DXD (li.) und MSPM (zentr.) inklusive der resultierenden Differenz (re.).

Anpassung des bodennahen Niederschlages bei tagesscharfer Betrachtung hauptsächlich in Form einer Reduktion stattfindet (vgl. Abbildung 44). Ursächlich sind dafür die bestehenden Rahmenbedingungen in Form von wenigen Messungen von Fahrzeugen in unmittelbarem Umfeld von Starkregenereignissen und ein damit verbundener entsprechend höherer Einfluss von Ereignissen geringer bis mittlerer Intensitäten.

Den Grundbaustein der Hochwasservorhersage bei EGLV bildet Delft-FEWS, ein weltweit eingesetztes und auch bei EGLV operationell betriebenes, modellgestütztes System zur Hochwasservorhersage und zur Prognose von Abflüssen, welches es ermöglicht, Wasserstände für verschiedene Zeiträume zu simulieren und Hochwasserverläufe vorherzusagen (Buchholz et al. 2011). Im Rahmen des Forschungsvorhabens wurde das bestehende System, welches als Drehscheibe für Produkte der numerischen Wettervorhersage des DWD, diverse Niederschlagsradarprodukte und Niederschlagsregistrierungen stationärer Regenschreiber verschiedener Betreiber fungiert, und somit ein breites Portfolio an Niederschlags-Inputdatensätzen für die hydrologische Modellierung bündelt, um die innerhalb von mobile[VIEW] generierten Produkte erweitert (vgl. Abbildung 45). Hierbei wurde nicht nur die Möglichkeit einer Integration von generierten MSPM-Datensätzen geschaffen, sondern auch eine direkte Schnittstelle zur Integration

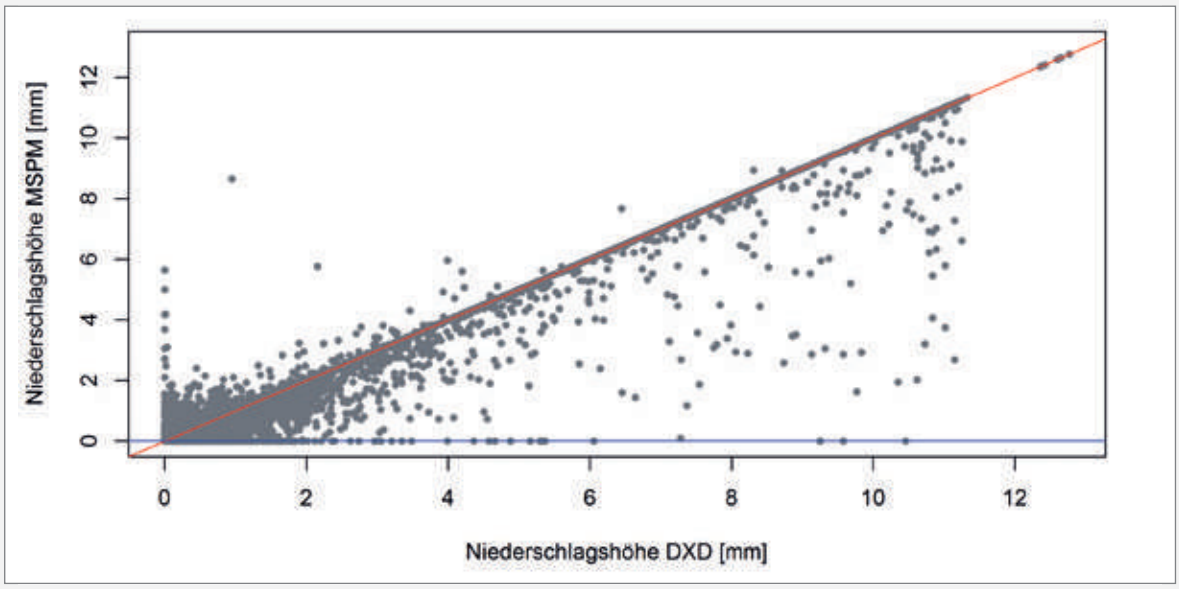


Abbildung 44: Pixelweise Gegenüberstellung von Niederschlagshöhen aus DXD-Niederschlagsradardaten gegenüber MSPM für den 28.05.2019. © EGLV

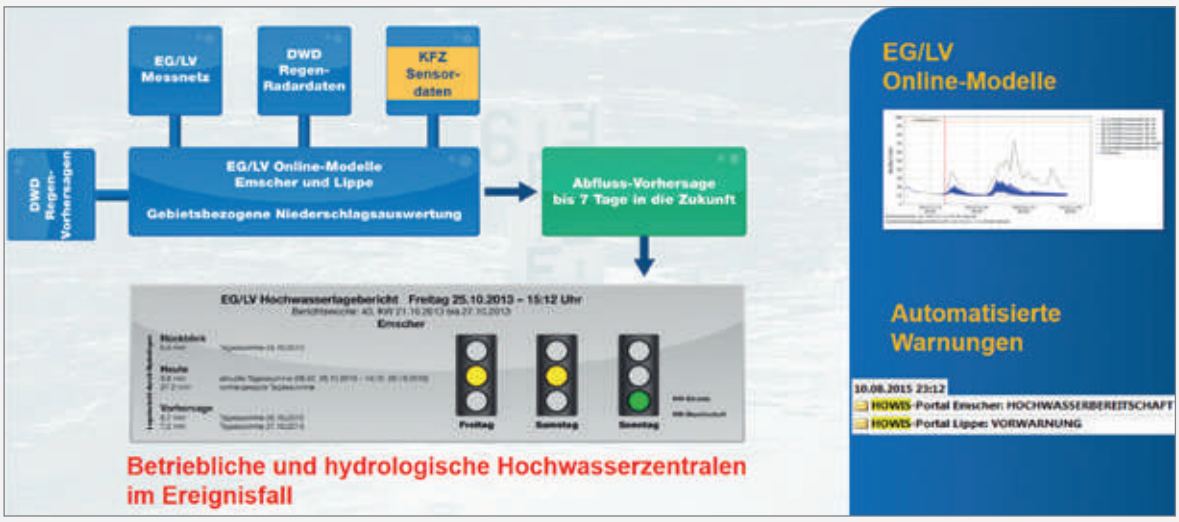


Abbildung 45: Schematische Darstellung des erweiterten Hochwasservorhersagesystem bei EGLV. © EGLV

von Kfz-Rohdaten, auf deren Basis sowohl qualitative Entscheidungen getroffen werden können, als auch eine quantitative Verarbeitung in bereits etablierten Workflows und Routinen möglich ist.

Die Kfz-Rohdaten liegen dabei in strukturierter Form fahrzeugscharf vor und erlauben aufgrund der zeitlichen Entwicklung eines Parameters (vgl. Abbildung 46) automatisierte Prozesse einzuleiten, so z. B. den Versand von Warnungen an einen zuvor definierten Empfängerkreis bei Überschreiten eines bestimmten Schwellenwertes auf Basis von Kfz-Sensorwerten. Das implementierte Warnmanagement sieht vor, auf Basis von fahrzeuggenerierten Intensitäten automatisierte SMS-Warnungen auszusprechen, erlaubt jedoch grundsätzlich auch APIs weiterer Messaging-Dienste anzusprechen.

Diese Warnungen können darüber hinaus – da es sich um Messungen in Form sich bewegender Fahrzeuge handelt – bereits räumlich differenziert eingesetzt werden. Zu jedem Zeitstempel einer Messung ist ein Standort des Sensorträgers zugehörig, woraus sich über die Zeit geläufige GPS-Fahrtstrecken ergeben (vgl. Abbildung 47). Somit kann nach aktuellem Stand bereits

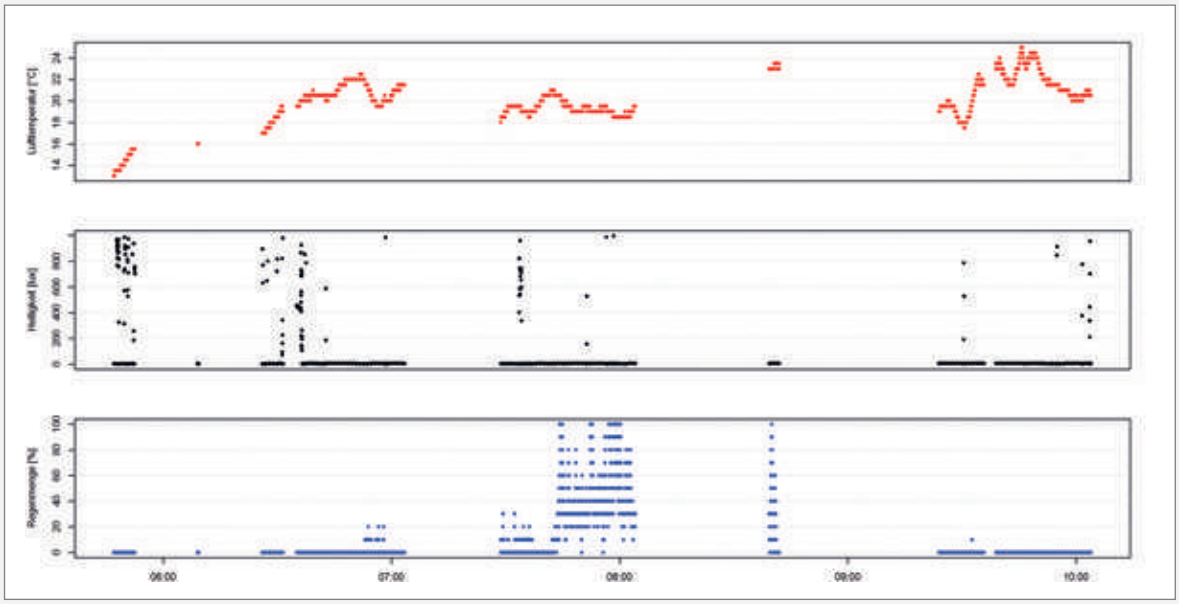


Abbildung 46: Exemplarische Darstellung des Zeitreihen-Displays innerhalb von Delft-FEWS. © EGLV

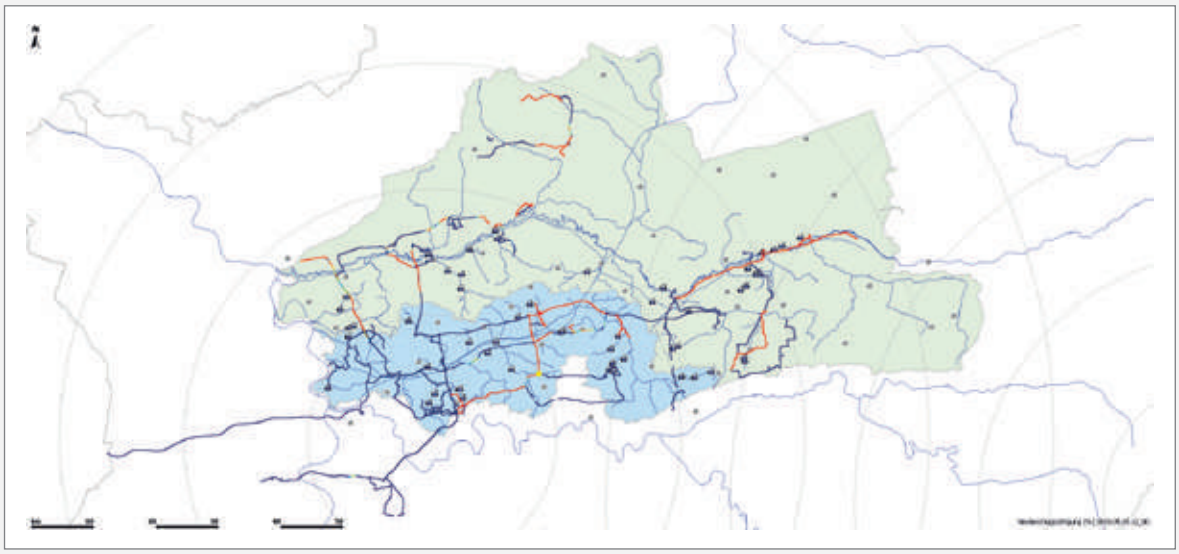


Abbildung 47: Exemplarische Darstellung des räumlichen Displays innerhalb von Delft-FEWS, aktuelle Position des gewählten Fahrzeugs hervorgehoben durch gelben Punkt entlang der Strecke. © EGLV

unter Berücksichtigung der Messposition gewarnt werden, sodass z. B. sich ein in Hamm befindendes Fahrzeug nur eine Warnung für diesen (Groß-)Raum ausspricht, nicht jedoch für weiter entfernte Gemeinden. Bei hinreichend vorhandener Datendichte aufgrund einer signifikant erhöhten Anzahl an Sensorträgern wäre es angesichts der sekündlichen Auflösung von Kfz-Messungen möglich durch Anwendung von geeigneten Interpolationsverfahren ein räumliches Bild des Niederschlagsgeschehens zu erzeugen. Dieses Produkt wäre an die ausgesprochen hohe zeitliche Auflösung der Kfz-Messungen angelehnt und könnte direkt im Rahmen der hydrologischen Modellierung verwendet werden, ohne hierfür etwaige Drittdaten zu benötigen.

Vor dem Hintergrund der bestehenden Messqualität der vorhandenen Regen-Licht-Sensoren ist dies aus inhaltlicher Sicht jedoch nicht sinnvoll umsetzbar. Stattdessen wurde das Warnwesen aktuell auf Basis der MSPM unter Verwendung eines Gebietsniederschlages umgesetzt (vgl. Abbildung 48).

Im Rahmen der Hochwasservorhersage mittels Delft-FEWS setzt EGLV auf das angebundene Niederschlags-Abfluss-Modell NASIM (Hydrotec 2019), welches als konzeptionelles, hydrologisches Grobmodell zur Simulation von Abflüssen an Emscher und Lippe zum Einsatz kommt. Das Modell setzt sich dabei aus räumlich abgegrenzten Einheiten, den sogenannten Teileinzugsgebieten, zusammen, ohne dass es zu einer räumlich kontinuierlichen Modellierung von hydrologischen Prozessen kommt. Vielmehr werden die einzelnen Teilprozesse auf Ebene der Teileinzugsgebiete in Form von Speicherkaskaden abstrahiert und zusammengefasst. Für das Einzugsgebiet der Emscher z. B. beläuft sich die Teileinzugsgebietsgröße auf durchschnittlich 8 km². Eingehender räumlich verteilter Niederschlag über dem Teileinzugsgebiet wird im Modell in einen Gebietsniederschlag umgerechnet.

Am Beispiel der zuvor festgestellten relativ hohen Differenzen zwischen den beiden Niederschlagsprodukten im Raum Gelsenkirchen/Bochum am 28.05.2019 wird

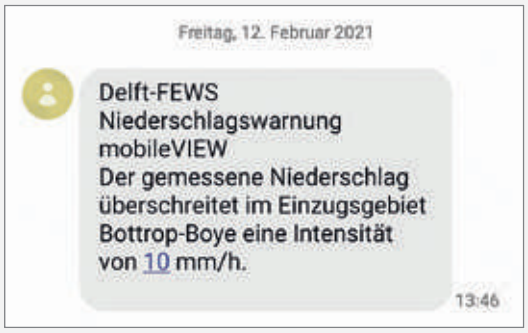


Abbildung 48: Beispiel für umgesetzte Niederschlagswarnung auf Basis von MSPM-Gebietsniederschlägen. © EGLV

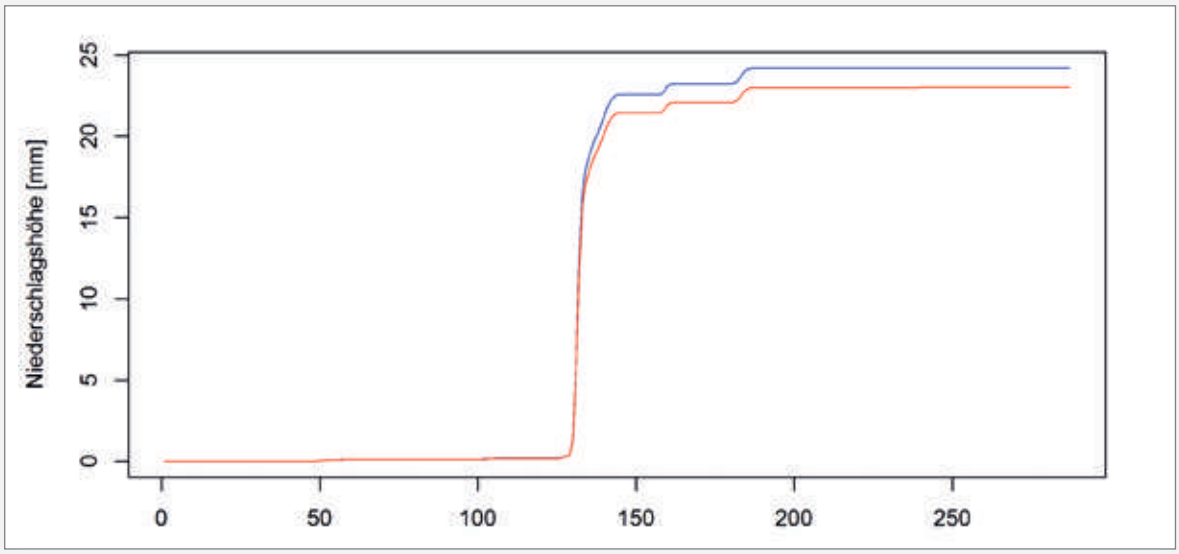


Abbildung 49: Kumulierter Gebietsniederschlag für ein Teileinzugsgebiet am 28.05.2019 (blau: DXD, rot: MSPM). © EGLV

nachfolgend der Gebietsniederschlag des abdeckenden Teileinzugsgebietes, welches eine Fläche von 13,4 km² umfasst und insg. 20 Rasterelemente anteilig schneidet, betrachtet. Auf die Fläche aggregiert ergibt sich für das Niederschlagsradarprodukt „DX-Dämpfungskorrigiert" eine kumulierte Tagesniederschlagshöhe von 24,2 mm (vgl. Abbildung 49), während sich die Summe aus den zugehörigen MSPM des Tages auf 23,0 mm beläuft. Die größte Differenz zwischen beiden Produkten ergibt sich hierbei um 10:55 UTC und beläuft sich auf 0,8 mm.

Unter Berücksichtigung von vereinfachten Stoff- und Energieflüssen wird folglich am Auslass des Teileinzugsgebietes ein simulierter Abfluss erzeugt. An dieser Stelle besteht jedoch keine Möglichkeit, die Güte des Ergebnisses anhand von unabhängigen Messungen abzuschätzen, da dort in aller Regel keine Messungen von Wasserstand bzw. Abfluss durchgeführt werden. Um zu beurteilen, wie sich simulierte zu gemessenen Abflüssen verhalten, bzw. ob eine Veränderung des Inputdatensatzes in Form eines flächendeckenden Niederschlagsproduktes zu einer Verbesserung der (Hochwasser-)Vorhersage führt, ist eine weitere Aggregation von Teileinzugsgebieten erforderlich bis zur Aggregationsstufe der Pegeleinzugsgebiete. Hiervon existieren im Einzugsgebiet der Emscher insgesamt 44, davon 20 entlang des Hauptlaufes. Die restlichen 24 werden an den Nebenläufen der Emscher betrieben. Dies hat zur Folge, dass sich die erforderliche Größe der Teileinzugsgebiete erhöht, um eine valide Aussage treffen zu können – wodurch jedoch auch modellbedingt durch Abstraktionen, Vereinfachungen und Aggregationen durch Erhöhung der räumlichen Betrachtungsebene deutlich höhere Unsicherheiten in das Modellergebnis eingebracht werden. Es gilt folglich die Auswirkungen von sich geändertem Niederschlag im Raum Gelsenkirchen/Bochum zu evaluieren, welcher sich auf eine Differenz von 1,2 mm im Verlauf von 24 Stunden summiert, kommt hierfür als nächstgelegener Pegel „Buer-Sutum" bei Flusskilometer 26,9 in Betracht, der jedoch bereits eine Einzugsgebietsfläche von 492 km² abdeckt – mit all ihren Unsicherheiten. Die notwendige Erweiterung der Einzugsgebietsgröße von 13,4 km² auf 492 km² unter Berücksichtigung der zugehörigen Niederschlagshöhen aus dem restlichen Einzugsgebiet ergibt hier eine 24-Stunden-Summe des Niederschlags in Höhe von 18,4 mm für das Niederschlagsradarprodukt „DX-Dämpfungskorrigiert" und eine Niederschlagshöhe von 17,5 mm für die MSPM, was in einer Differenz in Höhe von 0,9 mm für eines der vielversprechendsten Ereignisse des Jahres 2019 resultiert.

Unter Berücksichtigung der festgestellten Niederschlagsdifferenzen zwischen beiden Produkten, sowohl bezogen auf die beleuchtete Tagessumme größerer Teileinzugsgebiete, als auch bei Betrachtung von Einzelaufnahmen mit oben genannten mittleren Differenzen, sind bei dem eingesetzten hydrologischen Grobmodell daher keine signifikanten Abweichungen in der modellierten Zielgröße, dem Abfluss, zu erwarten. Dies wird im Detail auch an einer Gegenüberstellung zwischen Simulationsergebnissen am Pegel Buer-Sutum mit Fo-

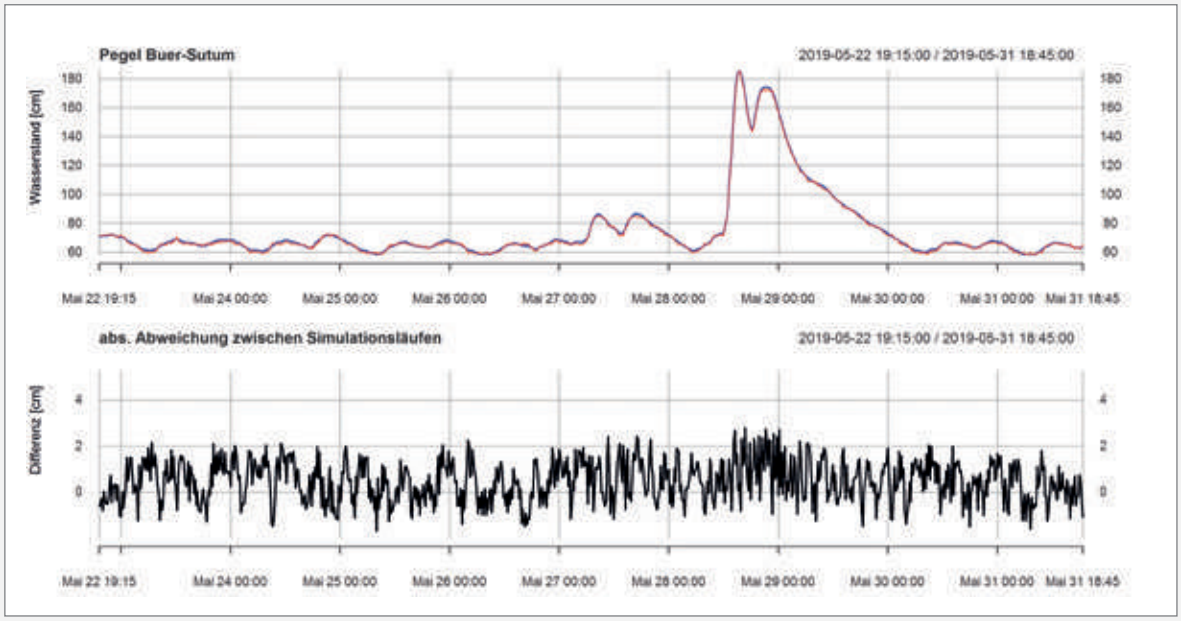


Abbildung 50: Vergleich der Simulationsergebnisse zwischen DXD (blau) und MSPM (rot) am Pegel Buer-Sutum am Beispiel des Niederschlagsereignisses vom 28.05.2019. © EGLV

kus auf das Niederschlagsereignis vom 28.05.2019 deutlich. Wie in Abbildung 50 zu erkennen ist, wird bei Verwendung des Niederschlagsradarproduktes „DX-Dämpfungskorrigiert" (blau) gegen ca. 14:30 Uhr (UTC) mit 185 cm der Scheitelpunkt des Wasserstandes infolge des Starkregenereignisses erreicht. Eine qualitative Änderung des Wasserstandverlaufs ist unter Verwendung der MSPM (rot) angesichts der geringfügigen quantitativen Differenzen zwischen beiden Produkten nicht zu erkennen. Bei einer mittleren Abweichung des Wasserstandes in Höhe von 0,44 cm zwischen beiden Produkten kann rechnerisch auch nur eine theoretische Differenz festgehalten werden.

Eine detaillierte Validierung der Hochwasservorhersagen anhand der charakteristischen Ereignisse, die Prüfung der Prognoseergebnisse im Hinblick auf die genauere Vorhersage des Abflussgeschehens sowie die konkrete Validierung der Ergebnisse im Hinblick auf Gewässer und auf Sturzfluten erscheint an dieser Stelle vor dem Hintergrund des Modellverhaltens sowie der Größenordnung durchgeführter Veränderungen aus hydrologischer Sicht daher als wenig vielversprechend.

Vergleich von Niederschlagsradar und Regen-Licht-Sensor-Werten

In Abbildung 42 zeigen die Punktmessungen in der überwiegenden Anzahl der betrachteten Zellen keinen Unterschied zu den Ergebnissen des Niederschlagsradarprodukts. Dort stimmen sie mit dem Niederschlagsradar überein. In Zellen, in denen die Niederschlagskarte aufgrund der unterschiedlichen Aussagen von Radar und Punktmessungen angepasst wurde, führte die Änderung häufiger zu einem niedrigeren Niederschlag. Tabelle 8 und Tabelle 15 zeigen, dass die Pluviometer und Fahrzeuge oft keinen Regen verzeichneten, wenn der Radarwert über 0 betrug.

Tabelle 15: Niederschlagsradarwerte über den Fahrzeugen.

RadarTop	0–0.1 mm/ 5 min	0.1–0.2 mm/ 5 min	0.2–0.3 mm/ 5 min	0.3–0.4 mm/ 5 min	0.4–0.5 mm/ 5 min	>0.5 mm/ 5 min
Anzahl der Werte	3110135	523318	310094	118871	66907	164156
PercRain30 > 0%	54 %	84 %	89 %	92 %	93 %	92.0 %
PercRain30 = 0 %	46 %	16 %	11 %	8 %	7 %	8.0 %

Die gefundene Abweichung führte bei der Berechnung der kumulierten Niederschlagsvolumina zu Unterschieden zwischen den Ergebnissen der Fahrzeugen und den Radarwerten über dem Fahrzeugen, wie in Abbildung 51.

Die Abweichung zwischen Bodenmessung und Radar ist konzentriert in den Bereich 0–0,1 mm/5min und wurde verursacht durch unterschiedliche Empfindlichkeiten der jeweiligen Informationsquellen.

Aus dieser Version der MSPM konnte somit keine Aussage abgeleitet werden, ob ein Regen geringer Intensität, definiert über eine Radarmessung, tatsächlich die Erdoberfläche erreicht hat. Dieses Defizit könnte in einer späteren Evolution mit einer Kalibrierung gegen Pluviometermessungen mit längeren Zeitabständen eliminiert werden.

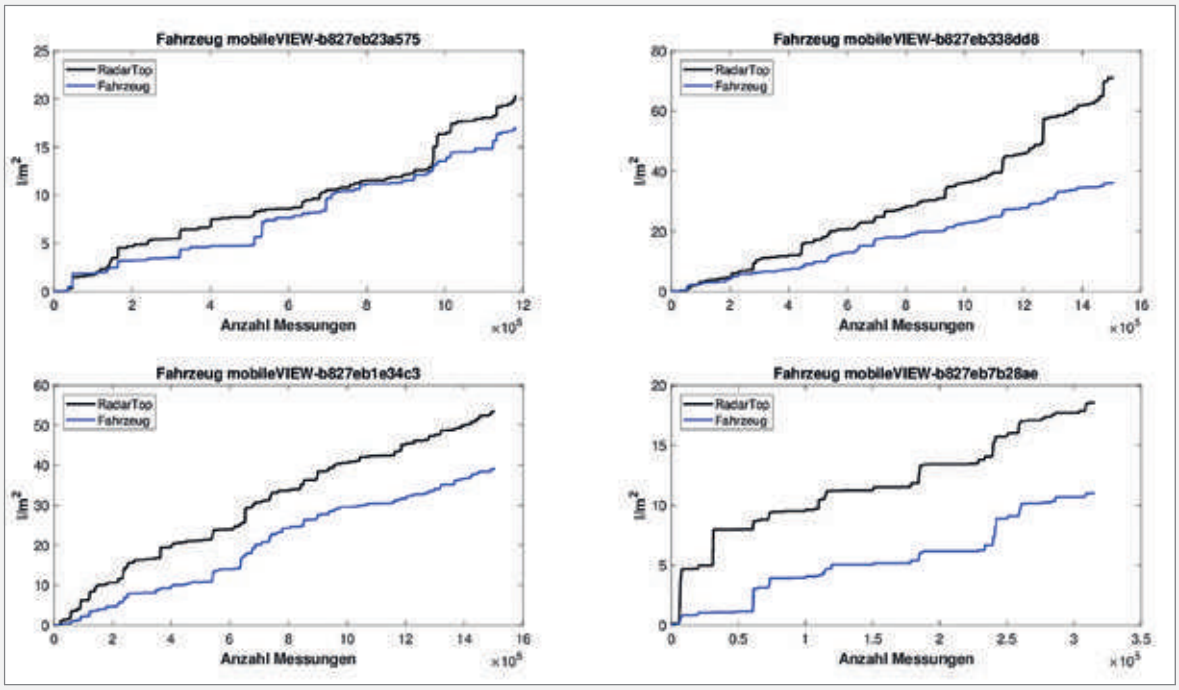


Abbildung 51: Vergleich des kumulierten Niederschlags von Niederschlagsradar- und Kfz-Messungen für vier Fahrzeuge. © IAV

TA 3.5 Bereitstellung der relevanten Ergebnisse und Bewertung der Verbesserung durch die Datenverdichtung und -verknüpfung

Insgesamt wurden vier Versionen der MSPM programmiert. Version 0.1 hatte Probleme mit der Lauf-Stabilität. In den Versionen 0.2 und 0.3 konnten diese Defizite behoben werden, sowie weitergehende Verbesserungen implementiert werden (Abbildung 52). Ab Version 0.3 war eine Evaluation des Modellergebnisses möglich.

Im Verlauf der Datenauswertung wurden fortlaufend Anpassungswünsche formuliert, die in einer Feedbackliste aufgenommen und in der nächsten Version integriert wurden. Für die Version 0.4 wurde so z. B. eine bessere Verwaltung von Ausreißern sowie ein neues Rainmodell impliziert.

Die Reaktion des Informationsmodells bei Starkregenereignissen wurde gesondert betrachtet. In insgesamt 30.592 Fahrzeugmessungen hat die Auswertung der Radarwerte einen Niederschlag größer als 4 mm pro 5 min innerhalb der Umgebung von drei Kilometern des Fahrzeuges ergeben. Dieser Radarwert wurde als *RadarMax*-Wert verzeichnet. Der Radarwert direkt über dem betrachteten Fahrzeug wurde als *RadarTop* bezeichnet. Der Suchradius wurde auf Basis der Verbreitungsradius ausgewählt.

Abbildung 53 zeigt, dass über einem Fahrzeug *(RadarTop)* nur selten ein Radarwert festgestellt wurde, der dem maximalen Radarwert entsprach *(RadarMax)*. Das bedeutet, dass nur in den seltensten Fällen ein Fahrzeug im Kern der maximalen Regenintensität gefahren ist und Daten aufgenommen hat. Für die Berechnung von MSPM bedeutete dies, dass die Anpassungen bei Starkregenereignissen eher während des Verbreitungsstatt des Verschneidungsschrittes stattgefunden haben.

Tabelle 16 zeigt Statistiken über die Übereinstimmung zwischen Fahrzeug- und Radarwerten in der Nähe von einer Starkregenzelle. In der Nähe von einem Starkregenereignis wurde angenommen, dass der Regensensor bei einem Niederschlagsereignis sehr hohe Werte zeigt (durchschnittlicher Wert über 30 Sekunden: *PercRain30* > 99 %). Eine Diskrepanz mit dem Radarwert über dem Fahrzeug *(RadarTop)* bedeutete in dieser Situation, dass die Verschneidung zu einer starken Anpassung geführt hat.

Release Inhalt

Aktualisiert: Abbildung Informationsmodell und ausführliche Release Beschreibung

Neu: Modelle werden durch Interface implementiert, um die Austauschbarkeit von Komponenten zu verbessern

State Definition
- **Neu: Implementierung GeoSplit**

Car Preprocessing:
- **Neu: wipKombi**
- **Neu: Grad and max function**
- **Neu: Gewicht für Car-Entries**

Pluviopreprocess
- **Pluvio als Punktmessung (Gewicht)**

Verschneidung
- **Neu: Aggregierung mit Gewicht**
- **Neu: Ausreißer Funktion**
- Aktualisiert: Impact Berechnung

Abbildung 52: Release Version 0.3 der MSPM. © IAV

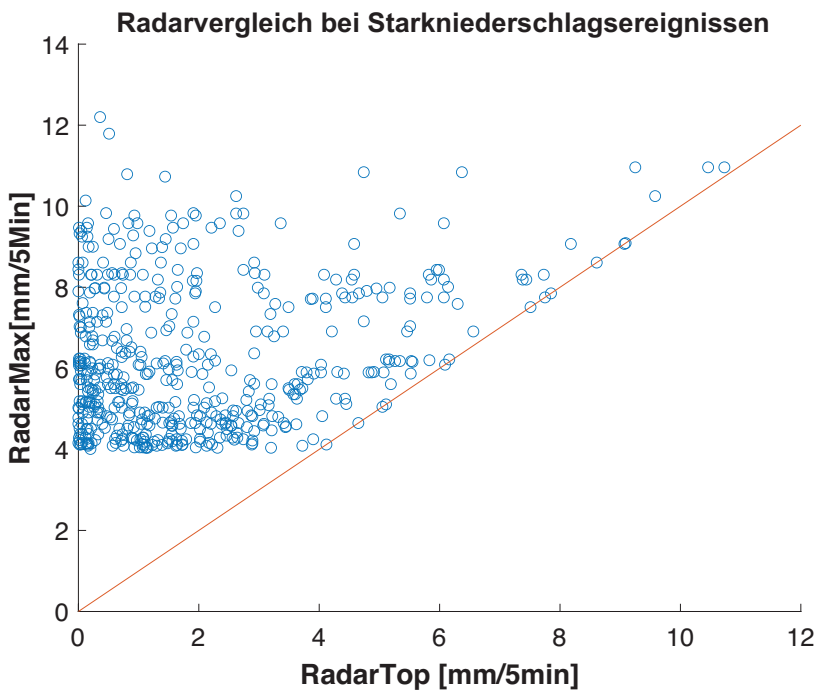


Abbildung 53: Darstellung des Parameters RadarMax im Vergleich zu RadarTop. © IAV

Tabelle 16: Starkregen Statistiken – RadarMax größer als 4 mm / 5 min.

Total: 30592	RadarTop < 0.1 mm / 5 min	RadarTop > 0.1 mm / 5 min
PercRain30 < 99 %	2418	12316
PercRain30 > 99 %	421	15393

Total: 30592	RadarTop < 1 mm / 5 min	RadarTop > 1 mm / 5 min
PercRain30<99 %	8481	6252
PercRain30>99 %	4734	10808

Statistische Auswertungen über diese 30.592 Fälle zeigten, dass in 22 % der Fälle die Fahrzeugmessungen zu einer starken Anpassung führten:

- Situation 1: 421 Mal zeigte das Fahrzeug 100 % beim Regen-Licht-Sensor, während *RadarTop* kein Regen gemessen hatte (bzw. unterhalb der Schwelle 0,1mm / 5 min definiert wurde)
- Situation 2: 6.252 Mal war *RadarTop* noch über 1 mm / 5 min, aber das Fahrzeug verzeichnete keinen Niederschlag.

Die beiden Situationen sind in Abbildung 54 und Abbildung 55 dargestellt. Beide Abbildungen zeigen oben die Radarkarte, in der Mitte die MSPM-Karte und unten die Differenzkarte (MPSM minus DXD). Die Position der Punkmessungen sind in der MSPM und der Differenzkarte gekennzeichnet.

Am 27. Juli 2019 waren einige Starkregenzellen über dem EGLV-Gebiet detektiert worden. Um 14:25 UTC wurde die Radarmessung östlich der Zelle stark modifiziert. Das zu diesem Zeitpunkt dort lokalisierte Fahrzeug und die Pluviometer verzeichneten Niederschlag an Stellen, an denen das Radar hingegen keinen Niederschlag verzeichnet hatte.

Am 13. August 2019 wurde das Gegenbeispiel aufgezeichnet. In diesem Fall bestätigte das Fahrzeug nicht die hohen Niederschlagsradarwerte. Beide Beispiele zeigen, wie die Kombination von Fahrzeugen und Pluviometern die Grenzen einer Regenzelle definieren.

Plausibilisierungsansätze

Das Informationsmodell durchlief verschiedene Iterationsstufen. Dabei wurden sowohl die Modellparameter als auch die Ergebnisse fortlaufend überprüft und entsprechende Anpassungen und Verbesserungen im Modell eingepflegt. Um dies zu ermöglichen, wurden verschiedene quantitative und qualitative Parameter in die MSPM eingepflegt und laufend evaluiert, dies umfasste die folgenden Parameter (Tabelle 17):

Tabelle 17: Qualitätsparameter der MSPM.

Parameter	Zweck
car-entries_used_in_the_model	Anzahl der Einträge durch Kfz in einer MSPM
pluvios-entries_used_in_the_model	Anzahl der Pluviometer in einer MSPM
Aggregation	Zellenwert [mm / 5 min], der allein aufgrund Pluviometer & Kfz berechnet wurde (Zellenwert ohne DX-d)
Average_impact_of_enrichment	Qualitativer Wert, wie weit die MSPM vom DX-d Produkt abweicht

Somit konnte zu jeder MSPM die Anzahl der verwendeten Fahrzeuge und Pluviometer ausgelesen werden, um einschätzen zu können, welche Informationsgrundlagen der jeweiligen MSPM zugrunde lagen. Eine Überprüfung fand ebenfalls durch die qualitative Überprüfung von Differenzenkarten (Abbildung 56) statt, welche die Differenz des DX-Produkts abzüglich der MSPM darstellen. So werden die Veränderungen durch die MSPM visuell verdeutlicht. Dies ermöglichte unplausible Werte schneller zu finden und die dafür verantwortlichen Fehler zu beheben.

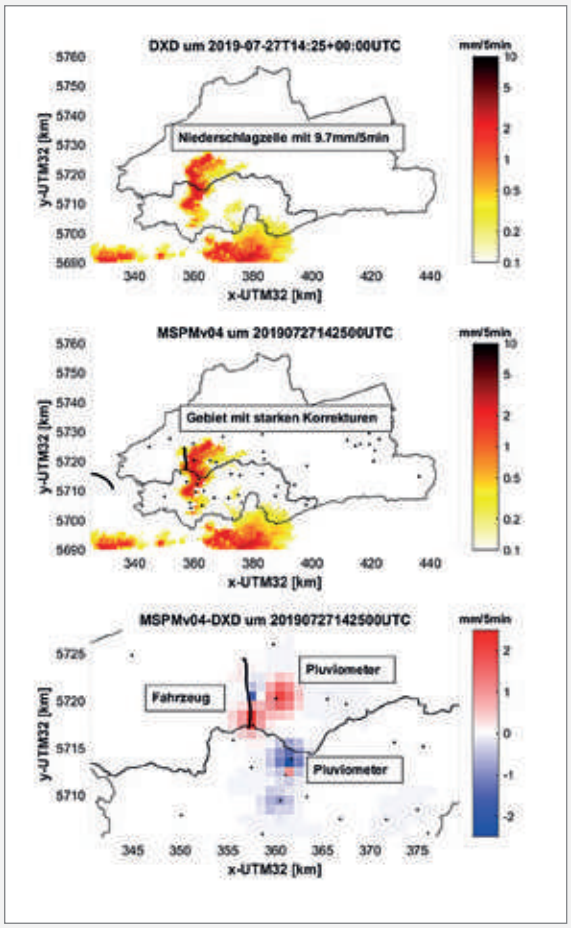


Abbildung 54: Starkregensituation Beispiel 1 am 27.07.2019. © IAV

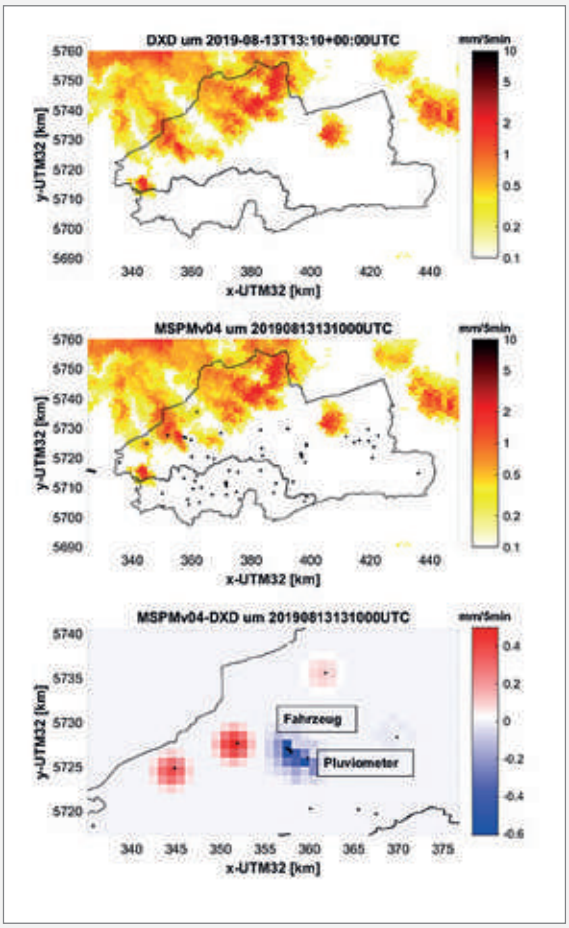


Abbildung 55: Starkregensituation Beispiel 2 am 13.08.2019. © IAV

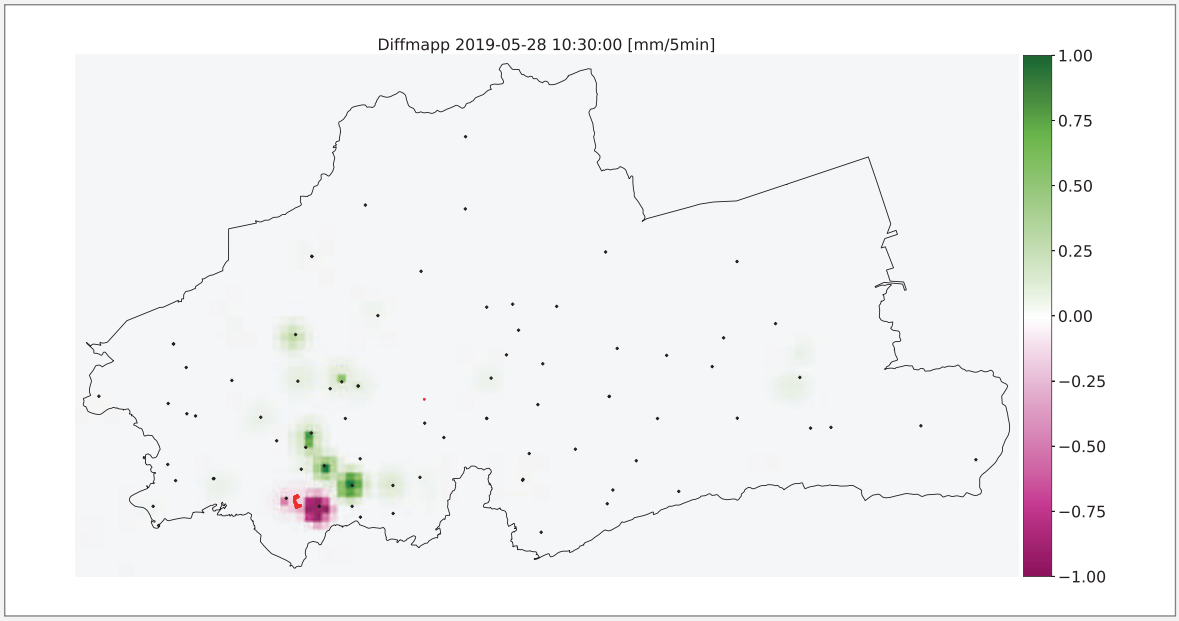


Abbildung 56: Differenzenkarte zur Plausibilisierung. © IAV

Weiterhin wurden die Annahmen, die für die Aggregierung getroffen wurden, statistisch überprüft. Für die Punktmessungen durch Kfz wurde eine Mindestanzahl an Messungen pro Zeitschritt *(minweight)* definiert, welche für Punktmessungen erreicht werden müssen bevor sie mittels einer Formel verrechnet werden. Der Parameter gleicht aus, dass eine einzelne Kfz-Messung eine geringe Qualität aufweist, ausgedrückt durch eine hohe Unsicherheit. In der Summe jedoch können Kfz-Messungen einen Mehrwert generieren. Minweight wurde auf 25 festgelegt, da dies einen guten Kompromiss aus zu erhaltender Datenqualität und zu tolerierendem Datenverlust (ca. 30 %) darstellt. 25 Sekunden entspricht der Zeit, die ein Fahrzeug, welches 70 km/h fährt braucht um eine halbe Radarrasterzelle zu durchfahren. Abbildung 57 zeigt eine Auswertung des Parameters Minweight beispielhaft für den 28.05.2019 als Histogramm links und Boxplot rechts.

Überlegungen zur weitergehenden Plausibilisierung und Validierung der Projektergebnisse

Eine Validierung der Ergebnisse im Hinblick auf die Aufbereitung von MSPM-Daten und Vorhersagen im Web-Interface kann durch eine entsprechende Erweiterung der bestehenden IoT-Plattform realisiert werden. Die exemplarische Auswertung der in Tab. 1 gelisteten Ereignisse, insbesondere des Starkniederschlages vom 28.05.2019, wurde bereits skriptbasiert auf Basis von R, einer freien Programmiersprache für statistische Berechnungen und Grafiken, umgesetzt, ließe sich jedoch auch mit wenig Aufwand in die Programmiersprache Python portieren. Für beide Programmiersprachen existieren mit *Shiny* und *Flask* entsprechende Frameworks für die webbasierte Darstellung von automatisch prozessierten Ergebnissen, wie z. B. eine Validierung nach vordefinierten Kriterien und Routinen. Grundsätzlich, da es sich bei den verwendeten Formaten um geläufige im Kontext von Geodaten handelt, ist es auch möglich, auf einer bestehenden JavaScript- bzw. TypeScript-Architektur aufzubauen und diese modular um notwendige Komponenten zu erweitern.

Weiterführende Plausibilisierungen der Modellergebnisse in Vor-Ort-Untersuchungen sind nach aktuellem Stand unter Berücksichtigung der technisch-konzeptionellen Rahmenbedingungen nicht zu leisten. Um diese zu ermöglichen, braucht es eines deutlich detaillierteren hydrologischen Modells als des bestehenden hydrologischen Grobmodells für die Vorhersage von Wasserständen an den Hauptläufen von Emscher und Lippe. Im Kontext von Starkregenstudien, und auch angesichts der in TA 3.4 festgestellten Differenzen zwischen bestehenden Niederschlagsradardaten und den generierten MSPM, ist es angeraten, ein entsprechend feingranulares hydraulisches Modell aufzubauen, bestenfalls mit gekoppelter Kanalnetzmodellierung, da ein großer Anteil entstehender Schäden durch die Überlastung des städtischen Entwässerungssystems auftritt, nicht zwingend durch abfließendes Oberflächenwasser. Somit wäre eine straßenscharfe Beurteilung möglich, wo und wann auf Grundlage von zuvor erfassten (Stark-)Niederschlägen mit einer Überflutung im urbanen Raum zu rechnen ist. An der Stelle ist zudem ein

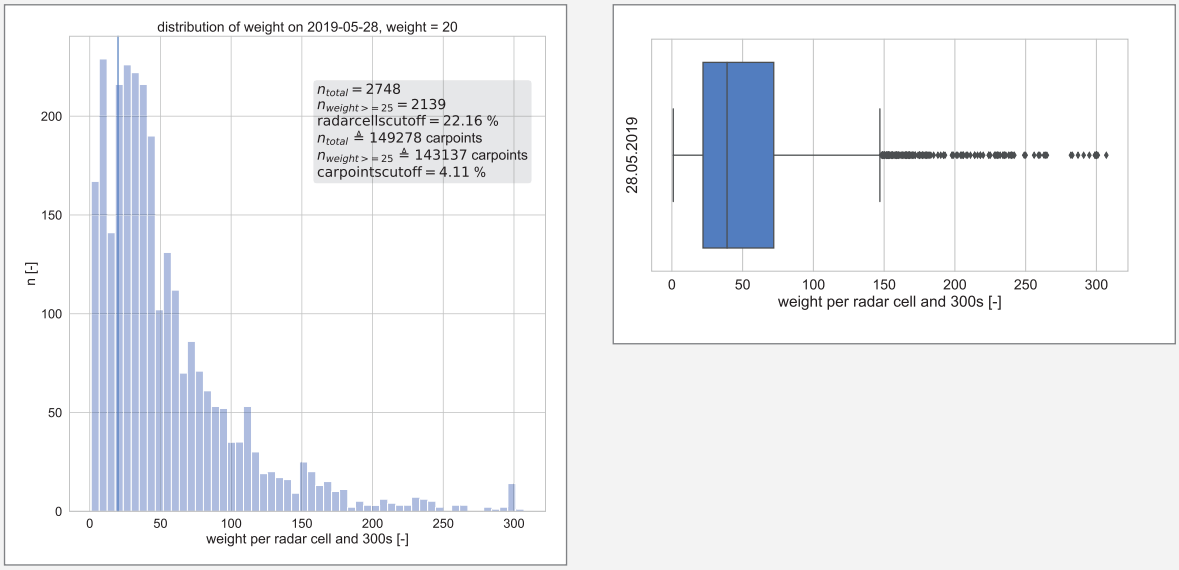


Abbildung 57: Anzahl der Messungen pro Zelle und Zeitschritt im Histogramm (links) und im Boxplot (rechts). © FiW e. V.

Kompromiss zwischen Detaillierungsgrad des Modells, abzubildendem Untersuchungsgebiet und der Anzahl an verfügbaren Fahrzeugen zu treffen. Ein gekoppeltes Feinmodell flächendeckend für beide Verbandsgebiete zu erstellen, erscheint nicht nur aus wirtschaftlicher Perspektive als nicht sinnvoll, sondern auch in Bezug auf die Rechenzeit des Modells, welche im Fall einer Echtzeitverarbeitung ebenfalls eine wesentliche Rolle spielt. Zudem setzt dies das Vorhandensein einer Vielzahl von einheitlich zusammengeführten und aktuellen Grundlagendaten zusammen. Zugleich wirkt sich eine Einschränkung des Untersuchungsgebietes auf einen kleineren und modelltechnisch abbildbaren Raum (z. B. nur Stadtgebiet Essen) nachteilig auf die Anzahl an im Projekt zur Verfügung stehenden Fahrzeugen aus. Existierte ein solch beschriebenes Modell, sollten weiterführende Vor-Ort-Test in der Form ausgeführt werden, als dass eine begleitende Messung von Wasserstand und/oder Abfluss im Entwässerungssystem einer Stadt durchgeführt werden könnte. Eine Nachrüstung entsprechender Messtechnik an geeigneter Stelle wäre im Gegensatz zu händischen Abflussmessungen im Gewässer sowohl realistischer umzusetzen, als auch fortlaufend möglich und, angesichts der beschriebenen Aggregationsstufen, zielführender im Hinblick auf die messtechnische Erfassung von starkregeninduzierten hydrologischen Auswirkungen. Als Vor-Ort-Tests sind auch die in TA 1.5 durchgeführten Messfahrten zu sehen, welche bei Vorhandensein valider Niederschlagsmessungen durch die Kfz-Sensorik entsprechend auszuweiten wären, um mehr Referenzmessungen bei vielversprechenden Wetterlagen auswerten zu können.

Operationell wurde eine SMS-Warnung umgesetzt, über die bei einer Überschreitung einer Intensität von 10 mm/h ein ausgewählter Nutzerkreis informiert wird. Über den gewählten Kommunikationskanal war jedoch kein bidirektionales Feedback möglich, sodass an dieser Stelle auch keine Auswertung von eingegangenem Feedback erfolgen konnte.

Eine Verbesserung der Aussagefähigkeit durch Nutzung von MSPM-Daten zur Verdichtung des Modells im Sinne eines Konzeptes zur künftigen Ausweitung und fortwährenden Nutzung der verdichteten Informationen wäre in einer iterativen Überarbeitung des Informationsmodells zu sehen. Anstelle eines Vorgehens auf Pixelbasis erscheint das Modell in Richtung eines objektorientierten Ansatzes optimierbar. Dies könnte verwendet werden, um als Objekt identifizierte Niederschlagszellen bei nicht übereinstimmenden Messungen aus verschiedenen Quellen entlang eines Vektors zu verschieben, was der modelltechnischen Nachbildung einer Winddrift von fallendem Niederschlag näherkommen würde als die Verringerung eines Niederschlages in einem Radius von z. B. 3 km um eine punktuelle Kfz-Messung. Auf diese Weise ließen sich auch leeseitige Auswirkungen realisieren anstelle von lediglich lokalen Korrekturen. Ebenso erscheint es vielversprechend, wenn aufgezeichnete Niederschläge stationärer Messtechnik insofern mit einer stärkeren Gewichtung in das bestehende Modell eingehen würden, als dass die Messungen im Sinne eines „ground truth" verwendet werden könnten, um die mittels Niederschlagsradar gemessene 2D-Niederschlagsmatrix mittels Ausgleichsfunktionen global in Richtung der Bodenmessungen zu transformieren – und erst anschließend durch Kfz erfasste lokale Abweichungen in das finale Niederschlagsprodukt zu implementieren. Dies ist solange erforderlich, wie eine ausreichende Abdeckung durch Fahrzeugen noch nicht erreicht ist. Bei einer Vielzahl an Fahrzeugen, die in Echtzeit Messdaten erheben, ergäbe sich dadurch eine eigenständige und unabhängige Datenquelle zur Niederschlagserfassung – entsprechende Sensorqualitäten vorausgesetzt.

6.1.4 Teilprojekt TP 4: Projektmanagement und Qualitätssicherung

Projektmanagement

Das FiW fungierte während der 39-monatigen Bearbeitungszeit als Projektkoordinator. Darunter fielen die Koordination und Organisation von Statusmeetings, Arbeitstreffen, Workshops und Tagungsbesuche sowie die Unterstützung der Projektpartner bei administrativen Fragestellungen. Dazu gehörte u. a. auch die Dokumentation der Projektfortschritte, das Controlling des Berichtswesens im Konsortium, sowie die Überwachung von Zeitplänen und Rahmenbedingungen der Förderrichtlinie. Weiterhin war das FiW für die Einhaltung der förderpolitischen Vorgaben des Vergaberechts bei allen Bestellprozessen verantwortlich.

Zu Beginn des Projektes erarbeitete das FiW eine Kooperationsvereinbarung, die in mehreren Iterationsschritten mit den Projektpartnern aktualisiert und schließlich unterschrieben wurde.

Nach Anpassung des Arbeitsplans aufgrund von neuen Erkenntnissen im Projektverlauf sowie neuen technologischen Möglichkeiten, die sich zwischen der Antragsstellung und der Bewilligung ergeben haben, aktualisierte das FiW den Finanzierungsplan der Kfz-Hardware und legte die begründeten Änderungen dem Projektträger vor. Projektintern erarbeiteten die FiW Mitarbeiter das Dokument „aktualisierte Vorhabensbeschreibung"

sowie eine Begründung für die durchgeführte Mittelumwidmung für das Konsortium als Vorlage für das Berichtswesen.

Öffentlichkeitsarbeit

Ein wichtiger Bestandteil des Forschungsprojektes war die Vernetzung in der Fachwelt. Das FiW koordinierte die regelmäßige Teilnahme und Vorbereitung von Fachtagungen sowie die Erstellung von Fachbeiträgen für diverse Journals. Das Konsortium besuchte während der Projektlaufzeit eine Reihe themenrelevanter Veranstaltungen:

- YoWaReS Aachen 2018: Kurzvortrag und Posterpräsentation durch Herrn Braun
- Urban Rain Conference 2018: Einreichung Fachbeitrag durch Herrn Treis
- Tag der Hydrologie 2019: Einreichung Fachbeitrag und Vortrag durch Herrn Braun
- Messe hanseWasser 2019: Veröffentlichung Projektflyer und Vortrag durch Herrn Falk
- IWASA 2020: Einreichung Fachbeitrag und Vortrag durch Herrn Braun
- Tag der Hydrologie 2020: Einreichung Fachbeitrag, Vortrag von Herrn Braun konnte aufgrund der Absage der Veranstaltung infolge der Corona-Pandemie nicht stattfinden)
- AGIT Symposium 2020: Einreichung Fachbeitrag durch Herrn Falk und Online-Vorträge durch Herrn Braun und Herrn Costa-Patry

Auch an den Veranstaltungen der mFUND-Begleitforschung nahmen Vertreter des mobile^{VIEW}-Konsortiums regelmäßig teil:

- mFUND-Konferenz 2017: Teilnahme durch Herrn Dr. Bolle
- mFUND-Konferenz 2018: Teilnahme Herr Braun und Herr Pätsch, Vortrag durch Herrn Braun
- mFUND-Fachaustausch Wetterdaten 2019: Einreichung Kurzbeitrag und Vortrag durch Herrn Tabatabaei
- mFUND Fachaustausch Wetterdaten II 2020: Teilnahme von Herrn Falk und Herrn Braun

Neben den wissenschaftlichen Beiträgen zu diversen Fachtagungen und Journals brachte das FiW die Projektinhalte auch in die Öffentlichkeitsarbeit der Johannes-Rau-Forschungsgemeinschaft (JRF) sowie der ZUSE-Forschungsgemeinschaft ein.

Vernetzung in der Fachwelt

Das Konsortium führte während der Projektlaufzeit eine rege Kommunikation zu anderen Forschungsprojekten. Neben dem bereits beschriebenen fachlichen Austausch mit dem mFUND-Projekt FloWKar stellte u. a. Frau Dr. Marlene Willkomm (Stadtentwässerungsbetriebe Köln) das bereits abgeschlossene Forschungsprojekt sd-kama dem Konsortium vor. Dieses Projekt behandelte das Katastrophenmanagement im Hochwasserfall in der Stadt Köln durch Nutzung von Crowd-Sourcing-Informationen über eine Smartphone-App.

Im Anschluss an die Messe von hanseWasser in Bremen erfolgte auch ein Austausch von Projekterkenntnissen mit dem mFUND-Projekt TransData. TransData beschäftigte sich bis Juni 2019 mit der Optimierung von Transportplanung/-steuerungen, u. a. für die Leerung von Textilcontainern. Dabei wurden Prozess-/Betriebsdaten mit externen Mobilitäts-, Geo- und Wetterdaten verschnitten. Für TransData war primär die Erkenntnis relevant, ob die in mobile^{VIEW} angestrebte Erhebung von Kfz-Sensordaten für die Generierung von kleinräumigen Niederschlagsdaten umgesetzt werden konnte.

Im Januar 2019 wurde in den Scientific Reports des Nature-Magazins ein interessanter Artikel über ein Forschungsprojekt in den USA veröffentlicht, mit dem Titel: „Windshield wipers on connected vehicles produce high-accuracy rainfall maps“ (Bartos et al. 2019). Die Autoren beschreiben darin ein Verfahren, in dem Scheibenwischer als binäre Indikatoren für Bodenniederschlag verwendet werden. Die Autoren setzten dabei auf einen anderen Ansatz zur Verdichtung von Niederschlagsinformationen, dennoch ergaben sich verschiedenen Parallelen zu mobile^{VIEW}. Ein wissenschaftlicher Austausch kam trotz der mehrmaligen Kontaktversuche leider nicht zustande.

7 FAZIT UND AUSBLICK

Hintergrund und Zielsetzung

Das Forschungsprojekt mobile^{VIEW} hat wichtige Erkenntnisse im Hinblick auf die Nutzung von Synergieeffekten von Daten der Wasser- und Verkehrswirtschaft geliefert. Erstmals wurden Kfz-Sensoren zur Generierung von wasserwirtschaftlichen Mehrwerten im Praxiseinsatz verwendet. Für eine massentaugliche Einsatzmöglichkeit existieren aber noch verschiedene Herausforderungen.

Zentrales Ziel des Verbundvorhabens war die Generierung einer weiteren Datenquelle über das Niederschlagsgeschehen im Einzugsgebiet von Emschergenossenschaft und Lippeverband in NRW. Hierfür sollte ein Machbarkeitsnachweis erbracht werden, dass heutzutage bereits in Kfz erhobene Sensordaten zu Niederschlagsinformationen verarbeitet werden können. Hierzu sollten im täglichen praktischen Einsatz Aufzeichnungen von Kfz-Sensoren dahingehend untersucht werden, inwieweit sich Niederschlagsinformationen aus den gewonnenen Daten ableiten lassen. Nach Möglichkeit sollte die Erhebung, Übermittlung und Verarbeitung der erhobenen Kfz-Daten in Echtzeit erfolgen und die neue Datenquelle in die Hochwasservorhersagesysteme von EGLV integriert werden.

Das zugrundeliegende Konzept wurde bereits in einem Projekt an der Leibniz Universität Hannover erprobt, allerdings konnten keine weiterführenden Erkenntnisse zur Ableitung von Niederschlagsinformationen aus Kfz-Sensordaten gewonnen werden, wobei man sich im Projekt RAINCARS hauptsächlich mit Scheibenwischerfrequenzen beschäftigte. Parallel zur Laufzeit von mobile^{VIEW} veröffentlichten Bartos et al. 2019 Ergebnisse einer Studie aus den USA, bei der es gelang, Niederschlagsintensitäten aus Scheibenwischerinformationen anhand eines Bayesschen Algorithmus zu generieren.

Erkenntnisse / Errungenschaften

Der Machbarkeitsnachweis zur Ableitung von Niederschlagsinformationen aus Kfz-Sensordaten konnte durch das Projekt mobile^{VIEW} nun auch in Deutschland erbracht werden. Darüber hinaus konnte ein statistisches Modell zur Verarbeitung von Kfz-Sensordaten zu Niederschlagsinformationen in Echtzeit implementiert werden.

Das Konsortium stellte für den Abgriff von Sensordaten aus der Fahrzeugelektronik eine Übertragungstechnik mit einer Datenbox zusammen, die den Versand über Mobilfunk an einen zentralen Server übernahm. Bei der Auswahl einer geeigneten Übertragungstechnik musste darauf geachtet werden, dass der Eingriff nichtinvasiv erfolgte, um ein Erlöschen der Allgemeinen Betriebserlaubnis der Fahrzeuge zu verhindern. Für die Auswahl geeigneter Fahrzeuge wurde ein detaillierter Anforderungskatalog aufgesetzt, der hard- und softwareseitige Anforderungen erfüllen musste, aber auch organisatorische und standortbezogene Eigenschaften beinhaltete.

Da die Erhebung von GPS-Daten der ausgewählten Fahrzeuge eine Verarbeitung von personenbezogenen Daten im Sinne der Datenschutzgrundverordnung der Europäischen Union darstellte, musste das Konsortium darüber hinaus ein umfassendes Datenschutzkonzept entwickeln und vielzählige Einwilligungserklärungen einholen, um die Daten zu erheben, verarbeiten und ggf. (anonymisiert) publizieren zu dürfen. Hierfür wurden verschiedene technisch-organisatorische Maßnahmen (TOM) implementiert. Unter anderem wurde es den Fahrzeugführenden ermöglicht, über das Aufrufen eines QR-Codes (privacy switch) die Datenerhebung für einen zu definierenden Zeitraum zu unterbinden.

Ein Erfolg des Vorhabens ist somit die Verfügbarmachung von massenhaft dezentral vorliegenden Messdaten im Sinne von Floating Car Data (FCD) für spezifische Anwendungsfälle. Hier ist zukünftig damit zu rechnen, dass neben einer möglichen Integration in lokale Warnmanagementsysteme vor allem eine intelligente Verkehrssteuerung und gegenseitige Warnung von (autonom-)vernetzten Fahrzeugen vor lokalen Gefahrensituationen vor dem Hintergrund des Ausbaus von 5G und der Car2x-Kommunikation vorangetrieben wird.

Nach einer eingehenden Rohdatenanalyse ergaben sich für das Konsortium weitere Herausforderungen im Hinblick auf die weitere Datenverarbeitung. Untersuchungen zur Sensorqualität des Regen-Licht-Sensors zeigten, dass die ausschließliche Verwendung dieses Sensors nicht zu einer Ableitung von Niederschlägen höherer Intensität geeignet war. Vor diesem Hintergrund wurde die Implementierung eines statistischen Modells unter Verwendung weiterer Sensordaten umgesetzt. Zunächst mussten über 70 Sensoren darauf untersucht werden, ob sie in einen Zusammenhang mit Niederschlagsinformationen gebracht werden konnten. Dafür waren sowohl Kenntnisse über physikalische Zusammenhänge des Niederschlags als auch Informationen über die Funktionsweise und Aufgabe der unterschiedlichen Kfz-Sensoren erforderlich. Irrelevante Sensoren konnten eliminiert werden, sodass lediglich ein Bruchteil der Gesamtanzahl in das spätere Modell

zur Ableitung von Niederschlagsinformationen einging. Dabei lagen einige der Sensorsignale in einer Form vor, die eine Verarbeitung erschwerte, und mussten somit im Vorfeld des Modells zu sogenannten Features weiterverarbeitet werden (Feature Engineering). Dies umfasste bspw. die Zusammenfassung verschiedener Scheibenwischersignale zu einem Gesamtwert, aber auch die Normierung von Temperatur- und Lichtsensorwerten.

Nach dem erfolgten Pre-Processing wurden die Sensordaten in das sogenannte Car-Modell eingespeist, das aus den Kfz-Sensordaten Niederschlagsinformationen in Form von Intensitätswerten ableitete. Das Car-Modell wurde im Vorfeld mit einem Trainingsdatensatz an Niederschlagsradardaten angelernt. Dies bedeutete, dass über einen Zeitraum von zwei Jahren erhobene Kfz-Sensordaten den jeweils für diesen Zeitpunkt in der Atmosphäre erhobenen radargestützten Niederschlagsintensitätswerten zugeordnet wurden. Die Varianz der abgeleiteten Intensitäten aus verschiedenen Parameterkombinationen ermöglichte außerdem die Definition eines Konfidenzintervalls für jeden ausgegebenen Intensitätswert. Je größer das Konfidenzintervall, mit desto höheren Unsicherheiten war der abgeleitete Intensitätswert behaftet. Dies wurde in der folgenden Verschneidung mit Niederschlagsradar- und Pluviometerdaten entsprechend berücksichtigt. Die beschriebene Methodik ermöglichte dem Konsortium den Machbarkeitsnachweis zur Nutzbarmachung von Kfz-Sensordaten im Hinblick auf Niederschlagsinformationen. Zur Plausibilisierung der Ableitung von Niederschlagsintensitäten aus dem Car-Modell wurden unter anderem weitere Methoden der künstlichen Intelligenz angewandt.

Um die Niederschlagsintensitäten aus den Kfz-Sensoren in Echtzeit mit zum gleichen Zeitpunkt erhobenen Niederschlagsradar- und Pluviometerdaten zu verschneiden, setzte das Konsortium ein Informationsmodell ein, welches in mehreren Versionen iteriert und weiterentwickelt wurde. Bei der Verschneidung wurden die in einem Raster von 1 km x 1 km vorliegenden Radarniederschlagsdaten um die in den jeweiligen Rasterzellen aggregierten Niederschlagsinformationen aus Kfz-Sensordaten ergänzt, für die Pluviometermessungen wurde analog verfahren. Als zeitliche Auflösung wurden fünf Minuten gewählt, da dies zum Zeitpunkt der Implementierung der höchsten verfügbaren zeitlichen Auflösung der Niederschlagsradarprodukte von DWD und EGLV entsprach. Bei der Verschneidung verwendete das Konsortium für jede Erhebungsmethode spezifische Konfidenzintervalle. Die der konventionellen Erhebungsmethoden wurden aus der Literatur entnommen, diejenigen der Kfz-Werte wurden durch das Car-Modell gemeinsam mit den Intensitätswerten generiert. So wurde die im 1 km x 1 km-Raster vorliegende Niederschlagskarte mit den erhobenen Niederschlagsinformationen aus Kfz-Sensoren zu einer Multi-Source-Precipitation-Map (MSPM) angereichert.

Im Rahmen der Projektdurchführung wurde durch das Konsortium auch die Veröffentlichung von Live-Daten umgesetzt. Dazu wurde eine Internet-of-Things-Plattform aufgesetzt, auf der historische Daten eines gewählten Zeitraums oder die Live-Erhebung von Daten eingesehen werden konnten. Aufgrund der datenschutzrechtlichen Hürden bei der Verarbeitung der Rohdaten musste das Konsortium letztendlich aber auf den öffentlichen Zugang zur Plattform verzichten. Stattdessen wurde ein Produkt für die mCLOUD erarbeitet, das die datenschutzrechtlichen Vorgaben erfüllte und dort öffentlich zugänglich ist. Auch auf diversen Veranstaltungen und in einer Reihe von Fachbeiträgen präsentierte das Konsortium seine Zwischenstände und Ergebnisse.

Die Verwertung der Daten im Rahmen der Hochwasservorhersage ist insofern gelungen, als dass die im Projekt generierten Niederschlagsprodukte, sowohl in Form der Kfz-Rohdaten, als auch MSPM, in das bestehende Vorhersagesystem Delft-FEWS zur Visualisierung und automatisierten Verarbeitung integriert werden konnten. Bei entsprechend guter Sensorqualität und einer hohen Anzahl von Messungen können in diesem Rahmen Kfz-Daten skalierbar als unabhängige Datenquelle verarbeitet werden. Auf dieser Basis ist auch damit zu rechnen, dass der Einfluss der Kfz-Messungen auf die MSPM deutlich größer ausfällt. Aufgrund des vorliegenden hydrologischen Grobmodells und der lediglich relativ geringen hydrologischen Auswirkungen auf die Vorhersage an den Pegeln an Emscher und Lippe ist kein signifikanter Einfluss auf die Modellierungsergebnisse festzustellen, sondern hauptsächlich eine punktuelle Differenzierung im Eingangsdatensatz. Hier wären kleinräumige, gekoppelte hydraulische Feinmodelle als künftig geeignetere Modellgrundlage zu sehen, um im Kontext von Starkregen zu belastbareren Ergebnissen zu gelangen. In Punkto Warnwesen konnten jedoch automatisierte SMS-Warnungen auf Teileinzugsgebietsebene realisiert werden, die bei Überschreitung eines definierten Schwellenwertes an einen festgelegten Empfängerkreis versendet werden.

Herausforderungen und offene Punkte

In einem Forschungsprojekt werden stets unterschiedliche Wege zur Zielerreichung geprüft. So mussten auch in mobile^{VIEW} Änderungen am 2017 entwickelten Arbeitsplan in Anbetracht sich ändernder externer Umstände vorgenommen werden.

Die im Rahmen der Vorüberlegungen angedachte Übertragungstechnik wurde nach Projektbeginn verworfen. Durch die zusätzliche Nutzung nachgebauter Regen-Licht-Sensoren wäre eine Erhebung dieser Daten über eine Bluetooth-Schnittstelle nicht möglich gewesen. Als Vorteil erwies sich dabei, dass damit diese Schnittstelle für den Fahrer nicht blockiert wurde. Bei der Suche nach einer geeigneten Übertragungstechnik musste u. a. die allgemeine Betriebserlaubnis der Fahrzeuge für den Straßenverkehr berücksichtigt werden, wodurch viele invasive Methoden bereits als ausgeschlossen galten. Die letztendlich verwendete Übertragungstechnik stellte eine angemessene Lösung für die Anwendung in einem Forschungsprojekt mit einer begrenzten Anzahl an Fahrzeugen dar, für eine Praxisanwendung wäre sie aufgrund des anspruchsvollen Einbaus und des Preises der Datenbox ungeeignet. Aufbauend auf den Erkenntnissen dieses Projektes ist zukünftig eine günstigere und pragmatischere Übertragungstechnik zu recherchieren. Da diese Daten allerdings bereits intern im Fahrzeug erhoben werden und moderne Fahrzeuge vor dem Hintergrund der Einführung der eCall-Pflicht seit 2018 mit entsprechender Übertragungstechnik ausgestattet sind, ist dies weniger eine technische Herausforderung, sondern vielmehr eine rechtliche Frage zwischen den OEMs und den Fahrzeugführenden im Hinblick auf die Hoheit von im Fahrzeug erhobenen Daten.

Zentraler Diskussionspunkt der Ergebnisse bleibt aber die gewählte Methodik zur Ableitung von Niederschlagsinformationen anhand der Anlernung eines statistischen Modells (Entscheidungsbaum) an Niederschlagsradardaten. Wie zuvor ausführlich beschrieben, unterscheiden sich Niederschlagsinformationen des Niederschlagsradars und von Bodenmessungen erheblich in ihrer Charakteristik. Selbst bei den konventionellen Datenerhebungsmethoden, Niederschlagsradar und Pluviometer, unterscheiden sich die im gleichen Zeitraum aufgenommenen Intensitätswerte bei Starkregenereignissen um bis zu 25 %. Dies hat vielfältige Gründe, die in Kapitel 6 näher erläutert wurden.

Die Messungen von Kfz-Sensoren sind in ihren Rahmenbedingungen den Pluviometermessungen ähnlich, da es sich in beiden Fällen um die Erhebung des bodennahen Niederschlages handelt. Den entscheidenden Nachteil der Pluviometer stellt die geringe Repräsentativität in der Fläche dar. Die Rohdatenanalyse des Konsortiums zeigte auf, dass bereits bei einer Entfernung von über einem Kilometer kein eindeutiger Zusammenhang mehr zwischen den Werten der Pluviometer und der Kfz-Sensoren bestand. Somit hätten sich für ein Training des Modells auf Grundlage von Pluviometerwerten lediglich eine sehr geringe Menge der erhobenen Daten geeignet, nämlich nur die, die innerhalb eines Umkreises von einem Kilometer um ein Pluviometer aufgezeichnet wurden.

Somit wählte das Konsortium für das Anlernen des Car-Modells die einzige andere zeitnah verfügbare Datenquelle mit hoher Qualität, die Niederschlagsradardaten. Ein großer Vorteil dieser Daten ist ihre flächendeckende Verfügbarkeit, sodass eine große Datenmenge für das Training des Car-Modells zur Verfügung stand. Nichtsdestotrotz wären auch bei einer idealen Zuordnung von Kfz-Sensoren zur Niederschlagsradarwerten erhebliche Abweichungen zu erwarten, u. a. aufgrund der folgenden Gegebenheiten:

- Beim Training des Car-Modells werden Punktmessungen (GPS-Koordinaten des Kfz) mit einer Flächeninformation (1 km x 1 km Raster des Niederschlagsradarproduktes) verglichen.
- Weiterhin wird eine Bodenmessung mit einer Messung in der Höhe (im Extremfall bis zu 2 km) verglichen. Durch Winddrift kommt es dazu, dass die in der Höhe gemessenen Niederschläge an anderer Stelle auf den Boden auftreffen.
- Niederschlagsradare sind indirekte Messverfahren und bilden lediglich einen diskreten Zeitpunkt des Niederschlagsgeschehens ab.
- Die zeitliche Auflösung des Niederschlagsradars beträgt 5 min. Die Kfz-Werte werden jedoch sekündlich erhoben und müssen entsprechend zusätzlich zur Aggregation auf die Fläche auch zeitlich aggregiert werden.

Für das Konsortium ergab sich somit die Herausforderung, eine geeignete Methodik zur Plausibilisierung der abgeleiteten Niederschlagsinformationen zu finden. Dies gestaltete sich insofern als schwierig, als dass eine weitere, in der Fläche vorliegende Datenquelle in ausreichender Qualität nicht existent ist. Einzige Möglichkeiten zur Plausibilisierung waren somit, das Pre-Processing der Eingangsdaten sowie die statistische Methode zur Ableitung von Niederschlagsintensitäten im Car-Modell möglichst zu optimieren. Für das Pre-Processing inklusive des Feature Engineerings stützte sich das Konsortium dabei auf die wasserwirtschaftliche Expertise über die Zusammenhänge zwischen Kfz-Sensoren und dem Niederschlagsgeschehen. Bezüglich des Car-Modells lieferte die Überprüfung der abgeleiteten Werte durch einen vom Modell unabhängigen Entscheidungswald sowie ein neuronales Netz zufriedenstellende Ergebnisse.

Bezüglich der Erhebung von Kfz-Sensordaten als weitere Datenquelle für Niederschlagsinformationen können folgende Aussagen getroffen werden: Generell waren abweichende Intensitäten der Kfz-Daten zu den bestehenden konventionellen Erhebungsmethoden aufgrund deren spezifischen Fehlerquellen zu erwarten. Dies stellte schließlich eine der ursprünglichen Motivationen für das Forschungsprojekt dar. Jedoch kann zum Zeitpunkt des Projektendes keine eindeutige Quantifizierung vorgenommen werden, inwieweit diese Abweichungen aus den Unterschieden der Messmethoden-Charakteristika oder aus einem in der Realität über- oder unterschätzt vorliegendem Niederschlagsereignis resultieren.

Die fehlende alleinige Aussagekraft der Regen-Licht-Sensoren war ein zuvor definiertes technisches Risiko und konnte vom Konsortium durch die Entwicklung einer aufwendigen künstlichen Intelligenz zur Generierung von Niederschlagsintensitäten aufgewogen werden. Eine Validierung anhand unabhängiger Referenzmessungen an Pluviometern, die nicht in die Modellierung eingespeist wurden, konnte aufgrund der geringen Anzahl zeitgleicher Messungen von Pluviometern und vorbeifahrenden Kfz nicht weiterverfolgt werden. In Anlehnung an die Ergebnisse von Bartos et al. (2019), dass Kfz-Sensoren zur rein binären Niederschlagsdetektion sehr gut geeignet sind, ist davon auszugehen, dass in mobile^{VIEW} aus den Kfz-Sensoren abgeleitete Nullwerte eine höhere Sicherheit beinhalten, vor allem im Vergleich mit dem Niederschlagsradar, eine Quantifizierung von Niederschlagsintensitäten jedoch mit einer hohen Unsicherheit behaftet ist.

Somit ist für die Bearbeitung in zukünftigen Projekten darauf hinzuweisen, dass für die Erhebung von validen Niederschlagsinformationen die entsprechenden Kfz-Sensoren so weiterentwickelt werden müssen, dass möglichst direkt Intensitäten des Niederschlages gemessen und ausgelesen werden können. Darüber hinaus wäre eine umfassende Messkampagne mit Bodenstationen zur Validierung der Kfz-Sensordaten durchzuführen.

Sollte eine Weiterentwicklung der Kfz-Sensoren nicht möglich sein, kann im Rahmen einer Weiterverfolgung des Ansatzes mit einem statistischen Modell eine hohe Anzahl an Bodenmessstationen für das Anlernen des Modells zu besseren Ergebnissen beitragen.

8 LITERATUR

Bartos, M., et al. (2019). „Windshield wipers on connected vehicles produce high-accuracy rainfall maps." Scientific Reports **2019-01**.

Boylu Uz, F., et al. (2018). „GPUs vs CPUs for deployment of deep learning models." https://azure.microsoft.com/de-de/blog/gpus-vs-cpus-for-deployment-of-deep-learning-models/ 2020.

Breiman, L. (2001). „Random Forests." Machine Learning **45**(1): 5–32.

Buchholz, O., et al. (2011). „Anforderungen an Hochwasservorhersagesysteme – Erfahrungen mit der Software Delft-FEWS." Korrespondenz Wasserwirtschaft (KW) **10**: 564-568.

DIN 4710 (2003). DIN 4710 – Statistiken meteorologischer Daten zur Berechnung des Energiebedarfs von heiz- und raumlufttechnischen Anlagen in Deutschland D. I. f. N. e.V.

DIN 5034-2 (2019). DIN 5034-2. Tageslicht in Innenräumen – Teil 2: Grundlagen. D. I. f. N. e.V.

DIN 17037 (2019). DIn 17037 – Tageslicht in Gebäuden. D. I. f. N. e.V.

Fu, T. C. (2011). „A Review on Time Series Data Mining." Engineering Applications of Artificial Intelligence **24**: 164-181.

Hydrotec (2019). Niederschlag-Abfluss-SimulationNA-SIM -Kurzbeschreibung. Aachen, Hydrotec Ingenieurgesellschaft für Wasser und Umwelt mbH.

IPCC (2013). „Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change." 1535 ff.

Krämer, S., et al. (2019). „Radarregendaten für die Stadtentwässerung – Welche Güte ist erreichbar?" Korrespondenz Wasserwirtschaft (KW) **12/2019**: 100–108.

Lanaro, G. (2017). Python high performance. Build robust application by implementing concurrent and distributed pro-cessing techniques. Birmingham, Packt Publishing.

Rabiei, E., et al. (2013). „Rainfall estimation using moving cars as rain gauges – laboratory experiments." Hydrol. Earth Syst. Sci. **17**: 4701-4712.

Schellart, A., et al. (2012). High Resolution Rainfall Measurement and Analysis in a Small Urban Catchment. St. Moritz, University of Bradford.

Schmitt, T., et al. (2018). „Einheitliches Konzept zur Bewertung von Starkregenereignissen mittels Starkregenindex." Korrespondenz Wasserwirtschaft (KW) **2**: 113-120.

Tipaldi, G. D. (2015). Notes on Univariate Gaussian Distributions andOne-Dimensional Kalman Filters. Department of Computer Science. Freiburg, University of Freiburg.

Zheng, A. and A. Casari (2018). Feature engineering for machine learning: Principles and techniques for data scientists. Peking, O'Reilly.

ABBILDUNGSVERZEICHNIS

Abbildung 1: Standorte der im Rahmen des Forschungsvorhabens ausgerüsteten Fahrzeuge im Verbandsgebiet von EGLV. © EGLV 19

Abbildung 2: Messkette der Übertragungstechnik mit einer Datenbox der Firma Axotec. © IAV 21

Abbildung 3: Embedded Software Test Station. © IAV 22

Abbildung 4: Muster eines im Fahrzeug angebrachten „Privacy Switches“ zur temporären Unterbrechung der Datenaufzeichnung. © EGLV 23

Abbildung 5: Web-Interface bei Aktivierung des „Privacy Switch“ über ein Smartphone. © IAV 23

Abbildung 6: Übersicht über die räumliche Verteilung von EGLV-Niederschlagsstationen in den Verbandsgebieten nach Bereitstellungsintervall (rot = minütlich, blau = stündlich). © EGLV 28

Abbildung 7: Verwendeter Datenlogger in Form einer Axotec-Datenbox (links) und verwendete induktive CAN-Bus-Klemme zum nicht-invasiven Mitlesen der Datenströme im Fahrzeug (rechts). © EGLV 29

Abbildung 8: Serienmäßig in den Rückspiegelfuß integrierter Regen-Licht-Sensor (links) und nachträglich montierte Montagefassung für das Forschungsvorhaben (rechts). © EGLV 29

Abbildung 9: Konzept der IoT-Plattform. © IAV 32

Abbildung 10: Schaubild einer Lambda-Architektur. © IAV 32

Abbildung 11: Gemessener Niederschlag an einem Pluviometer und in den umliegenden Radarzellen. © FiW e. V. 36

Abbildung 12: Temperaturaufzeichnung an einer DWD-Klimastation in einer Höhe von 5 cm (rot) und 2 m (blau). © FiW e. V. 36

Abbildung 13: Heatmap von Fahrstrecken im EGLV-Gebiet zwischen 25.09.2018 und 05.10.2018. © IAV 37

Abbildung 14: Fahrspuren verschiedener Autos (farbige Linien), Pluviometer (blaue Punkte) sowie Radarrasterzellen. © FiW e. V. 37

Abbildung 15: Überprüfung und Visualisierung von Ergebnissen in ArcGIS Pro. © FiW e. V. 38

Abbildung 16: Zeitlicher Verlauf der Scheibenwischerstufen gegenüber Regensensorwerten. © FiW e. V. 38

Abbildung 17: Zeitreihe der Fahrzeugtemperaturmesswerte an einem Morgen. © FiW e. V. 39

Abbildung 18: Histogramme der Lichtsensoren, 12.06.2019, x-Achse: Helligkeit in lux, y-Achse absolute Häufigkeit. © FiW e. V. 40

Abbildung 19: Histogramm der Regen-Licht-Sensoren aller Fahrzeuge am 12.06.2019. Auf der x-Achse die Werte des Regen-Licht-Sensors, auf der y-Achse die absolute Häufigkeit der Werte. © FiW e. V. 40

Abbildung 20: Streudiagramme vom 12.06.2019 mit dem Temperatursensor auf der Scheibe und dem Luftfeuchtesensor. © FiW e. V. 41

Abbildung 21: Zeitlicher Verlauf von FSTemp und TempScheibe. © FiW e. V. 42

Abbildung 22: Korrelationsmatrix ausgewählter Kfz-Sensoren für den 28.05.2019. © FiW e. V. 44

Abbildung 23: Flowchart der Signale der Scheibenwischeranlage. © FiW e. V. 44

Abbildung 24: Messfahrt eines Fahrzeugs vom Standort Essen zu einer Klimastation des DWD. © FiW e. V. 45

Abbildung 25: Vergleich der Messwerte Regen-Licht-Sensor (links) und Pluviometer (rechts) für eine Messfahrt von Essen zu einer Klimastation des DWD. © FiW e. V. 45

Abbildung 26: In das Informationsmodell eingehende Einflussgrößen. © IAV 46

Abbildung 27: Gesamtkonzept des Informationsmodells. © IAV 47

Abbildung 28: Abschnitt eines Entscheidungsbaums zur Klassifizierung von Niederschlagsintensitäten. © FiW e. V. 51

Abbildung 29: Teil eines Entscheidungsbaums zur Regression innerhalb des Kfz-Modells. © FiW e. V. 52
Abbildung 30: Beispielhafte Architektur eines mehrlagigen Perzeptrons. © FiW e. V. 53
Abbildung 31: Architektur eines rekurrierenden neuronalen Netzwerks. © FiW e. V. 53
Abbildung 32: Vergleich einer Informationspositionierung von FEWS und dem Informationsmodell. © FiW e. V. 56
Abbildung 33: Gemittelte Rohsignale am 28.05.2019. © FiW e. V. 56
Abbildung 34: QoS-Auswertung für den Zeitraum 12.02.2019 – 19.02.2019. © IAV 57
Abbildung 35: Login-Maske Webserver für interne Projektmitarbeiter. © IAV 58
Abbildung 36: Visualisierungsoptionen für die Kartendarstellung im Webservice. © IAV 58
Abbildung 37: Kartendarstellung von Daten der Kfz, Pluviometer und des Niederschlagsradars im Webservice. © IAV 59
Abbildung 38: Übersicht über durchgeführte physikalische Korrekturen im Rahmen der Prozessierung des Niederschlagsradarproduktes „DX-Dämpfungskorrigiert". © EGLV 61
Abbildung 39: Verteilung von Pluviometerwerten und Radarwerten. © IAV 63
Abbildung 40: Horizontale Beleuchtungsstärke E_a [klx] und horizontale Bestrahlungsstärke E_e [W/m²] bei bedecktem Himmel für 51° nördlicher Breite in Abhängigkeit von Jahres- und Tageszeit. Quelle: Deutsche Norm DIN 5034-2 71
Abbildung 41: Verbreitungsbeispiel eines veränderten Niederschlagswertes auf Nachbarzellen. © IAV 74
Abbildung 42: Visualisierung des DX-Produktes (oben), der um die veränderten Werte angepassten MSPM (Mitte) sowie der Differenzenkarte der beiden Produkte (unten). © IAV 74
Abbildung 43: Kumulierter Tagesniederschlag für den 28.05.2019 auf Basis von DXD (li.) und MSPM (zentr.) inklusive der resultierenden Differenz (re.). 76
Abbildung 44: Pixelweise Gegenüberstellung von Niederschlagshöhen aus DXD-Niederschlagsradardaten gegenüber MSPM für den 28.05.2019. © EGLV 77
Abbildung 45: Schematische Darstellung des erweiterten Hochwasservorhersagesystem bei EGLV. © EGLV 77
Abbildung 46: Exemplarische Darstellung des Zeitreihen-Displays innerhalb von Delft-FEWS. © EGLV 78
Abbildung 47: Exemplarische Darstellung des räumlichen Displays innerhalb von Delft-FEWS, aktuelle Position des gewählten Fahrzeugs hervorgehoben durch gelben Punkt entlang der Strecke. © EGLV 78
Abbildung 48: Beispiel für umgesetzte Niederschlagswarnung auf Basis von MSPM-Gebietsniederschlägen. © EGLV 79
Abbildung 49: Kumulierter Gebietsniederschlag für ein Teileinzugsgebiet am 28.05.2019 (blau: DXD, rot: MSPM). © EGLV 79
Abbildung 50: Vergleich der Simulationsergebnisse zwischen DXD (blau) und MSPM (rot) am Pegel Buer-Sutum am Beispiel des Niederschlagsereignisses vom 28.05.2019. © EGLV 80
Abbildung 51: Vergleich des kumulierten Niederschlags von Niederschlagsradar- und Kfz-Messungen für vier Fahrzeuge. © IAV 82
Abbildung 52: Release Version 0.3 der MSPM. © IAV 83
Abbildung 53: Darstellung des Parameters RadarMax im Vergleich zu RadarTop. © IAV 83
Abbildung 54: Starkregensituation Beispiel 1 am 27.07.2019. © IAV 85
Abbildung 55: Starkregensituation Beispiel 2 am 13.08.2019. © IAV 85
Abbildung 56: Differenzenkarte zur Plausibilisierung. © IAV 85
Abbildung 57: Anzahl der Messungen pro Zelle und Zeitschritt im Histogramm (links) und im Boxplot (rechts). © FiW e. V. 86

TABELLENVERZEICHNIS

Tabelle 1: Auflistung der ausgelesenen Fahrzeugsignale. 25
Tabelle 2: Komponenten der IoT-Plattform. 33
Tabelle 3: Statistische Auswertung der Temperatursensoren. 43
Tabelle 4: State Definition. 48
Tabelle 5: Radar-Modell. 48
Tabelle 6: Pluvio-Modell. 48
Tabelle 7: Für eine mögliche Integration in das Informationsmodell untersuchte Drittdaten. 62
Tabelle 8: Vergleich von Pluviometer- und Radarwerten bei verschiedenen Niederschlagsintensitäten. 63
Tabelle 9: Informationen für die Datenassimilation. 64
Tabelle 10: Im Informationsmodell nicht verwendete Kfz-Sensordaten. 65
Tabelle 11: Im Informationsmodell verwendete Sensordaten mit einer Kurzbeschreibung. 68
Tabelle 12: Entwickelte Features für das KI-Modell. 72
Tabelle 13: Zeitliche Bedeutung des Zeitstempels bei verschiedenen Datenquellen. 72
Tabelle 14: Übersicht über exemplarische Starkregensituationen mit vorhandenen Kfz-Daten. 75
Tabelle 15: Niederschlagsradarwerte über den Fahrzeugen. 81
Tabelle 16: Starkregen Statistiken – RadarMax größer als 4 mm/5 min. 84
Tabelle 17: Qualitätsparameter der MSPM. 84

BILDNACHWEISE

Titel, obere Abbildung ..... Heatmap von Fahrstrecken im EGLV-Gebiet. © IAV

Titel, untere Abbildung ..... Serienmäßig in den Rückspiegelfuß integrierter Regen-Licht-Sensor. © EGLV

Seite 8 ..... © Igor Schubin/pixabay.com

Seite 32 ..... © Dibyendu Joardar/pixabay.com

Seite 50 ..... © Igor Schubin/pixabay.com

Seite 63 ..... © Holger Schué/pixabay.com

Seite 97 ..... © tookapic/pixabay.com

Projektpartner

mobileVIEW

Verbesserung der Kurzfristvorhersage von Niederschlagsereignissen mittels Fahrzeugsensoren