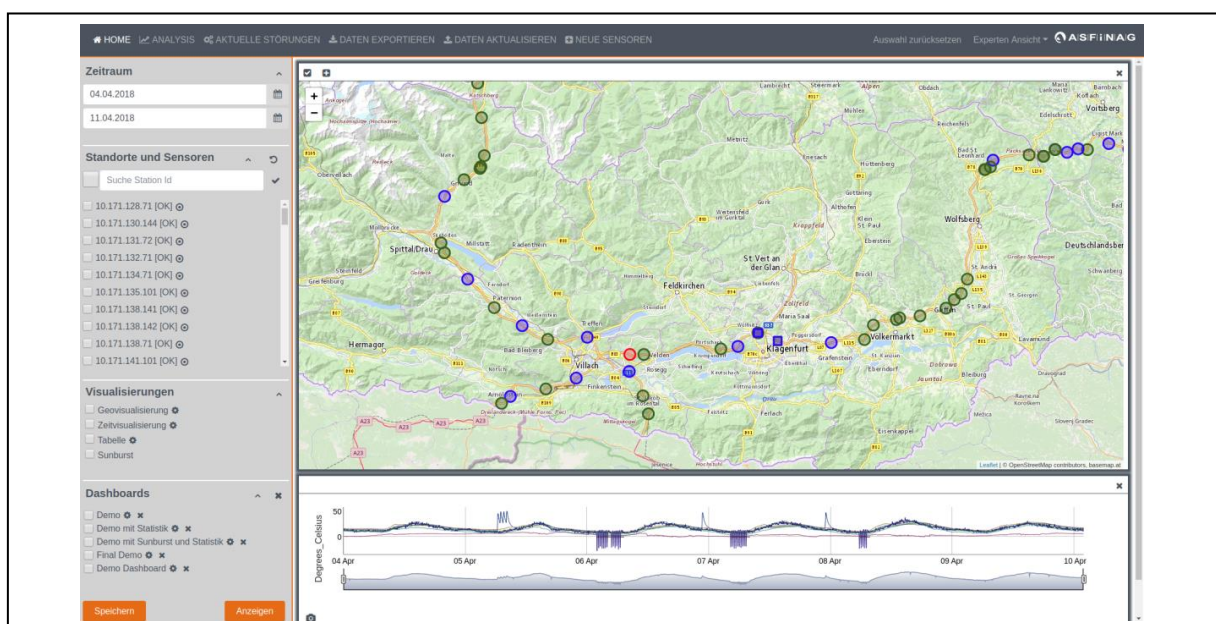


Qualitätssicherung der Umfelddatenerfassung UDEQI

Ein Projekt finanziert im Rahmen der
Verkehrsinfrastrukturforschung 2015
(VIF2015)

September 2018



Impressum:

Herausgeber und Programmverantwortung:

Bundesministerium für Verkehr, Innovation und Technologie
Abteilung Mobilitäts- und Verkehrstechnologien
Radetzkystraße 2
A – 1030 Wien



ÖBB-Infrastruktur AG
Nordbahnstraße 50
A – 1020 Wien



Autobahnen- und Schnellstraßen-Finanzierungs-
Aktiengesellschaft
Rotenturmstraße 5-9
A – 1010 Wien



Für den Inhalt verantwortlich:

Technische Universität Graz
Institut für Straßen- und Verkehrswesen
Rechbauerstraße 12/II
A-8010 Graz

Know-Center GmbH
Inffeldgasse 13/6
A-8010 Graz



Programmmanagement:

Österreichische Forschungsförderungsgesellschaft mbH
Thematische Programme
Sensengasse 1
A – 1090 Wien



Qualitätssicherung der Umfelddatenerfassung UDEQI

Ein Projekt finanziert im Rahmen der
Verkehrsinfrastrukturforschung
(VIF2015)

AutorInnen:

Michael CIK
Martin FELLENDORF
Roman KERN
Mark KRÖLL
Manuel LIENHART

Auftraggeber:

Bundesministerium für Verkehr, Innovation und Technologie
ÖBB-Infrastruktur AG
Autobahnen- und Schnellstraßen-Finanzierungs-Aktiengesellschaft

Auftragnehmer:

Technische Universität Graz (Institut für Straßen- und Verkehrswesen)
Know-Center GmbH

Abkürzungen

API	Application programming interface
BORRMA	Boschung Road & Runway Management
BüS	Betriebsüberwachungs System
CSV	Comma-separated values
Ctrl	Control (Strg-Taste)
DAG	Datenausgabegerät
DB	Database
DE	Daten-Endgeräte
DEG	Datenerfassungsgerät
DOF	Degrees of freedom
FBT	Fahrbahntemperatur
FG	Funktionsgruppe
FGSV	Forschungsgesellschaft für Straßen- und Verkehrswesen
FTP	File Transfer Protocol
GUI	Graphical User Interface
HK	Helligkeitssensor
hPa	Hektopascal
IP-Adresse	Internet Protocol Adresse
JSON	JavaScript Object Notation
KDD	Knowledge Discovery in Databases
NI	Niederschlagsintensität
NOK	Not OK
OpenWIS	Open Weather Information System
PM	Particulate matter (Feinstaub)
PNG	Portable Network Graphics
QSS	Qualitätssicherungs Sensorik Tool
RH	Relative humidity (Relative Luftfeuchtigkeit)
SQL	Structured Query Language
TLS	Transport Layer Security
UDE	Umfelddatenerfassung
VMIS	Verkehrsmanagement und Informationssystem
WFD	Wasserfilmdicke

INHALTSVERZEICHNIS

ABBILDUNGSVERZEICHNIS.....	6
TABELLENVERZEICHNIS.....	9
1 Einleitung.....	10
1.1 ASFiNAG Grundlagen	10
1.2 Umweltdatenerfassung	11
1.3 Sensorik	17
2 Aufbau wissenschaftlicher Prototyp	19
2.1 Datenverarbeitung und Datenhaltung	20
2.1.1 Datenvorverarbeitung	23
2.1.2 Zeitreihendatenbank	24
2.2 Algorithmen und Verfahren	26
2.2.1 Verfügbarkeit	27
2.2.2 Plausibilität	35
2.3 Graphische Benutzeroberfläche.....	44
2.4 Systemarchitektur.....	49
2.4.1 Module & Workflow.....	50
2.4.2 Vorteile der gewählten Architektur	51
3 Referenztestfeld.....	53
4 Validierung der prototypischen Ergebnisse anhand eines realen Anwendungsszenarios	59
5 Zusammenfassung	76
6 Ausblick	79
7 Referenzen	81

ABBILDUNGSVERZEICHNIS

Abbildung 1: Funktionsschema einer Verkehrsbeeinflussungsanlage	11
Abbildung 2: Überblick über das UDEQI System mit den wesentlichen funktionalen Komponenten wie (i) der Datenverarbeitung und Datenhalten, (ii) den Verfahren zur Berechnung der Verfügbarkeit sowie der Plausibilität im Analysemodul sowie (iii) der Interaktion des Benutzers mit dem System über die Client Web-GUI.....	19
Abbildung 3: Datenquellen Umfelddatenerfassung ASFiNAG.	21
Abbildung 4: OpenWIS Dateistruktur	22
Abbildung 5: UDE Dateistruktur.	22
Abbildung 6: Beispiel fehlende OpenWIS-Messwerte.	23
Abbildung 7: Beispiel fehlerhafte OpenWIS-Messwerte.	24
Abbildung 8: Beispielhafter Datenauszug aus der InfluxDB als CSV.....	25
Abbildung 9: Schematischer Aufbau der UDEQI InfluxDB - OpenWIS und UDE Daten werden getrennt abgespeichert.	26
Abbildung 10: Potentielle Fehler, die innerhalb eines drahtlosen Sensornetzwerkes auftreten können.....	26
Abbildung 11: Beispiel für die OpenWIS Station „10.171.34.71“ für das Jahr 2016.	29
Abbildung 12: Anzahl an Files pro Jahr bei OpenWIS Stationen für den Zeitraum 2012 - 2017.	30
Abbildung 13: Anzahl an leeren Files pro Jahr bei OpenWIS Stationen für den Zeitraum 2012 - 2017.	31
Abbildung 14: Beispiel für die UDE Station “A02_1 161,282”.	32
Abbildung 15: Fehlende Werte pro UDE Station in %.	33
Abbildung 16: Beispiele für kontextueller Ausreißer (links) und Punktausreißer (rechts).	35
Abbildung 17: Exemplarische Darstellung von kollektiven Ausreißern, die mittels einer Clusteranalyse als nicht plausibel eingestuft werden können.....	37
Abbildung 18: Vergleich von verschiedenen State-of-the-Art Algorithmen zur Outlier-Detection mit Bezug zu den Anforderungen im Projekt. Klammern bedeuten, dass die Anforderungen nur teilweise erfüllt werden.....	38
Abbildung 19: Anwendung linearer Regression zur Erkennung von Ausreißern. In grün: der Verlauf der gemessenen Sensorwerte; in blau der Sensorverlauf, wie er von der Regression ($T_2 + T_3$) vorhergesagt wird. Der Bereich in magenta zeigt die Stelle der Abweichung, in der der Ausreißer gefunden wurden.....	41

Abbildung 20: Links: Verlauf der Fahrbahntemperatur (FBT). Rechts: Verlauf der FBT (in rot) im Vergleich zu den FBT der benachbarten Stationen.	42
Abbildung 21: Links: Verlauf der Fahrbahntemperatur (FBT) und in rot die gekennzeichneten Ausreißer. Rechts: Verlauf der FBT (in rot) im Vergleich zu den FBT der benachbarten Stationen.	42
Abbildung 22: Links: Verlauf der Wasserfilmdicke (WFD) und in rot die gekennzeichneten Ausreißer. Rechts: Verlauf der WFD (in rot) im Vergleich zu den WFD der benachbarten Stationen.	43
Abbildung 23: Differenz zwischen den Werten von A02_1 004,128 (WFD) und der Schätzung basierend auf A02_1 002,358_RN_4 (links) sowie A23_2 001,198_RN_2 (rechts) am 2017-04-05.	43
Abbildung 24: „Einfache“ Ansicht von UDEQI mit Anzeige des Zeitraums, der Standorte und Sensoren und der Visualisierungen im linken Bereich und der Geo-Visualisierung aller OpenWIS sowie UDE Stationen in Österreich. Die entsprechende Färbung der Station spiegelt deren Zustand wieder.	45
Abbildung 25: Die Geo-Visualisierungskomponente der Dashboard-Anzeige.	46
Abbildung 26: Ganglinien zweier Sensoren (Wasserfilmdicke, Lufttemperatur) mittels LineChart Visualisierungen.	47
Abbildung 27: Manuelle Annotation von Sensorbereichen innerhalb eines LineCharts.	47
Abbildung 28: Einfacher Vergleich aller Sensoren zweier OpenWIS Stationen mittels Sunburst Visualisierungen.	48
Abbildung 29: Überblick über die UDEQI Systemarchitektur.	49
Abbildung 30: OpenWIS-Station auf der A1 an Streckenkilometer 246,52 in Richtung Walserberg, die Testsensorik ist in Rot hervorgehoben	54
Abbildung 31: Sensorenvergleich - Parameter Lufttemperatur (Genauigkeit: +/- 0.2°C bei -20°C bis +50°C) für den Zeitraum 01.01.2018 bis 30.06.2018	55
Abbildung 32: Sensorenvergleich - Parameter Luftdruck (Genauigkeit: +/- 0.5hPa bei 0 bis +40°C) für den Zeitraum 01.01.2018 bis 30.06.2018, exklusive von Messwerten außerhalb der Genauigkeit (bei Temperaturen von unter 0°C).....	56
Abbildung 33: Sensorenvergleich - Parameter Relative Feuchtigkeit (Genauigkeit: +/- 2% RH) für den Zeitraum 01.01.2018 bis 30.06.2018.....	57
Abbildung 34: Sensorenvergleich - Parameter Taupunkttemperatur (Genauigkeit: +/- 0.7°C) für den Zeitraum 01.01.2018 bis 30.06.2018.....	58

Abbildung 35: Übersicht aller OpenWIS und UDE Stationen. Die jeweilige Färbung codiert den aktuellen „Zustand“ einer Station (OK = grün, NOK = rot).	60
Abbildung 36: Übersicht aller Sensoren der Station nahe Velden mittels der Sunburst Visualisierung.	61
Abbildung 37: Auswahl aller Sensoren der Station nahe Velden mit der Einheit „Degree_Celsius“.	62
Abbildung 38: Auswahl des gewünschten Zeitraums und der gewünschten Visualisierung.	63
Abbildung 39: Zeitliche Darstellung der gewählten Sensoren der Station nahe Velden.	63
Abbildung 40: Zoomen in einem Bereich um die Daten genauer zu analysieren.	64
Abbildung 41: Zoomen in einem Bereich um die Daten genauer zu analysieren.	65
Abbildung 42: Markierung (Annotation) eines gewünschten Bereichs (in orange) für eine spätere Analyse.	66
Abbildung 43: Auswertung zur Verwendung in einem Bericht als PNG exportieren.	66
Abbildung 44: Auswahl der Sensoren, welche nicht in Ordnung sind („rot“ dargestellt)	67
Abbildung 45: Statistische Informationen zu den beiden Sensoren, die "rot" markiert waren.	67
Abbildung 46: Speichern der aktuellen Dashboard Ansicht zu späteren Wiederverwendung.	68
Abbildung 47: Exportieren des Dashboards (Daten und Statistiken) als CSV.	69
Abbildung 48: Daten Update - Status ändern.	70
Abbildung 49: Wiederherstellen des zuvor gespeicherten Dashboards.	70
Abbildung 50: Auswahl desselben Sensors (Cu_Temperature) zweier benachbarter Stationen.	71
Abbildung 51: Konfiguration um Sensoren von benachbarten Stationen in einer Auswertung gemeinsam anzeigen zu können.	72
Abbildung 52: Vergleich derselben Sensoren von zwei benachbarten Stationen – zoomen zeigt die größere Fluktuation eines Sensors.	72
Abbildung 53: Vergleich derselben Sensoren von zwei benachbarten Stationen – zoomen zeigt eine detailliertere Darstellung von Ausreißern.	73
Abbildung 54: Beschreibung (Annotation) eines gewünschten Bereichs.	74
Abbildung 55: Beschreibung (Annotation) eines gewünschten Bereichs.	74
Abbildung 56: Navigieren zwischen den einzelnen Annotationen erleichtert den Vergleich und die Analyse von Ausreißern.	75

Abbildung 57: Benutzerdefinierte Annotationen werden bei einem Export ebenfalls mitexportiert.....	75
Abbildung 58: Übersicht vorgeschlagene Funktionalitäten und deren Status innerhalb UDEQI	79
Abbildung 59: Mögliche Schnittstellenfunktion eines zukünftigen UDEQI-Produktsystems	80

TABELLENVERZEICHNIS

Tabelle 1: Statistiken für beide Sensordatentypen (OpenWIS, UDE)	28
Tabelle 2: Häufigkeit der einzelnen Sensortypen über alle UDE-Stationen	32
Tabelle 3: Überblick über alle im Projekt entwickelten und implementierten algorithmischen Ansätze, die auf ihre Anwendbarkeit hin untersucht wurden	39
Tabelle 4: Mögliche Zustände, die eine Station bzw. ein Sensor annehmen kann mit der entsprechenden farbliche Darstellung in der web-basierten Benutzeroberfläche	51
Tabelle 5: Vergleich der t-Test-Ergebnisse für den Vergleich der unterschiedlichen ermittelten Parameter beider Sensoren	55

1 EINLEITUNG

1.1 ASFiNAG Grundlagen

Zur Erhöhung der Verkehrssicherheit und Verbesserung des Verkehrsflusses werden auf Autobahnen und Schnellstraßen Verkehrsbeeinflussungsanlagen eingesetzt. „Durch die Vermeidung von Staus tragen Verkehrsbeeinflussungsanlagen auch zur Verringerung der Umweltbelastung durch den Straßenverkehr sowie von Zeit- und Energieverlusten bei.“ [1]

Im Netz der ASFiNAG bestehen diese meist „aus dynamischen Verkehrszeichen, die bei Bedarf vor Gefahren wie Stau, Unfällen oder witterungsbedingten Einschränkungen warnen und variable, situationsangepasste Geschwindigkeitsbeschränkungen oder Überholverbote anordnen und darüber hinaus Fahrstreifen sperren oder freigeben können.“ [1]

Die im Rahmen von Verkehrsbeeinflussungsanlagen (VBA) auf Autobahnen und Schnellstraßen ermittelten Umfelddaten stellen Eingangsgrößen einerseits für situationsabhängige automatische Schaltvorschläge dar, ebenso können diese durch menschliche Operatoren bei Ereignissen wie Unfällen, Pannen oder verlorenen Teilen auf der Fahrbahn vorgenommen werden (vgl. Abbildung 1). Mittels dynamischen Warnhinweisen, Geschwindigkeitsbeschränkungen oder Spurzuweisungen werden die Verkehrsteilnehmer vor potenziell gefährlichen Umfeldbedingungen (wie z.B. Nässe und Schneefall) gewarnt. [2]

Die Schaltungen aufgrund kritischer Fahrbahnzustände, Niederschlag und Sichtweite sind Elemente des Automatikbetriebs von VBAs. Auf die Verkehrssicherheit wirken sich Fahrbahn-nässe/-glätte und/oder Niederschlag in zweierlei Hinsicht negativ aus: zum einen wird durch die Nässe auf der Fahrbahn der Kraftschluss zwischen Reifen und Fahrbahn verringert, wodurch der Bremsweg erhöht und die aufnehmbaren Radialkräfte bei Kurvenfahrten reduziert werden. Zum anderen wird durch Niederschlag und Sprühfahnenbildung die Sicht des Kraftfahrers beeinträchtigt. Diesen negativen Einflüssen wird durch die Anzeige einer angepassten zulässigen Höchstgeschwindigkeit und ggf. von Warnhinweisen mittels Wechselverkehrszeichen (WVZ) begegnet.

Eingeschränkte Sicht ist für Verkehrsteilnehmer gefährlich, da Hindernisse bei zu hoher Geschwindigkeit nicht rechtzeitig erkannt werden können. Die Wahrnehmung möglicher Gefahrensituationen verzögert sich, woraus Auffahrunfälle resultieren können. Durch die Anzeige von Warnhinweisen und zulässigen Höchstgeschwindigkeiten soll der Anhalteweg eines Fahrzeuges an die vorhandene Sichtweite angepasst und so ein Auffahren auf verspätet erkannte Hindernisse vermieden werden.

Funktionsschema einer VBA

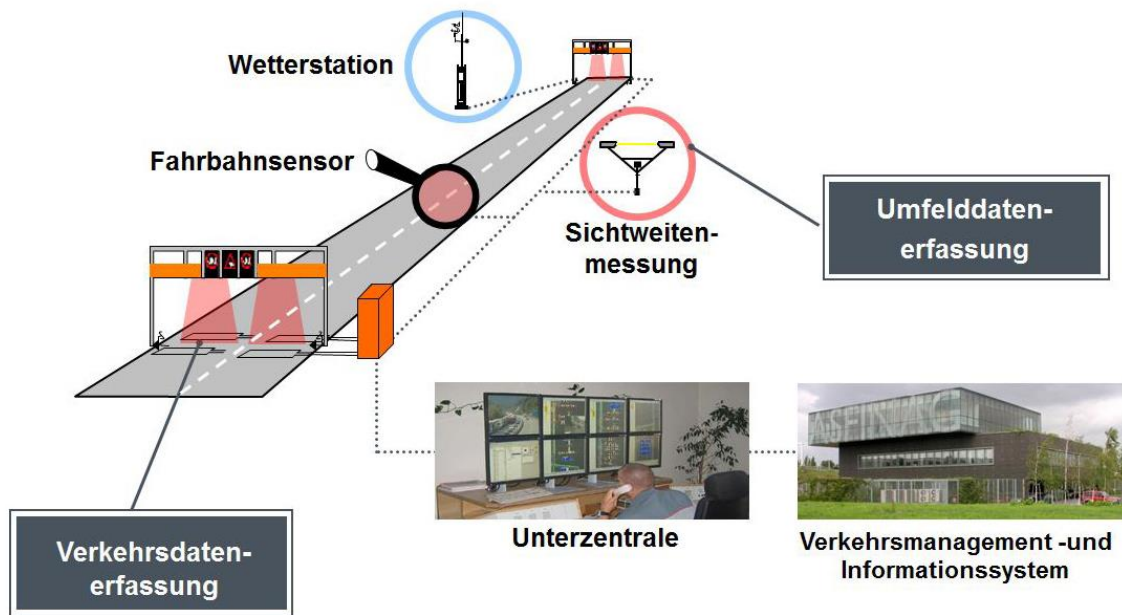


Abbildung 1: Funktionsschema einer Verkehrsbeeinflussungsanlage

Quelle: [2]

Witterungsbedingte Schaltungen können nur dann einen positiven Einfluss auf die Verkehrssicherheit erzielen, wenn die Anzeigen der VBA durch die Verkehrsteilnehmer befolgt werden. Hierfür ist es wiederum erforderlich, dass die stationär erfassten Umfelddaten eine hohe Qualität aufweisen und die aktuellen Umfeldbedingungen zeitnah im System abgebildet werden. Nur so kann in Abhängigkeit der vorliegenden Verkehrs- und Umfeldsituationen eine sinnvolle und wirksame Beeinflussung des Verkehrsablaufs stattfinden.

1.2 Umweltdatenerfassung

Aktuell wird im Streckennetz der ASFiNAG die Qualität der Systeme (bzw. Sensoren) zur Umfelddatenerfassung weder systematisch noch vergleichend untersucht und speziell im täglichen Betrieb der Umfelddatenerfassung werden Fehler häufig nicht oder erst spät bzw. zufällig erkannt.

Aufgrund der Wichtigkeit der Umfelddatenerfassung und des für die Praxis zu erwartenden Nutzens bestand das Ziel dieses Forschungsvorhabens einerseits darin, eine Verfügbarkeitsaussage pro Sensor und für alle Sensoren gemeinsam zu haben, andererseits eine Performance-Kennzahl pro Sensor (Wie gut sind die Messwerte und deren Qualität tatsächlich?) zu entwickeln. Dabei wurde eine Plausibilisierung zwischen den

unterschiedlichen Sensoren sowie ein Längsabgleich (entlang der Strecke, zwei benachbarte Stationen, Datenfusion) angestrebt.

Witterungsereignisse sind inhomogene und instationäre Ereignisse, deren genaue Ausprägung schwer überprüfbar ist. Um die Umfeldgrößen, bzw. die Sensoren zu deren Erfassung praxisnah und längerfristig zu untersuchen, wird in Deutschland seit 2003 bei München ein Testfeld für Umfelddatenerfassung betrieben [3]. In diesem Testfeld werden Umfeldsensoren verschiedener Hersteller mit zum Teil stark variierenden Messverfahren zur Bestimmung ein und derselben Messgröße unter gleichen Bedingungen untersucht und verglichen. Dadurch sollen die Erfassungssysteme bzgl. ihrer Eignung für den Einsatz in Streckenbeeinflussungsanlagen (SBA) bzw. Verkehrsbeeinflussungsanlagen (VBA) eingestuft und gleichzeitig die Erfassungssysteme (Hardware und Software) verbessert werden. Dieser Punkt wurde im Projektkonsortium konzeptuell diskutiert. Eine solche Vorgehensweise erlaubt die Ableitung von Anforderungen für Ausschreibung, Abnahme und Betrieb von Umfeldsensoren.

Der Nutzen des Projekts liegt für die Betreiber von SBA/VBA in einer verbesserten, einheitlichen und wirtschaftlichen Erfassung von Umfelddaten. Für die Verkehrsteilnehmer ergeben sich hieraus nachvollziehbare und plausible Schaltungen, was durch entsprechende Akzeptanz zu einer Erhöhung der Verkehrssicherheit führt.

Umfelddatenerfassung in Verkehrsbeeinflussungsanlagen

Die Erfassung und Nutzung der Umfelddaten im Rahmen der SBA (VBA) ist in dem Merkblatt für die Ausstattung von Verkehrsrechnerzentralen und Unterzentralen – MARZ [4] und den Technischen Lieferbedingungen für Streckenstationen – TLS [5] festgelegt. Neben verbindlichen Anforderungen an die Funktionen und Schnittstellen der Funktionsebenen wird folgende Architektur für SBA definiert:

- Verkehrsrechnerzentrale (VRZ),
- Unterzentrale (UZ) und
- Streckenstation (SSt) mit Steuermodulen (SM) und Eingabe-/Ausgabekonzentratoren (EAK) einschließlich Datenerfassungsgerät (DEG) und Datenausgabegerät (DAG).

Zwischen den Ebenen erfolgt die Kommunikation über eine Bus-Topologie.

Die Wetter- und Umfelddaten werden in der TLS der Funktionsgruppe (FG) 3 und den entsprechenden DE (Daten-Endgeräte-Kanal)–Typen zugeordnet, die erfassten Daten sind

binär kodiert. Im Regelbetrieb erfolgt die Abfrage durch reines Polling (SSt->Detektor) mit einer Übertragung der Messwerte im 1-Minutenintervall.

In Testfeld Deutschland [3] wurden die sog. Primärgrößen

- Niederschlagsintensität,
- Wasserfilmdicke,
- Sichtweite

und weitere Messgrößen (zur Plausibilisierung der Primärgrößen und als Stützgrößen für Prognosemodelle) im Wissenspapier „Hinweise zur Erfassung von Umfelddaten in Streckenbeeinflussungsanlagen“ des FGSV AK 3.2.1 [6] beschrieben. In dem Dokument werden zudem die auf meteorologischen Zusammenhängen basierenden Plausibilitätsüberprüfungen der Messgrößen, die Ermittlung der Zustände „Nässe“, „Sichtweite“ und „Glätte“ sowie die Schaltalgorithmen dargestellt. Erste Anforderungen an Umfeldgrößen zur Nässeerfassung wurden im „Merkblatt für Nässeerfassung in Streckenbeeinflussungsanlagen“ [7] definiert.

Exemplarisch sind einige Messgrößen und deren Auswirkungen auf die Verkehrsteilnehmer dargestellt:

(1) Sichtweite

Einschränkungen der Sichtweite können durch unterschiedliche Ereignisse hervorgerufen werden:

Eingeschränkte Sicht aufgrund von Niederschlag

Niederschlag führt, insbesondere bei Nacht, zu einer zum Teil erheblichen Sichteinschränkung für den Kraftfahrer. Dies resultiert einerseits aus den Regentropfen in der Luft und andererseits aus dem Wasser auf der Windschutzscheibe, welches sich gegebenenfalls mit Schmutz, Öl oder ähnlichem vermischt. Für beide Fälle kommt es darüber hinaus zu einer verstärkten Blendung. Untersuchungen von SCHARSCHING [8] zufolge hängt die Beeinträchtigung der Sicht bei Regen weitgehend von der Niederschlagsintensität ab. Bei Nacht spielt die Tröpfchengröße eine noch bedeutendere Rolle als die Stärke des Regens. So bewirken besonders viele kleine Tröpfchen, z. B. bei Nieselregen mit Durchmesser kleiner als 0,5 mm, eine Verringerung des Kontrastes und der Strukturierung des Hintergrundes, wodurch es zu einer Sichtweitenbeschränkung auf bis unter 200 m kommen kann. Bei größeren

Tröpfchendurchmessern liegt die Sichtweite hingegen kaum unter 3000 m. In stärkerem Ausmaß als bei Regen beeinträchtigen auch bei Schneefall mehrere kleine Flocken die Sicht bei gleicher Niederschlagsintensität stärker als eine geringere Anzahl großer Flocken. In Verbindung mit Wind (Schneetreiben) wird die Sicht noch weiter verschlechtert. Hierbei wird je nach Windstärke auch noch der am Boden liegende Schnee (falls vorhanden) aufgewirbelt, was bei Sturm und starkem Schneefall zu Sichtweiten unter 10 m führen kann. Bei vorhandener Schneedecke ist überdies auch die Wahrnehmung der Umwelt verändert, da Konturen und Kontraste schwerer zu erkennen sind.

Eingeschränkte Sicht aufgrund von Fahrbahnnässe (Sprühfahnen)

Auf nasser Fahrbahn wird die Sicht durch Sprüh- und Spritzwasser vorausfahrender Fahrzeuge beeinträchtigt. Für die Bildung von Sprühwasser wird das Wasser zuerst vom Reifen aufgenommen und dann in die Luft gewirbelt (Sprühfahnen), bei Spritzwasser wird das Wasser aus der Kontaktzone zwischen Reifen und Fahrbahn verdrängt. Einfluss auf das Ausmaß der Bildung von Sprüh- und Spritzwasser haben Fahrgeschwindigkeit, Wasserfilmdicken, sowie Fahrbahn- und Reifenoberfläche. Für die Steuerung von SBA/VBA kann diese Art von Sichtweitenreduktion nicht direkt gemessen werden, gerade sie stellt jedoch ein großes Risiko für die Verkehrssicherheit dar. Die Sichteinschränkung durch Sprühfahnen wird i. d. R. durch die Nässeerfassung abgedeckt, so dass diese hier nicht gesondert betrachtet wird.

Eingeschränkte Sicht aufgrund von Nebel

Für den Verkehrsteilnehmer „plötzlich“ eintretende Sichtweitenreduktionen werden häufig durch Nebel hervorgerufen. Nebel wird meteorologisch als eine auf dem Boden aufliegende Wolke aus Wassertröpfchen mit Durchmessern von 0,1 µm bis 50 µm und teilweise schwarzem Kondensationskern bezeichnet. Es wird zwischen Advektionsnebel und Strahlungsnebel unterschieden: Advektionsnebel besteht aus driftenden Nebelmassen, die entstehen, wenn warme Luft über kalten Boden streicht und es dabei zu einer Kondensation der Luftfeuchtigkeit kommt. Strahlungsnebel tritt am Häufigsten auf und entsteht, wenn es bei hoher Luftfeuchtigkeit zu einer stärkeren Abkühlung des Bodens als der Luft kommt (z. B. nachts) und dadurch die Luftfeuchte kondensiert. Der entscheidende Unterschied ist, dass Advektionsnebel wandert, wohingegen Strahlungsnebel nur lokal begrenzt auftritt. Die Wirkung von Nebel auf die Sichtweite wird von MANGOLD [9] wie folgt beschrieben: viele kleine in der Luft befindlichen Aerosole werfen einen Schatten oder decken Bereiche aus dem Sichtfeld des Verkehrsteilnehmers vollständig ab. Zusätzlich wird das Licht an diesen

Wassertröpfchen reflektiert und gestreut. Dies verursacht die Einschränkung der Sicht für den Fahrer. Zudem kommt die subjektive Einschätzung der Sichtweite durch den Fahrer hinzu, was im Prinzip für alle Niederschlagsereignisse gültig ist. Für die streckenbezogene Erfassung von Nebel, wie sie für die Steuerung von SBA/VBA benötigt wird, besteht die Problematik, dass Nebel auch auf nur kurzen Abschnitten, jedoch mit sehr starker Sichtweitereinschränkung auftreten kann und deshalb von Messgeräten, die in zu großem Abstand positioniert sind, eventuell überhaupt nicht erfasst wird. Die Sichtweite wird z.B. in Deutschland im Hinweispapier für die Erfassung von Umfelddaten [10] hysteresebasiert in (mindestens) 6 Sichtweistestufen klassifiziert.

(2) Niederschlagsintensität

Die Niederschlagsintensität ist ein wesentlicher Parameter bei der Warnung der Kraftfahrer vor Nässe. Gemäß den Hinweisen zur Erfassung und Nutzung von Umfelddaten in SBA/VBA [10] bzw. dem Merkblatt "Nässeerfassung in SBA" ist die Messgröße Niederschlagsintensität bei der Steuerung zu berücksichtigen. Auf Grundlage der gemessenen Intensitäten werden durch Schwellenwerte Niederschlagsstufen festgelegt. Im Hinweispapier für Umfelddaten [10] werden hysteresebasierte Grenzwerte zur Klassifizierung der NI-Stufen vorgeschlagen.

(3) Wasserfilmdicke

Die Wasserfilmdicke (WFD) wird durch drei wesentliche Merkmale bestimmt [11]:

- Witterungsverhältnisse (Niederschlagsintensität, Niederschlagsdauer, Temperatur, Wind und Luftfeuchtigkeit);
- Fahrbahneigenschaften (Neigung, Oberflächentextur und Unebenheiten);
- Verkehrseinflüsse (Reibungswärme, Luftbewegung, Versprühen des Wassers).

Für den Einfluss der beiden erstgenannten Merkmale gibt es eine Vielzahl mathematischer Ansätze zur Bestimmung der Wasserfilmdicke.

Für verkehrstechnische Untersuchungen ist neben der Wasserfilmdicke auch die Dauer der Fahrbahnnässe von Bedeutung. Da die Dauer der Abtrocknungsphase der Fahrbahnoberfläche von weiteren Einflüssen abhängt, sind neben der Niederschlagsdauer und -intensität zudem die relative Luftfeuchte, die Lufttemperatur, die Windgeschwindigkeit und die Verkehrsbelastung relevant. Nach LANG [12] ist Nässe auf der Fahrbahn häufig doppelt so lange wie das eigentliche Niederschlagsereignis vorhanden.

Das Hinweispapier für die Erfassung von Umfelddaten [10]) sowie das Merkblatt zur Nässeerfassung beinhalten Vorschläge für hysteresebasierte Schwellenwerte zur Klassifizierung von WFD-Stufen.

Datengrundlage im ASFiNAG-Netz und daraus resultierende Problemstellung:

Entlang des ASFiNAG-Straßennetzes sind unterschiedlichste Umfelddatensensoren für die Bereiche Verkehrssteuerung und Winterdienst verbaut, welche als Grundlage für das Projekt zu verwenden sind:

System Verkehrssteuerung:

- Lufttemperatur
- Wasserfilmdicke
- Niederschlagsart
- Niederschlagsintensität
- Gefrierpunkttemperatur
- Sichtweite
- Fahrbahnzustand
- Windrichtung
- Windgeschwindigkeit
- Relative Luftfeuchte
- Helligkeit

System Winterdienst:

- Lufttemperatur
- Wasserfilmdicke
- Niederschlagsart
- Niederschlagsintensität
- Gefrierpunkttemperatur
- Sichtweite
- Fahrbahnzustand
- Windrichtung
- Windgeschwindigkeit
- Relative Luftfeuchte
- Taustoffkonzentration
- Taupunkttemperatur

1.3 Sensorik

Die Datengrundlage für Schaltvorgänge der Verkehrsbeeinflussungsanlagen und Streckenkenndaten für den ASFiNAG-Winterdienst werden über unterschiedliche Sensoriktypen zur Verfügung gestellt:

- Fahrbahnoberflächensensoren: Wasserfilmdicke, Restsalzmenge, Fahrbahnoberflächentemperatur, Gefriertemperatur, Bodentemp.1 bei 5cm, Bodentemp.2 bis 30 cm, Bodentemp.3 ab 30cm, Taustoffkonzentration
- Relative Luftfeuchte, Lufttemperatur, Taupunkttemperatur
- Windrichtung, Windgeschwindigkeit [Mitte], Windgeschwindigkeit [Spitze]
- Sichtweitensensoren (VBA, Nebelwarnanlage)
- Sichtweite, Niederschlagsart, Niederschlagsintensität
- OpenWIS-Standorte (VBA/Glättemeldeanlagen)
- Glättemeldeanlagen-Standorte

Das Ziel des Forschungsvorhabens war, einerseits eine Verfügbarkeitsaussage pro Sensor und über alle Sensoren zu haben, andererseits eine Performance-Kennzahl (Plausibilisierung) pro Sensor zu entwickeln. Dabei sollte eine Plausibilisierung zwischen den unterschiedlichen Sensoren sowie ein Längsabgleich (entlang der Strecke, zwei benachbarte Stationen, Datenfusion) angestrebt werden.

Dies ist in einem lauffähigen, performanten Prototyp umzusetzen, in dem die Werte der einzelnen Sensoren (je nach Sensortyp + Verfügbarkeitswerte) sowie deren Qualität (Plausibilisierung) bewertet werden. Die Daten sollten in Echtzeit abgerufen und bewertet werden können und fehlerhafte oder auffällige Sensoren sollten in einer eigenen Auswertung ausgegeben werden. Zusätzlich sind die ermittelten Minutendaten einerseits einem Monitoringtool und die userspezifischen Auswertungen auf einem Graphical User Interface (GUI) anzuzeigen (Langzeitdaten, Verfügbarkeit und Performance).

Abschließend ist anzumerken, dass sowohl in den geltenden nationalen als auch in internationalen Richtlinien noch keine Methoden zur Prüfung der Genauigkeit von Umfelddaten definiert sind.

Weiters wurde auf Basis der Ergebnisse von international ähnlich orientierten Projekten ersichtlich, dass eine formale Statistik-orientierte Vorgehensweise zur Analyse der Umfelddatensensordaten nicht durchgängig anwendbar und auch nicht zweckmäßig ist [3].

Gründe dafür sind:

- Die Heterogenität der Ergebnisse
Beim Vergleich der Zeitreihen der verschiedenen Sensoren zeigte sich häufig eine heterogene Verteilung der Messwerte unterschiedlicher Sensoren, so dass eine statistische Auswertung mit der Ermittlung von Abweichmaßen nicht zielführend ist.
- Fehlende bzw. zu wenige Referenzmessungen und -beobachtungen
Für viele der Messgrößen standen keine (Bodentemperatur, Windgeschwindigkeit, Windrichtung) bzw. statistisch nicht aussagekräftige (Temperatur der Fahrbahnoberfläche, Wasserfilmdicke) oder "ungleiche" (Niederschlagsart: große Stichprobe für Schnee durch Video, praktisch keine für Regen) Referenzmessungen zur Verfügung. Somit konnten keine bzw. nur statistisch nicht aussagekräftige Angaben zu Abweichungen von der "Realität" getroffen werden.

Unter Berücksichtigung der o.g. Erkenntnisse konnte in diesen Untersuchungen eine Bewertung der zu prüfenden Systeme eher qualitativ durchgeführt werden, die Messwerte der verschiedenen Sensoren wurden inhaltlich und funktional/technisch untersucht.

Basierend auf diesen Erkenntnissen sollte in diesem Forschungsprojekt die Bewertung der Messwerte hinsichtlich inhaltlicher Kriterien folgende zusätzliche Auswertungen umfassen:

- Abweichung von Referenzmessungen (wurde im Projekt im Rahmen eines „virtuellen“ Testfeldes umgesetzt)
- Abweichung von Beobachtungen vor Ort
- Plausibilität im Vergleich zu anderen Sensoren.

Diese zusätzlichen Datenerhebungen dienen der Validierung der unterschiedlichen Messgrößen der verschiedenen Sensortypen als Referenzmessgrößen.

2 AUFBAU WISSENSCHAFTLICHER PROTOTYP

In diesem Kapitel wird der Aufbau des wissenschaftlichen Prototyps näher erläutert – das UDEQI-System ist schematisch in Abbildung 2 dargestellt. Den Input von UDEQI stellen dabei die bestehende Sensorinfrastruktur sowie schon verfügbare Sensordatenbanken dar. Während die Datenbanken nur einmal in das UDEQI System importiert werden, stellt der Sensorinput einen kontinuierlichen Prozess dar. Der oder die BenutzerIn interagiert mit der UDEQI-Web-GUI und wird somit in den täglichen Arbeitsabläufen unterstützt. Wesentliche Operationen, wie beispielsweise schnell einen Überblick über den aktuellen Zustand des Sensornetzwerks zu bekommen, werden zur Verfügung gestellt. Zudem besteht die Möglichkeit, Auswertungen für spätere Analysen, Diskussionen oder auch Berichte zu erstellen. Dazu wird nicht direkt auf die Datenbank zugegriffen, sondern ausschließlich auf das UDEQI-Analysemodul. Dieses stellt das zentrale Modul dar und ist verantwortlich für die Implementation der heuristischen, statistischen und maschinellen Lernverfahren, die Aussagen über die Verfügbarkeit sowie die Plausibilität der Sensorik zulassen. Zudem kapselt diese Komponente den Zugriff auf die Datenbank, sodass die Datenbankzugriffe eingeschränkt werden können. Erfahrungsgemäß wiederholen sich gleiche oder sehr ähnliche Zugriffsmuster häufig. Daher ist eine Caching-Komponente vorgesehen, in der häufig wiederkehrende Auswertungen zwischengespeichert werden können.

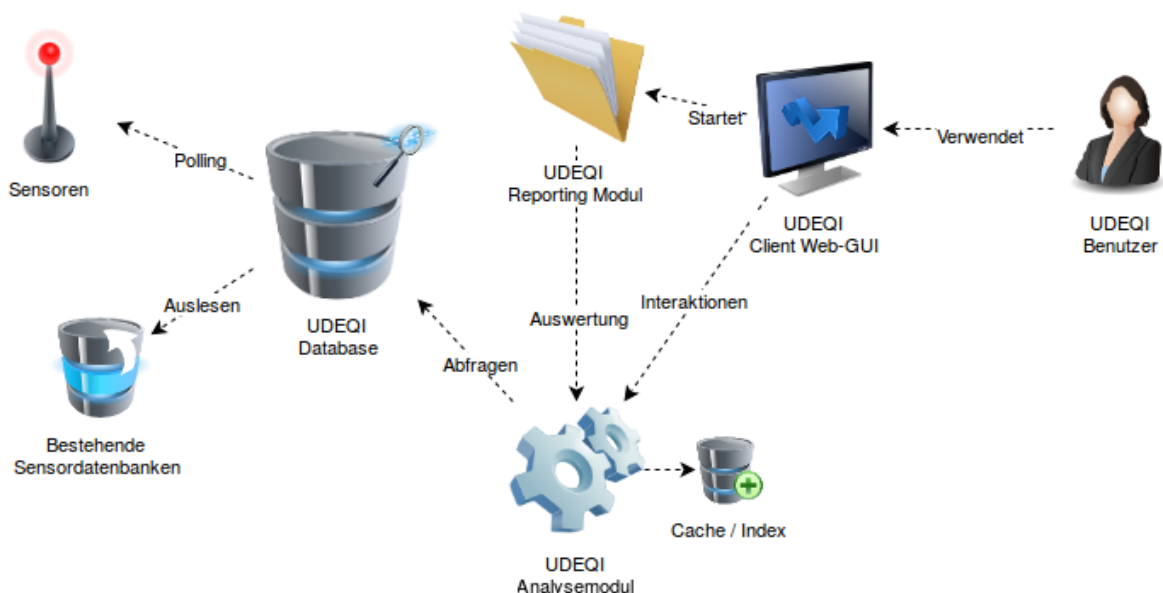


Abbildung 2: Überblick über das UDEQI System mit den wesentlichen funktionalen Komponenten wie (i) der Datenverarbeitung und Datenhalten, (ii) den Verfahren zur Berechnung der Verfügbarkeit sowie der Plausibilität im Analysemodul sowie (iii) der Interaktion des Benutzers mit dem System über die Client Web-GUI.

Die einzelnen UDEQI Systemkomponenten wurden jeweils aus Microservices¹ konzipiert. Dadurch können die Komponenten separat entwickelt, getestet und auch ausgerollt werden. Das Microservice-Architekturmuster stellt den aktuell besten Kompromiss aus Flexibilität und Funktionalität dar. Die Softwareentwicklung selbst erfolgte iterativ und agil, basierend auf tagtäglichen Anwendungsszenarien der ASFiNAG. Diese Art der Softwareentwicklung hat sich in den letzten Jahren als „*Best Practice-Methode*“ etabliert.

In den folgenden Abschnitten werden die drei Hauptkomponenten des wissenschaftlichen Prototyps detailliert beschrieben:

- Datenverarbeitung und Datenhaltung (vgl. 2.1)
- Verfahren zur Berechnung der Verfügbarkeit sowie der Plausibilität (vgl. 2.2)
- Interaktion des Benutzers über die graphische Benutzeroberfläche (vgl. 2.3)

Das Zusammenwirken der drei Hauptkomponenten sowie eine Beschreibung der Vorteile der gewählten Architektur sind in Abschnitt 2.4 zusammengefasst.

2.1 Datenverarbeitung und Datenhaltung

Aktuell werden auf Seiten der ASFiNAG drei verschiedene Systeme verwendet (Abbildung 3). In Absprache mit den Projektverantwortlichen der ASFiNAG werden Sensordaten aus zwei unterschiedlichen Datenquellen in der Datenbank gespeichert und weiterverarbeitet, 1.) dem OpenWIS² System, bei dem die Daten als JSON³-Dateien bereitgestellt werden 2.) sowie einer SQL⁴-Datenbank, aus der die UDE⁵-Daten stammen. Das System BORRMA⁶-web, welches Daten der älteren Boschung-Sensorik bereitstellt, wird nicht integriert, da die Sensorik mittelfristig ausgetauscht werden soll.

¹ <http://microservices.io/>

² Open Weather Information System

³ JavaScript Object Notation

⁴ Structured Query Language

⁵ Umfelddatenerfassung

⁶ Boschung Road & Runway Management

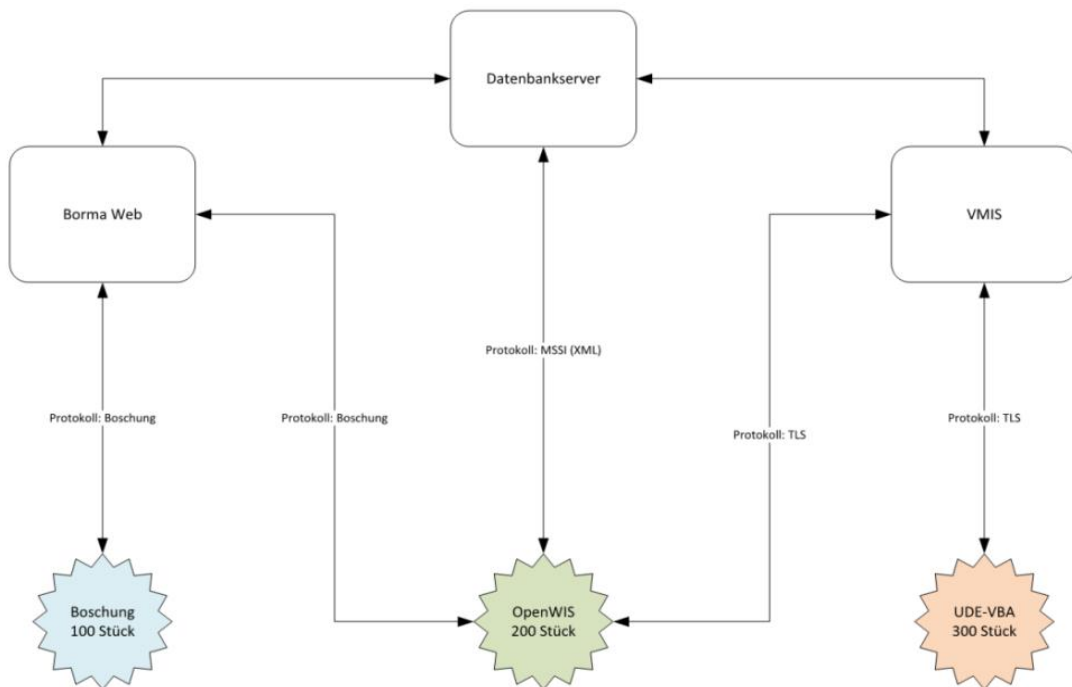


Abbildung 3: Datenquellen Umfelddatenerfassung ASFiNAG.

Im Folgenden wird die Struktur der Dateiablage bei OpenWIS Daten als auch bei UDE Daten erläutert.

OpenWIS-Daten

„OpenWIS ist eine Offene Hard- und Software Plattform für Wetter Informations Systeme. OpenWIS möchte so allgemein wie möglich sein um für eine breite Palette von Anwendungen verwendbar zu sein. Die OpenWIS Software-Architektur ist modular aufgebaut und besteht aus einzelnen Modulen die in Python implementiert sind. Die einzelnen Module werden über separate Projektseiten verwaltet um möglichst unabhängig voneinander weiterentwickelt werden zu können. Ein definiertes API stellt die Schnittstelle unter den Modulen zur Verfügung.“ [13]

Die Input-Dateien liegen in einem Dateisystem vor, das wie folgt strukturiert ist:

ID der Sensorstation → Jahr → Monat

In Monatsordnern liegen die Sensor-Dateien, tageweise gruppiert (Abbildung 4). OpenWIS Daten sind nach folgendem Schema aufgebaut: Jedes Sensorenbündel (bestimmt durch eine

ID) beinhaltet Sensordaten, sortiert nach dem Zeitpunkt der Aufnahme. Pro Tag wird eine eigene CSV-Datei angelegt.

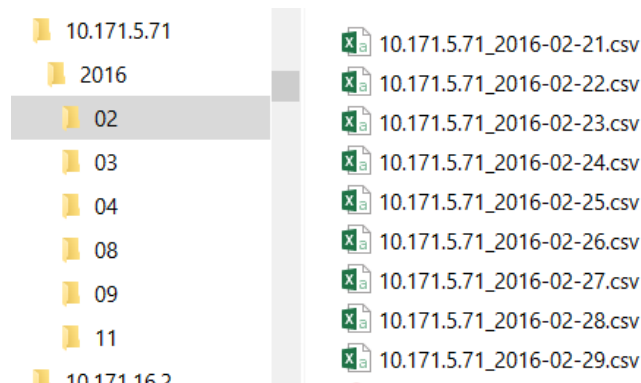


Abbildung 4: OpenWIS Dateistruktur

Jede Zeile in einer solchen Sensor-Datei besteht dabei aus einem Zeitstempel und einem Messwert pro Spalte. Einzelne Zellen können dabei Messwerte, Übertragungsfehler oder leere Felder beinhalten, was eine Vorverarbeitung der Daten notwendig macht (vgl. 2.1.1).

UDE-Daten

Die Daten des UDE-Systems stehen über eine SQL-View zur Verfügung und können mittels einer entsprechenden SQL-Abfrage ausgelesen werden (Abbildung 5).

Eine Zeile des Abfrage-Ergebnisses besteht dabei unter anderem aus:

Zeitstempel, Name der Autobahn bzw. Schnellstraße plus Fahrtrichtung, Kilometerangabe des Messpunktes, Typ des Sensors und Messwert.

S_Stamp	NAME	ROUTE	KM_POS	RO...	WERT	STA...	W_NAME	W_EINH	REF
2016-10-04 09:05:00.0	WTA_A21_1_034.595-HK2	A21_1	034.595	-1	942	0	HK	Lx	21643
2016-10-04 09:06:00.0	AQ_A021_1_006.955-HK2	A021_1	006.955	1.5	687	0	HK	Lx	20305
2016-10-04 09:06:00.0	AQ_A021_1_006.955-HK1	A021_1	006.955	1.5	340	0	HK	Lx	20306
2016-10-04 09:05:00.0	AQ_A02_1_006.955-HK2	A021_1	006.955	2.5	257	0	HK	Lx	20321
2016-10-04 09:06:00.0	AQ_A02_1_006.955-HK1	A021_1	006.955	2.5	383	0	HK	Lx	20322
2016-10-04 09:06:00.0	AQ_A02_2_016.450-HK1	A02_2	016.450	2.5	597	0	HK	Lx	21016
2016-10-04 09:06:00.0	AQ_A02_2_016.450-HK2	A02_2	016.450	2.5	383	0	HK	Lx	21017
2016-10-04 09:06:00.0	AQ_A21_1_000.383.R38.1-HK2	A02.Ra4.6	000.383	1.5	774	0	HK	Lx	21704
2016-10-04 09:06:00.0	AQ_A21_1_000.383.R38.1-HK1	A02.Ra4.6	000.383	1.5	467	0	HK	Lx	21705
2016-10-04 09:06:00.0	AQ_A02_1_000.259.R4.4-HK2	A02.Ra4.5	000.259	1.5	130	0	HK	Lx	21684
2016-10-04 09:06:00.0	AQ_A02_1_000.259.R4.4-HK1	A02.Ra4.5	000.259	1.5	340	0	HK	Lx	21685
2016-10-04 09:06:00.0	AQ_A21_1_037.330_AS-HK2	A21_1	037.330	0	0	0	HK	Lx	21659
2016-10-04 09:06:00.0	AQ_A21_1_037.330_AS-HK1	A21_1	037.330	0	0	0	HK	Lx	21660
2016-10-04 09:06:00.0	AQ_A21_1_037.330_IS-HK2	A21_1	037.330	9	1.042	0	HK	Lx	21661
2016-10-04 09:06:00.0	AQ_A21_1_037.330_IS-HK1	A21_1	037.330	9	512	0	HK	Lx	21662

Abbildung 5: UDE Dateistruktur.

2.1.1 Datenvorverarbeitung

Um die Daten konsistent und einheitlich auszubreiten, wurden mittels einer SpringBoot⁷ Applikation in Abhängigkeit vom Datentyp (OpenWIS vs. UDE) folgende Vorverarbeitungsschritte eingeführt:

UDE-Daten

Die UDE-Daten wurden nach Zeitstempel gruppiert und aggregiert, sodass alle Messwerte, die an einer Sensorstation zum selben Zeitpunkt empfangen wurden, als ein Messpunkt mit „1...n“ Messwerten dargestellt wurde. Diese Gruppierung wurde eingeführt, um ein einheitliches Datenlayout der UDE-Daten und OpenWIS-Daten zu gewährleisten.

OpenWIS-Daten

Für die OpenWIS-Daten wurde eine Filterung implementiert, die nicht vorhandene (Abbildung 6) beziehungsweise fehlerhafte (Abbildung 7) Datenpunkte aus den finalen Messpunkten entfernen soll.

Timestamp (UTC)	Timestamp (Europe/Vienna)	ARS31 Threat of icing (n/a)/ARS31_TOI(0)	temp external 1 (Degrees_Celsius)/BT(0)
4/18/2016 0:00	4/18/2016 2:00	2 (threat not detected)	16.66573143
4/18/2016 0:01	4/18/2016 2:01	2 (threat not detected)	16.65079117
4/18/2016 0:02	4/18/2016 2:02	2 (threat not detected)	16.63882256
4/18/2016 0:03	4/18/2016 2:03	2 (threat not detected)	16.62686729
4/18/2016 0:04	4/18/2016 2:04	2 (threat not detected)	16.61789131
4/18/2016 0:05	4/18/2016 2:05	2 (threat not detected)	16.60592842
4/18/2016 0:06	4/18/2016 2:06	2 (threat not detected)	16.59694672
4/18/2016 0:07	4/18/2016 2:07		
4/18/2016 0:08	4/18/2016 2:08	2 (threat not detected)	16.57598877
4/18/2016 0:09	4/18/2016 2:09	2 (threat not detected)	16.56401062
4/18/2016 0:10	4/18/2016 2:10	2 (threat not detected)	16.55501747
4/18/2016 0:11	4/18/2016 2:11	2 (threat not detected)	16.54303169
4/18/2016 0:12	4/18/2016 2:12	2 (threat not detected)	16.53403282
4/18/2016 0:13	4/18/2016 2:13	2 (threat not detected)	16.52503400

Abbildung 6: Beispiel fehlende OpenWIS-Messwerte.

⁷ <https://projects.spring.io/spring-boot/>

Timestamp (UTC)	Timestamp (Europe/Vienna)	ARS31 Threat of icing (n/a)/ARS31_TOI(0)	temp external 1 (Degrees_Celsius)/BT(0)	R2S LT (Degrees_Celsius)/LT(0)
4/18/2016 0:00	4/18/2016 2:00	2 (threat not detected)	16.66573143	12.15999985
4/18/2016 15:56	4/18/2016 17:56	2 (threat not detected)	21.64646149	12.63000011
4/18/2016 15:57	4/18/2016 17:57	2 (threat not detected)	21.61956024	12.47000027
4/18/2016 15:58	4/18/2016 17:58	2 (threat not detected)	21.58255768	12.47000027
4/18/2016 15:59	4/18/2016 17:59	2 (threat not detected)	21.54549789	12.47000027
4/18/2016 16:00	4/18/2016 18:00	2 (threat not detected)	21.50168419	12.47000027
4/18/2016 16:01	4/18/2016 18:01	2 (threat not detected)	21.45780182	12.55000019
4/18/2016 16:02	4/18/2016 18:02	2 (threat not detected)	21.41386986	12.47000027
4/18/2016 16:03	4/18/2016 18:03	2 (threat not detected)	21.36651802	12.47000027
4/18/2016 16:04	4/18/2016 18:04	2 (threat not detected)	21.31910896	12.47000027
4/18/2016 16:05	4/18/2016 18:05		21.27502251	12.47000027
4/18/2016 16:06	4/18/2016 18:06	2 (threat not detected)	21.23091316	12.55000019
4/18/2016 16:07	4/18/2016 18:07			
4/18/2016 16:08	4/18/2016 18:08	2 (threat not detected)	21.14932251	ERROR (was not able to read 7 bytes, got: ")
4/18/2016 16:09	4/18/2016 18:09	2 (threat not detected)	21.11527252	12.63000011
4/18/2016 16:10	4/18/2016 18:10	2 (threat not detected)	21.07777786	12.55000019
4/18/2016 16:11	4/18/2016 18:11	2 (threat not detected)	21.04366684	12.69999981
4/18/2016 16:12	4/18/2016 18:12		21.0129528	ERROR (was not able to read 7 bytes, got: ")

Abbildung 7: Beispiel fehlerhafte OpenWIS-Messwerte.

2.1.2 Zeitreihendatenbank

Da es sich bei den Daten um Messreihen handelt, entschieden wir uns für die Speicherung der Daten eine Zeitreihen-Datenbank zu verwenden. Zeitreihen-Datenbank Systeme sind Datenbank-Systeme, die speziell zum Zweck von Zeitreihenanalysen entwickelt und optimiert wurden. In diesen Systemen werden Daten als Tupel von Zeitpunkt und Messwert abgespeichert, um darauf anschließend mathematische Operationen ausführen zu können.

Im Rahmen des Projekts wurden die zurzeit gängigsten Zeitreihen-Datenbanken evaluiert unter anderem die DalmatinerDB⁸, die InfluxDB⁹ und Elasticsearch¹⁰. Wir entschieden uns dafür, die vorverarbeiteten Daten beider Systeme in einer InfluxDB abzulegen. Die InfluxDB ist eine Zeitreihen-Datenbank, die sich im Zuge dieser Evaluierung als sehr geeignet für den vorhandenen Anwendungsfall gezeigt hat. Ausschlaggebende Gründe dafür waren unter anderem (i) die gute Performance, sowohl beim Import als auch bei Abfragen, sowie (ii) die gute Kompressionsrate bei der Speicherung von Sensordaten. Als weiterer Vorteil der InfluxDB stellte sich die Abfragesprache *InfluxQL* heraus, die sehr nah an SQL angelehnt ist, was die Einarbeitungszeit für die Arbeit mit der InfluxDB reduzierte. Darüber hinaus hat die InfluxDB mit Chronograf¹¹ eine einfache GUI-Komponente mit deren Hilfe eine schnelle Analyse der Input-Daten ermöglicht wird.

⁸ <https://dalmatiner.io/>

⁹ <https://docs.influxdata.com/influxdb/v1.2/>

¹⁰ <https://www.elastic.co/de/products/elasticsearch>

¹¹ <https://docs.influxdata.com/chronograf/v1.2/>

Die OpenWIS und UDE Daten werden somit in der InfluxDB als Zeitreihen gehalten, wobei eine Zeitreihe aus einer beliebigen Anzahl an Datenpunkten bestehen kann. Ein Datenpunkt wiederum besteht aus einem Zeitstempel, Messwerten und Metadaten. Eine Zeitreihe entspricht dabei einer Art SQL Tabelle und ein Datenpunkt entspricht einer Zeile in dieser Tabelle. Der Vorteil gegenüber herkömmlichen SQL-Datenbanken ist hierbei, dass die InfluxDB schemalos agiert, das heißt eine beliebige Anzahl an Spalten (Messwerten) kann ohne „a priori“ Wissen über die Struktur der Daten zu besitzen, angelegt werden. Sobald eine unbekannte Spalte importiert wird, wird sie automatisch angelegt. Dieser Fall kann beispielsweise eintreten, falls einer Station ein neuer Sensor hinzugefügt wird.

Für den wissenschaftlichen Prototyp wurde ein Ansatz gewählt, der OpenWIS- und UDE-Daten in zwei getrennten Datenbanken innerhalb der InfluxDB abspeichert, um eine bessere Performance und mehr Flexibilität hinsichtlich eventueller künftiger Entwicklungen zu erlangen.

Im Folgenden wird die Struktur der OpenWIS als auch der UDE Daten beschrieben. Für jede Sensorstation (z.B.: „10.171.161.73“ bzw. „A10_1 km|11,26“) wird eine eigene Zeitreihe angelegt, in die dann die Daten für diese Sensorstation eingetragen werden. Dabei beinhaltet ein Datenpunkt die Messwerte aller Sensoren, die zu diesem Zeitpunkt an dieser Sensorstation gemessen wurde. Abbildung 8 zeigt einen beispielhaften Datenauszug aus der InfluxDB im CSV-Format¹².

```
name,tags,time,Alarm_1(0),Alarm_2(0),Alarm_3(0),BT(1),BT(2),BT(4),BT_defect(0),FB_Feucht(0),Frost_Detected(0),GT(0),GT(2),GT_F(0),Sensors_defect(0),TP(0),WFD(0)
10.171.161.73,,1427324400000000000,false,false,false,8.770000457763672,8.43999995803833,9.3999999618530273,,,,,,,,,,,,,
10.171.161.73,,1427324460000000000,false,false,false,8.760000228881836,8.430000305175781,9.3999999618530273,,,,,,,,,,,,,
10.171.161.73,,1427324520000000000,false,false,false,8.760000228881836,8.430000305175781,9.3999999618530273,,,,,,,,,,,,,
10.171.161.73,,1427324580000000000,false,false,false,8.75,8.420000076293945,9.3999999618530273,,,,,,,,,,,,,
10.171.161.73,,1427324640000000000,false,false,false,8.739999771118164,8.40999984741211,9.3999999618530273,,,,,,,,,,,,,
10.171.161.73,,1427324700000000000,false,false,false,8.729999542236328,8.3999999618530273,9.3999999618530273,,,,,,,,,,,,,
10.171.161.73,,1427324760000000000,false,false,false,8.729999542236328,8.3999999618530273,9.3999999618530273,,,,,,,,,,,,,
10.171.161.73,,1427324820000000000,false,false,false,8.720000267028809,8.390000343322754,9.3999999618530273,,,,,,,,,,,,,
10.171.161.73,,1427324880000000000,false,false,false,8.710000038146973,8.380000114440918,9.3999999618530273,,,,,,,,,,,,,
10.171.161.73,,1427324940000000000,false,false,false,8.699999809265137,8.380000114440918,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325000000000000,false,false,false,8.699999809265137,8.36999985559082,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325060000000000,false,false,false,8.6899995803833,8.359999656677246,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325120000000000,false,false,false,8.680000305175781,8.350000381469727,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325180000000000,false,false,false,8.670000076293945,8.350000381469727,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325240000000000,false,false,false,8.670000076293945,8.34000015258789,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325300000000000,false,false,false,8.65999984741211,8.34000015258789,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325360000000000,false,false,false,8.649999618530273,8.34000015258789,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325420000000000,false,false,false,8.649999618530273,8.34000015258789,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325480000000000,false,false,false,8.640000343322754,8.329999923706055,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325540000000000,false,false,false,8.640000343322754,8.329999923706055,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325600000000000,false,false,false,8.630000114440918,8.319999694824219,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325660000000000,false,false,false,8.630000114440918,8.319999694824219,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325720000000000,false,false,false,8.619999885559082,8.3100004196167,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325780000000000,false,false,false,8.619999885559082,8.3100004196167,9.300000190734863,,,,,,,,,,,,,
10.171.161.73,,1427325840000000000,false,false,false,8.609999656677246,8.3100004196167,9.300000190734863,,,,,,,,,,,,,
```

Abbildung 8: Beispielhafter Datenauszug aus der InfluxDB als CSV.

¹² Comma-separated values

Zur Veranschaulichung eine schematische Darstellung der Datenstruktur innerhalb der InfluxDB (Abbildung 9). Beide Datenbanken (OpenWIS und UDE) beinhalten eine beliebige Anzahl an Zeitreihen. Jede dieser Zeitreihen besteht aus einer zeitlich sortierten Anzahl an Datenpunkten, die eine beliebige Anzahl an Messwerten beinhalten können.

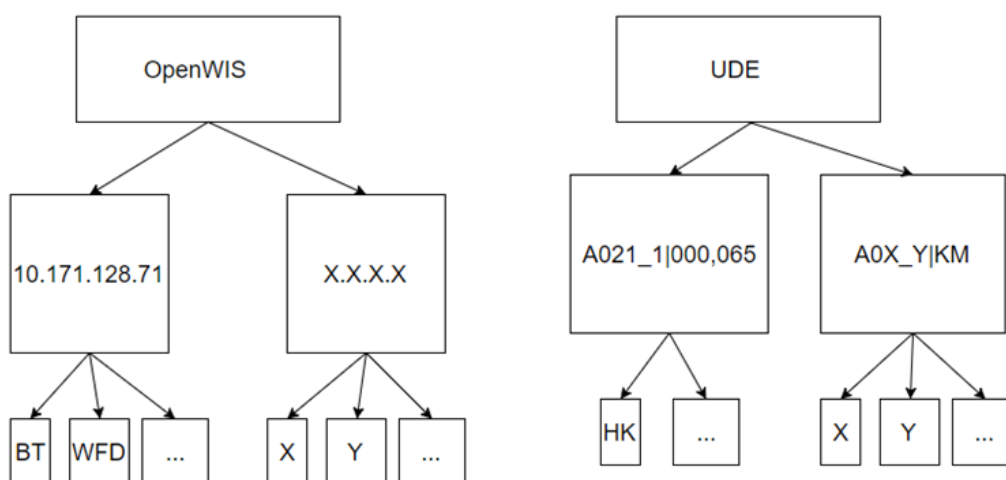


Abbildung 9: Schematischer Aufbau der UDEQI InfluxDB - OpenWIS und UDE Daten werden getrennt abgespeichert.

2.2 Algorithmen und Verfahren

Abbildung 10 skizziert ein drahtloses Sensornetzwerk und deutet die vielfältigen Ursachen an, die in dem Netzwerk zu Fehlern führen können. Jede Komponente in diesem Netzwerk kann ausfallen und dieser Ausfall kann zu unterschiedlichsten Fehlern führen - beispielsweise zu fehlerhaften Sensormessungen oder zu überhaupt keinen Messungen.

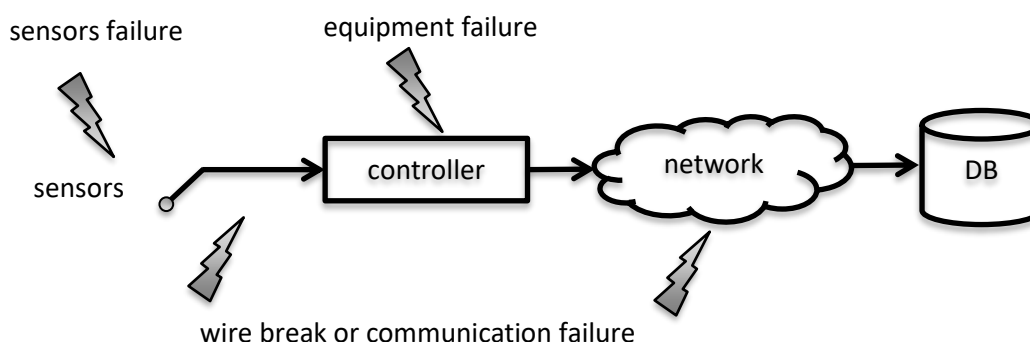


Abbildung 10: Potentielle Fehler, die innerhalb eines drahtlosen Sensornetzwerkes auftreten können.

Gerade in Hinblick auf die Thematik Verfügbarkeit und Plausibilität ist es wichtig, sich dieser Zusammenhänge bewusst zu werden. Fehlerhaftes Equipment kann sich beispielsweise

sowohl auf die Verfügbarkeit (Sensor defekt bzw. Leitungsunterbrechung) als auch auf die Plausibilität (Driften der Werte bei einem älteren Sensor, Herstellungsfehler bzw. wurde ein Sensor möglicherweise unzureichend geeicht) auswirken. Individuelle Fehler können sich demzufolge unterschiedlich in den gespeicherten Werten in der Datenbank manifestieren. Dieses Forschungsvorhaben hat es sich nun zum Ziel gesetzt, mit hoher Güte Aussagen zu erhalten

- (i) über die **Verfügbarkeit eines Sensors**, d.h. in welchem Ausmaß stehen Messwerte eines Sensors überhaupt zur Verfügung (vgl. Abschnitt 2.2.1), sowie
- (ii) über die **Plausibilität eines Sensors**, d.h. wie gut sind die Messwerte und deren Qualität tatsächlich? (vgl. Abschnitt 2.2.2)

Um diese Aussagen treffen zu können, werden in den nächsten beiden Abschnitten Algorithmen und Verfahren aus dem Bereich der „Knowledge Discovery in Databases (KDD)“ angewandt und evaluiert. Das Ergebnis von Verfügbarkeits- sowie Plausibilitätsberechnungen wird zwischengespeichert und dient dem Statusupdate von OpenWIS- sowie UDE-Stationen, die über die graphische Benutzeroberfläche angezeigt werden.

2.2.1 Verfügbarkeit

In einem ersten Schritt wurden umfassende Statistiken für beide Sensordatentypen (OpenWIS, UDE) erstellt. OpenWIS und UDE Daten unterscheiden sich voneinander und zwar einerseits in der Art der Speicherung:

- Die OpenWIS Dateien liegen in einem Ordnersystem vor, wobei sich in den Monatsordnern die Sensor-Dateien tageweise gruppiert befinden. Jede Zeile in einer solchen Sensor-Datei besteht dabei aus einem Zeitstempel und einem Messwert pro Spalte. In den einzelnen Zellen können dabei Messwerte, Übertragungsfehler oder leere Felder beinhalten.
- Die Daten des UDE-Systems stehen über eine SQL-View zur Verfügung und können mittels einer entsprechenden SQL-Abfrage ausgelesen werden.

und andererseits in den existierenden Feldern eines Sensors. Letzteres wirkt sich auf die unterschiedlichen Statistiken aus, die in folgender Tabelle zusammengefasst sind:

	#Stationen	Zeitraum	Generelle Statistiken	Detaillierte Statistiken
OpenWIS	220	2012 – 2017	<ul style="list-style-type: none"> - #Files per Year (= fehlender Files) + #leerer Files pro Jahr (= #Files ohne Inhalt) - Zeitperioden mit / ohne Messungen pro Jahr 	<ul style="list-style-type: none"> - Pro Station + pro Sensortyp + pro Jahr - Duplikate, ungültige Werte - Min, Max, Mittelwert, Standardabweichung, Quantile
UDE	783	2016 – 2017	<ul style="list-style-type: none"> - Vorkommen - Zeitperioden mit / ohne Messungen pro Jahr - Fehlende Werte 	<ul style="list-style-type: none"> - Pro Station + pro Sensortyp + pro Jahr - Min, Max, Mean, Standardabweichung, Quantile

Tabelle 1: Statistiken für beide Sensordatentypen (OpenWIS, UDE)

Statistische Auswertungen für OpenWIS Stationen

Abbildung 11 zeigt detaillierte Statistiken für die OpenWIS Station „10.171.34.71“ für das Jahr 2016. Daraus ist zu ersehen, dass diese Station eine ganze Reihe von Sensoren mit unterschiedlichen Einheiten umfasst wie beispielsweise Temperaturwerte für die Gefriertemperatur (Sensor ARS31) bzw. die Fahrbahnoberflächentemperatur (Sensor IRS31) oder boolesche Werte, ob Reif detektiert wurde oder nicht. Für jeden Sensor dieser Station wird detailliert aufgelistet, wie viele Werte für diesen Zeitraum gemessen werden sollten, wie viele Messungen fehlen und wie viele von den vorhandenen Messwerten ungültig sind. Des Weiteren werden für jeden Sensor der Station der minimale Wert, der maximale Wert, der Mittelwert sowie die Streuung um diesen Mittelwert für den angegebenen Zeitraum angegeben.

Sensor name	Data Type	Appearance in files %	Expected amount	Missing values	Invalid Values	Minimum	Maximum	Mean	Streuung
Timestamp (UTC)	Timestamp	100	198489	3111	139	01.01.2016 00:00	31.12.2016 23:59		
Timestamp (Europe/Vienna)	Timestamp	100	198489	3111	139	01.01.2016 01:00	01.01.2017 00:59		
IRS31 TT1 (Degrees_Celsius)/BT(2)	Float	100	198489	3926	261	-9,100000381	28,70000076	5,903496749	5,962239922
IRS31 FBT (Degrees_Celsius)/BT(6)	Float	100	198489	5021	836	-10,89999962	30,89999962	5,766051488	6,276944875
WSx LT (Degrees_Celsius)/LT(0)	Float	100	198489	22302	1384	-14,3805666	25,88579369	3,786902034	5,485400746
LPM_M1_SYNOP_4680 (None)/NS_SYNOP(0)	String	100	198489	6545					
WSx RF (Percent)/RF(0)	Float	100	198489	20087	30029	-3,550000191	100	81,19178924	13,84090803
IRS31 WFD (Micrometers)/WFD(0)	Float	100	198489	6766	25770	-16,25777435	734	3,021142202	11,32940043
A1 (None)/Alarm_1(0)	Boolean	100	198489	18142	139024				
A2 (None)/Alarm_2(0)	Boolean	100	198489	5962	44496				
A3 (None)/Alarm_3(0)	Boolean	100	198489	18148	32310				
Defekte BT (None)/BT_defect(0)	Boolean	100	198489	5970	44493				
FB Feucht (None)/FB_Feucht(0)	Boolean	100	198489	18155	32310				
Reif detektiert (None)/Frost_Detected(0)	Boolean	100	198489	5985	44478				
Defekte GT (None)/GT_defect(0)	Boolean	100	198489	18134	32324				
NS Regen (None)/NS_Regen(0)	Boolean	100	198489	3867	46600				
NS Schnee (None)/NS_Schnee(0)	Boolean	100	198489	3859	46602				
Defekter NS (None)/NS_defect(0)	Boolean	100	198489	3859	46599				
Defekte Sensoren (None)/Sensors_defect(0)	Boolean	100	198489	3843	135699				
ARS31 GT (Degrees_Celsius)/GT(2)	Float	99,28571429	197071	5112	105352	-25,93000031	30,10000038	-7,02429096	10,72383499
FBT 1wire (Degrees_Celsius)/BT(6)	Float	72,85714286	144613	3111	0 All values are invalid	All values are invalid	All values are invalid	All values are invalid	All values are invalid
TP WSx (Degrees_Celsius)/TP(0)	Float	72,85714286	144613	3111	0 All values are invalid	All values are invalid	All values are invalid	All values are invalid	All values are invalid
ARS31 Status (n/a)		72,85714286	144613	3111					
FBT 1Wire (Degrees_Celsius)/BT(4)	Float	27,14285714	53875	22943	215	-9,1875	29,0625	7,949895554	5,743928698
ARS31 GT-Status (n/a)		27,14285714	53875	3111					
GT valid (None)/GT_valid(0)	Boolean	27,14285714	53875	5942	44519				
NS leicht (None)/NS_leicht(0)	Boolean	27,14285714	53875	3864	32445				
NS stark (None)/NS_stark(0)	Boolean	27,14285714	53875	3847	135699				
ARS31 GT (None)/GT(2)	Float	11,42857143	22684	3111	0 All values are invalid	All values are invalid	All values are invalid	All values are invalid	All values are invalid
IRS31 TT1 (None)/BT(2)	Float	10	19848	3111	0 All values are invalid	All values are invalid	All values are invalid	All values are invalid	All values are invalid
IRS31 FBT (None)/BT(6)	Float	10	19848	3111	0 All values are invalid	All values are invalid	All values are invalid	All values are invalid	All values are invalid
WSx LT (None)/LT(0)	Float	10	19848	3111	0 All values are invalid	All values are invalid	All values are invalid	All values are invalid	All values are invalid
WSx RF (None)/RF(0)	Float	10	19848	3111	0 All values are invalid	All values are invalid	All values are invalid	All values are invalid	All values are invalid
TP WSx (None)/TP(0)	Float	10	19848	3111	0 All values are invalid	All values are invalid	All values are invalid	All values are invalid	All values are invalid
IRS31 WFD (None)/WFD(0)	Float	10	19848	3111	0 All values are invalid	All values are invalid	All values are invalid	All values are invalid	All values are invalid
ARS31 Status (None)		10	19848	3111					
FBT ARS31pro (None)		10	19848	3111					
IRS31 FBZ (None)		10	19848	3111					
FBT 1wire (None)/BT(6)	Float	1,428571429	2835	3111	0 All values are invalid	All values are invalid	All values are invalid	All values are invalid	All values are invalid

Abbildung 11: Beispiel für die OpenWIS Station „10.171.34.71“ für das Jahr 2016.

Abbildung 12 enthält Auszüge über die Anzahl von vorhandenen Files für jene OpenWIS Stationen, bei denen am meisten Files (linker Bereich) bzw. am wenigsten Files (rechter Bereich) vorhanden sind.

#Files pro Jahr (Beginn)

Station ID	2012	2013	2014	2015	2016	2017	#Files
10.171.22.130	260	217	365	364	233	183	1622.0
10.171.24.249	260	217	365	364	183	183	1572.0
10.171.24.169	260	217	365	364	183	183	1572.0
10.171.26.34	260	217	365	364	183	183	1572.0
10.171.131.72	260	217	365	355	173	183	1553.0
10.171.24.217	260	217	365	302	183	183	1510.0
10.171.25.65	260	217	365	302	183	183	1510.0
10.171.25.67	260	217	365	302	183	183	1510.0
10.171.172.11	260	217	365	302	183	183	1510.0
10.171.30.1	259	217	365	302	183	183	1509.0
10.171.68.142	258	217	365	302	183	183	1508.0
10.171.68.73	258	217	365	302	183	183	1508.0
10.171.30.25	258	217	365	302	173	183	1498.0
10.171.24.97	260	217	365	302	183	170	1497.0
10.171.23.217	260	217	365	302	180	170	1494.0
10.171.26.129	260	217	365	302	164	170	1478.0
10.171.68.71	258	217	365	302	152	183	1477.0
10.171.24.33	260	217	365	302	152	170	1466.0
10.171.24.145	211	217	365	302	183	183	1461.0
10.171.68.101	258	217	365	302	183	135	1460.0
10.171.17.163	260	217	365	302	135	170	1449.0
10.171.30.201	258	217	365	302	183	84	1409.0
10.171.34.141	88	217	365	364	183	183	1400.0
10.171.24.73	260	217	365	302	140	104	1388.0
10.171.23.209	260	217	365	302	151	91	1386.0
10.171.33.71	70	217	365	364	183	183	1382.0
10.171.164.71	21	217	365	364	183	183	1333.0
10.171.16.34	70	217	365	363	183	112	1310.0
10.171.33.72	70	217	365	364	183	107	1306.0
10.171.39.145	21	217	365	364	160	170	1297.0
10.171.39.72	21	217	365	364	160	170	1297.0
10.171.39.172	21	217	365	364	160	170	1297.0
10.171.39.71	21	217	365	364	160	170	1297.0
10.171.39.171	21	217	365	364	160	170	1297.0
10.171.66.71	21	217	365	364	151	170	1288.0

#Files pro Jahr (Ende)

Station ID	2012	2013	2014	2015	2016	2017	#Files
10.171.51.101	0	0	0	0	0	71	71.0
10.171.224.104	0	0	0	0	0	71	71.0
10.171.24.81	0	0	0	0	0	71	71.0
10.171.130.143	0	0	0	0	0	71	71.0
10.171.224.74	0	0	0	0	0	71	71.0
10.171.227.104	0	0	0	0	0	71	71.0
10.171.227.102	0	0	0	0	68	0	68.0
10.171.228.101	0	0	0	0	62	0	62.0
10.171.227.73	0	0	0	0	0	58	58.0
10.171.228.103	0	0	0	0	0	55	55.0
10.171.56.72	0	0	0	0	0	53	53.0
10.171.188.105	0	0	0	0	0	21	21.0
10.171.186.71	0	0	0	0	0	21	21.0
10.171.208.71	0	0	0	0	0	21	21.0
10.171.56.73	0	0	0	0	0	16	16.0
10.171.160.72	0	0	0	0	0	13	13.0
10.171.130.141	0	0	0	0	0	13	13.0
10.171.55.101	0	0	0	0	0	13	13.0
10.171.34.171	0	0	0	0	0	13	13.0
10.171.55.103	0	0	0	0	0	13	13.0
10.171.55.102	0	0	0	0	0	13	13.0
10.171.130.101	0	0	0	0	0	13	13.0
10.171.174.71	0	0	0	0	0	13	13.0
10.171.135.73	0	0	0	0	0	13	13.0
10.171.130.142	0	0	0	0	0	13	13.0
10.171.135.72	0	0	0	0	0	13	13.0
10.171.174.76	0	0	0	0	0	13	13.0
10.171.160.103	0	0	0	0	0	13	13.0
10.171.150.72	0	0	0	0	0	13	13.0
10.171.150.71	0	0	0	0	0	13	13.0
10.171.130.72	0	0	0	0	0	13	13.0
10.171.174.73	0	0	0	0	0	13	13.0
10.171.135.71	0	0	0	0	0	13	13.0
10.171.174.74	0	0	0	0	0	13	13.0
10.171.160.74	0	0	0	0	0	13	13.0

Abbildung 12: Anzahl an Files pro Jahr bei OpenWIS Stationen für den Zeitraum 2012 - 2017.

Abbildung 13 enthält Auszüge über die Anzahl von Files für jene OpenWIS Stationen, bei denen die meisten Files zwar vorhanden jedoch leer (linker Bereich) sind bzw. bei denen die wenigsten Files leer (rechter Bereich) sind.

#Leerer Files pro Jahr (Beginn)

Station ID	2012	2013	2014	2015	2016	2017	#EmptyFiles
10.171.39.171	21	217	365	155	0	63	821.0
10.171.24.217	185	55	307	182	53	34	816.0
10.171.39.172	21	217	365	148	0	63	814.0
10.171.27.2	21	217	324	177	11	35	785.0
10.171.23.241	219	61	217	232	52	0	781.0
10.171.24.145	34	217	344	58	53	34	740.0
10.171.39.145	0	159	365	152	0	57	733.0
10.171.24.33	103	143	129	246	53	56	730.0
10.171.23.201	143	65	236	205	55	0	704.0
10.171.24.73	143	119	216	197	6	2	683.0
10.171.27.65	21	217	318	121	0	5	682.0
10.171.26.193	21	217	365	14	2	63	682.0
10.171.23.145	21	217	365	14	1	63	681.0
10.171.26.225	21	217	365	14	2	55	674.0
10.171.27.1	21	217	365	14	0	5	622.0
10.171.215.72	21	217	365	14	0	5	622.0
10.171.23.193	21	92	125	246	52	83	619.0
10.171.17.163	10	142	321	1	0	145	619.0
10.171.30.217	21	217	365	14	0	0	617.0
10.171.17.98	21	217	365	14	0	0	617.0
10.171.30.145	21	217	365	14	0	0	617.0
10.171.37.71	21	154	253	109	4	63	604.0
10.171.20.193	21	217	300	0	1	49	588.0
10.171.25.66	21	217	300	0	1	34	573.0
10.171.26.97	21	217	300	0	1	34	573.0
10.171.28.33	121	7	124	260	52	0	564.0
10.171.25.129	21	217	300	4	1	5	548.0
10.171.30.57	0	4	126	270	53	75	528.0
10.171.165.71	125	29	312	14	0	0	480.0
10.171.30.1	51	65	194	132	1	34	477.0
10.171.54.71	21	154	124	93	11	63	466.0
10.171.23.217	7	4	234	93	60	38	436.0
10.171.245.72	0	21	275	115	0	23	434.0
10.171.66.174	13	67	156	118	3	64	421.0
10.171.23.209	7	30	220	144	0	19	420.0

#Leerer Files pro Jahr (Ende)

Station ID	2012	2013	2014	2015	2016	2017	#EmptyFiles
10.171.144.11	0	0	0	0	1	20	21.0
10.171.235.72	0	0	0	0	1	20	21.0
10.171.234.101	0	0	0	0	1	19	20.0
10.171.235.71	0	0	0	0	1	11	12.0
10.171.67.173	0	0	0	0	7	1	8.0
10.171.5.71	0	0	0	0	2	5	7.0
10.171.230.141	0	0	0	0	1	5	6.0
10.171.188.105	0	0	0	0	0	5	5.0
10.171.186.71	0	0	0	0	0	5	5.0
10.171.208.71	0	0	0	0	0	5	5.0
10.171.227.102	0	0	0	0	1	0	1.0
10.171.234.72	0	0	0	0	1	0	1.0
10.171.228.101	0	0	0	0	1	0	1.0
10.171.160.72	0	0	0	0	0	0	0.0
10.171.130.141	0	0	0	0	0	0	0.0
10.171.55.101	0	0	0	0	0	0	0.0
10.171.34.171	0	0	0	0	0	0	0.0
10.171.55.103	0	0	0	0	0	0	0.0
10.171.55.102	0	0	0	0	0	0	0.0
10.171.130.101	0	0	0	0	0	0	0.0
10.171.174.71	0	0	0	0	0	0	0.0
10.171.135.73	0	0	0	0	0	0	0.0
10.171.130.142	0	0	0	0	0	0	0.0
10.171.135.72	0	0	0	0	0	0	0.0
10.171.56.72	0	0	0	0	0	0	0.0
10.171.56.73	0	0	0	0	0	0	0.0
10.171.174.76	0	0	0	0	0	0	0.0
10.171.160.103	0	0	0	0	0	0	0.0
10.171.150.72	0	0	0	0	0	0	0.0
10.171.150.71	0	0	0	0	0	0	0.0
10.171.130.72	0	0	0	0	0	0	0.0
10.171.174.73	0	0	0	0	0	0	0.0
10.171.135.71	0	0	0	0	0	0	0.0
10.171.174.74	0	0	0	0	0	0	0.0
10.171.160.74	0	0	0	0	0	0	0.0

Abbildung 13: Anzahl an leeren Files pro Jahr bei OpenWIS Stationen für den Zeitraum 2012 - 2017.

Abbildung 12 und Abbildung 13 geben eine erste Einschätzung bzgl. der Verfügbarkeit und liefern erste Kandidaten für weiterführende Analysen. Stationen, die „gut“ hinsichtlich ihrer Verfügbarkeit sind, eignen sich zur Generierung für Trainingsdaten und zur Evaluierung von Plausibilitätsalgorithmen. Die Anzahl der Stationen, die „schlecht“ abschneiden, schaffen ein Bewusstsein für die Problematiken im Sensornetzwerk der ASFiNAG. Eine diesbezügliche, erste Ursachenforschung ergab, dass viele dieser Stationen während der Wintermonate bzw. während Phasen von Baustellen abgeschaltet wurden. Auf Basis dieser Beobachtungen ist der graphischen Benutzeroberfläche eine Passivierungskomponente hinzugefügt worden, die es nun ASFiNAG Mitarbeitern erlaubt, einzelne Stationen von den Analysen auszunehmen.

Statistische Auswertungen für UDE Stationen

Abbildung 14 zeigt detaillierte Statistiken für die UDE Station „A02_1|161,282“. Auch bei dieser UDE Station sind mehrere Sensoren verbaut wie beispielsweise Sensoren zur Messung der Fahrbahntemperatur, der Gefriertemperatur oder der Wasserfilmdicke. Für jeden Sensor dieser Station wird detailliert aufgelistet, wie viele Messwerte vorhanden sind und wie viele

Werte fehlen. Des Weiteren werden wiederum für jeden Sensor der Station der minimale Wert, der maximale Wert, der Mittelwert sowie die Standardabweichung um diesen Mittelwert angegeben.

Sensor Type	Data Type	#Entries	#Missing Values	Minimum	Maximum	Mean	Standard deviation
Time Period	Timestamp	323004	0	2016-10-04T12:26:00Z	2017-05-21T23:59:00Z		
KM_POS	metadata						
ROUTE	metadata						
ROW_NR	metadata						
Fahrbahntemperatur	Float	323004	69	-13,30000019	42,90000153	6,926695672	9,219235474
Zustand der Fahrbahnoberfläche	Float	323004	66	0	66	3,145517547	13,17488041
Gefriertemperatur	Float	323004	66	-22	0	-0,08256688	1,003024348
Restsalz	Float	323004	66	0	100	0,442163642	4,760419139
Taustoffkonzentration	Float	323004	1	0	22	0,092346177	1,028072506
Temperatur in Tiefe 1	Float	323004	289529	0	0	0	0
Wasserfilmdicke	Float	323004	64	0	4,199999809	0,140960671	0,737692615

Abbildung 14: Beispiel für die UDE Station "A02_1|161,282".

Die folgende Tabelle 2 zeigt die Häufigkeit der einzelnen Sensortypen über alle UDE Stationen an; so sind beispielsweise der Helligkeitssensor (HK) mit ~77%, die Fahrbahntemperatur mit 18,9% und der Fahrbahnzustand mit 18,1% am häufigsten und die Globalstrahlung mit 0,1%, Schadstoffe / PM10 mit 0,5% und Stickoxid mit 0,8% am seltensten in UDE Stationen des ASFiNAG Sensornetzwerks verbaut.

Sensortyp	Häufigkeit [%]	Sensortyp	Häufigkeit [%]
Helligkeit	77,3	Taupunkttemperatur	10,3
Fahrbahntemperatur	18,9	Temperatur in Tiefe 1	9,2
Fahrbahnzustand	18,1	Temperatur in Tiefe 3	7,8
Relative Luftfeuchte	17,4	Windgeschwindigkeit (Mittel)	6,0
Lufttemperatur	16,9	Windrichtung	6,0
Sichtweite	16,6	Windgeschwindigkeit (Spitze)	4,7
Wasserfilmdicke	15,2	Stickstoffmonoxid	1,3
Gefriertemperatur	13,7	Stickstoffdioxid	1,3
Niederschlagsintensität	11,9	Luftdruck	1,3
Niederschlagsart	11,9	Stickoxid	0,8
Taustoffkonzentration	11,6	Schadstoffe / PM10	0,5
Restsalz	11,0	Globalstrahlung	0,1

Tabelle 2: Häufigkeit der einzelnen Sensortypen über alle UDE-Stationen

Abbildung 15 gibt einen Überblick über fehlende Messwerte (in %), d.h. bei welchen UDE Stationen die meisten Werte (rechter Bereich) fehlen und bei welchen die wenigsten.

Fehlende Werte in % (Beginn)

ID	Missing values %
A01_1_168_153	0
A12_1_008_497	0
A12_1_072_193	0
A231_1_001_031	0
A07_2_008_030	0
A09_2_172_188	0
S01_2_015_337	0
A12_2_039_557	0
A13_1_005_314	0
A13_2_008_850	0
A02_Ra4_12_003_873	0
A23_2_005_798	0
S01_1_012_835	0
A07_2_008_627	0
A10_1_005_700	0
A07_1_004_514	0
A10_2_034_500	0
A23_2_000_099	0
A22_2_001_200	0
A021_2_002_370	0
A02_2_172_275	0
A02_1_329_920	0
A02_2_004_128	0
A07_Ra6_4_000_169	0
A23_2_008_275	0
A09_1_194_053	0
A02_2_009_683	0
A10_2_025_640	0
A02_2_006_525	0
A02_Ra4_4_000_480	0
A12_1_002_975	0
A02_1_153_799	0
A02_2_190_814	0
S01_2_013_135	0
A01_2_170_205	0

Fehlende Werte in % (Ende)

ID	Missing values %
A10_2_007_390	18
A02_1_161_276	19
A04_2_006_463	19
S01_2_010_000	22
A12_2_093_264	23
A02_2_004_099	25
A07_1_007_495	33
A09_2_196_989	36
A04_1_011_602	42
A04_2_002_750	42
A04_1_008_700	42
A01_1_161_075	43
A04_2_004_197	43
A23_2_010_692	45
A12_1_144_747	45
A23_2_001_198	46
A01_2_157_712	46
A07_1_001_935	46
A01_2_175_001	46
A07_1_016_338	46
A07_1_014_682	46
A01_2_173_250	46
A01_1_165_891	46
A12_1_023_479	46
A10_1_024_805	46
A01_1_169_074	49
A10_2_014_330	49
A02_1_316_130	49
A10_2_002_340	53
A02_2_320_560	54
A10_1_014_330	60
A23_2_007_089	62
A23_2_004_897	62
A23_1_001_148	62
A23_2_013_955	62

Abbildung 15: Fehlende Werte pro UDE Station in %.

Wie bei den OpenWIS Stationen eignen sich jene UDE Stationen, die keine fehlenden Messwerte aufweisen, besser für weiterführende Plausibilitätsuntersuchungen. Die durchgeführten, statistischen Auswertungen lassen die Aussage zu, dass UDE Daten von höherer Verfügbarkeit zu sein scheinen als die OpenWIS Daten.

Anwendung der Verfügbarkeitsalgorithmen

Die Algorithmen für die Berechnung der Verfügbarkeit basieren auf den soeben vorgestellten, statistischen Auswertungen und daraus abgeleiteten Heuristiken; sie berücksichtigen neben historischen Aufzeichnungen zudem noch Sensorspezifikationen und Domänenwissen.

Diese Algorithmen zur Bestimmung der Verfügbarkeit werden bereits zu Beginn des Datenflusses implementiert, d.h. beim Verarbeiten und Einlesen der Sensordaten in die

InfluxDB, die ausgewählte Zeitreihendatenbank. Aus technischer Sicht wird diese Aufgabe von Kapacitor¹³, einer Datenverarbeitungs-Engine, übernommen, die ein Teil der InfluxDB Umgebung darstellt. Mit der Kapacitor-Komponente können eigene, benutzerdefinierte Logikfunktionen oder benutzerdefinierte Funktionen an die InfluxDB angeschlossen werden, um beispielsweise

- (i) Warnmeldungen mit dynamischen Schwellenwerten zu verarbeiten,
- (ii) Kennzahlen für Muster abzugleichen oder
- (iii) statistische Anomalien zu berechnen.

Die Heuristiken zur Erkennung von Verfügbarkeit werden über diese vorhandenen Logikstrukturen realisiert und führen zu Warnmeldungen, die weiterführende Aktionen triggern. Je nach Quelle, d.h. OpenWIS oder UDE, wurden unterschiedliche Logikfunktionen implementiert. Ein Grund dafür ist, dass sich OpenWIS und UDE darin unterscheiden, wie oft sie zur Abholung bereitgestellt werden. Werden die Daten dann abgeholt, werden sie vom Kapacitor anhand der Logikfunktionen automatisiert überprüft. Warnmeldungen werden beispielsweise ausgegeben, sobald

- an einem Tag ein bestimmter, einstellbarer Prozentsatz an Daten fehlt
 - pro Station
 - pro Sensor
- in einem nahe-Echtzeitbetrieb ein bestimmter, einstellbarer Prozentsatz an Daten innerhalb der letzten 5 / 10 Minuten fehlt
 - pro Station
 - pro Sensor

und führen zu einer Kennzeichnung der entsprechenden Zeiträume in der Datenbank. Diese Kennzeichnungen hinsichtlich fehlender Verfügbarkeit zeichnen schlussendlich für die Darstellung in der graphischen Benutzeroberfläche verantwortlich; d.h. im Fall von UDEQI werden die Stationen mit entsprechenden Fehlercodes versehen und entsprechend einfärbt.

¹³ <https://www.influxdata.com/time-series-platform/kapacitor/>

2.2.2 Plausibilität

In diesem Abschnitt werden Methodiken und Algorithmen vorgestellt, die die Frage nach der Qualität der gemessenen Sensorwerte beantworten sollen. Der Fokus lag dabei auf Algorithmen zur Detektion von Ausreißern (Outlier-Detection) in drahtlosen Sensornetzwerken. Ein drahtloses Sensornetzwerk besteht typischerweise aus einer großen Anzahl kleiner, kostengünstiger Sensorknoten, die über eine große Region verteilt sind. Aufgrund der begrenzten Ressourcen und Fähigkeiten der Sensoren, aufgrund von Umwelteinflüssen und böswilliger Angriffe sind die gemessenen Daten oft unzuverlässig und ungenau (vgl. [14]). Da Ausreißer einen großen Einfluss auf die Datenqualität haben, müssen sie vor der weiteren Verarbeitung erkannt und identifiziert werden, d.h. es muss eine Outlier-Detection durchgeführt werden.

Während im Allgemeinen die Outlier-Detection einen Vorverarbeitungsschritt darstellt, ist es in diesem Projekt eine der zentralen Forschungsfragen. Zentral dabei ist die Definition eines Ausreißers, der für gewöhnlich als „Abweichung vom Normalverhalten“ definiert wird. Die zu stellenden Frage ist nun: Was fällt unter Normalverhalten und was nicht? Zu diskutierende Beispiele umfassen: Baustellen, Verkehrsstaus, saisonale und auch regionale Wetterphänomene, etc. Ebenso stellt eine flächendeckende Dokumentation und Weitergabe einer solchen Abweichung vom Normalverhalten eine große Herausforderung dar.

Bekannte Kategorien von Ausreißern (vgl. Abbildung 16) in Sensordaten sind unter anderem (1) Punktausreißer, (2) kollektive Ausreißer, sowie (3) kontextuelle Ausreißer.

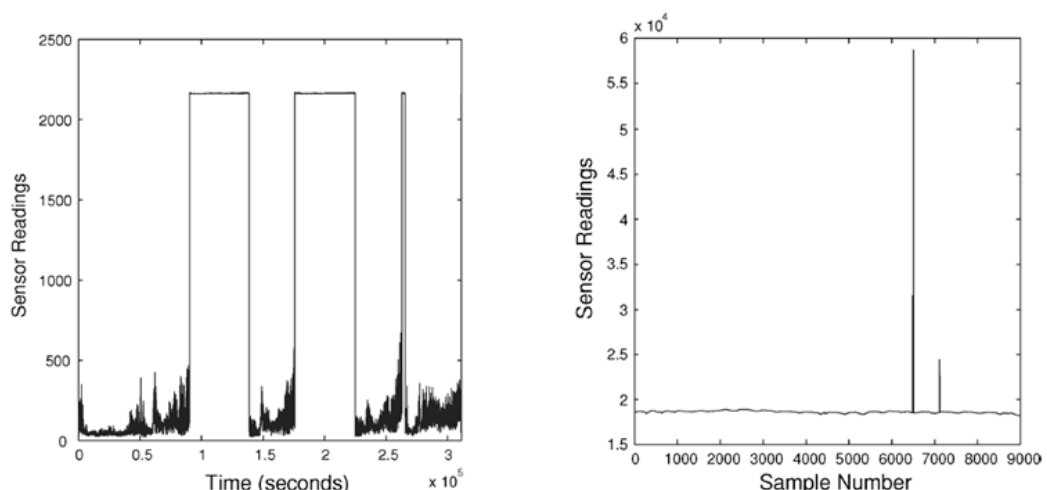


Abbildung 16: Beispiele für kontextueller Ausreißer (links) und Punktausreißer (rechts).

1. Punktausreißer bezeichnen kurze, abrupte Änderungen im Sensorverlauf zwischen aufeinander folgenden Sensorwerten; jeweils ein Messwert unterscheidet sich beispielsweise vom Rest der Zeitserie. (vgl. rechtes Bild in Abbildung 16)
2. Kollektive Ausreißer ist eine Gruppe von Ausreißern, die erst als Kollektiv, d.h. als eine Gruppe von aufeinanderfolgenden Datenpunkten, zu Ausreißern werden. (z.B. mehrere Messwerte folgen nicht einem saisonalen Muster)
3. Kontextuelle Ausreißer treten zum Beispiel auf, wenn ein Sensor einen konstanten Wert für mehrere aufeinanderfolgende Datenpunkte misst. Der gemessene Wert korreliert nicht mit der zugrundeliegenden, physikalischen Verteilung. (vgl. linkes Bild in Abbildung 16) bzw. unterscheiden sich Messwerte von zeitlich direkt vergleichbaren Messwerten.

Zusätzlich können Ausreißer noch in globale bzw. lokale Ausreißer eingeteilt werden, abhängig vom Detailgrad der Betrachtung bzw. dem Datenausschnitt, in dem die Abweichung erfolgt.

Outlier-Detection ist dabei ein Ansatz, um direkt nicht plausible Messwerte zu erkennen. Verfahren, um indirekt nicht plausible Messwerte zu erkennen, beruhen zumeist auf dem Vergleich eines vorhergesagten Messwertes zu dem tatsächlich ausgelesenen Messwert. Als Basis dient dazu ein Modell, das über unterschiedliche Variablen die Vorhersage berechnet. Für die Berechnung des Modells können unterschiedliche Verhältnisse verwendet werden:

- Verhältnis eines Sensors im Vergleich zu seinem historischen Verhalten
- Verhältnis (zum Teil unterschiedlicher) Sensoren am gleichen Standort zum gleichen Zeitpunkt
- Verhältnis Sensoren gleicher Art an unterschiedlichen (benachbarten) Standorten zum gleichen Zeitpunkt

Abbildung 17 zeigt schematisch, wie nicht plausible Messwerte über das Verhältnis von mehreren Sensoren am gleichen Standort erkannt werden können. In einer Clusteranalyse würden im ersten Schritt zwei Gruppen von Zeitserien erkannt werden ($\{1, 2, 3\}$ bzw. $\{4\}$, vgl. Abbildung 17). Anhand zusätzlicher Plausibilitätskriterien (z.B. aus historischen Daten kann die maximale plausible Abweichung abgeleitet werden) wird im zweiten Schritt die zweite Gruppe als nicht plausibel eingestuft.

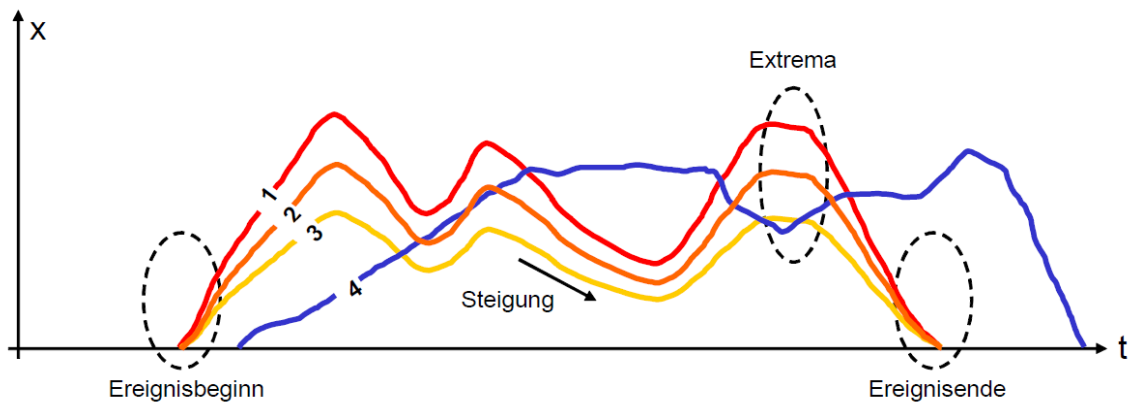


Abbildung 17: Exemplarische Darstellung von kollektiven Ausreißern, die mittels einer Clusteranalyse als nicht plausibel eingestuft werden können.

Sind alle einzelnen Messwerte auf ihre Plausibilität hin kontrolliert, kann eine logisch/physikalische Überprüfung folgen. Da einige Messgrößen voneinander meteorologisch abhängig sind, können nicht plausible Messwerte durch Vergleiche untereinander identifiziert werden. Die Langzeitprüfungen liefern Hinweise auf systematische Fehler. Für eine Langzeit-Plausibilitätsprüfung werden Ergebnisse der jeweils gleichen Sensortypen einer Messstelle mit denen der beiden Nachbarstellen verglichen. Des Weiteren werden Messwerte gleicher Messgrößen aber unterschiedlicher Sensoren korreliert und der Korrelationskoeffizient als dimensionsloses Maß für den Grad des linearen Zusammenhangs berechnet.

Im Kontext des Projektes werden jeweils jene Verfahren ausgesucht, die sich einerseits auf die bestehenden Zeitserien und verfügbaren Daten anpassen lassen sowie jene Verfahren, die auch in einem Online-Kontext eingesetzt werden können. Bei der Literaturrecherche wurden demzufolge existierende State-of-the-Art Algorithmen dahingehend bewertet, ob sie für die Anforderungen des Projekts geeignet und daher sinnvoll einzusetzen sind. Diese Recherche stellt somit die Grundlage für die Entwicklung der Plausibilitätsalgorithmen dar.

Algorithmen sollten unter anderem folgenden Anforderungen entsprechen:

- Verarbeitung multivariater Daten, d.h. Zeitreihen mehrerer Sensoren sollen miteinander verglichen werden können
- Verarbeitung von spatial bzw. temporal korrelierten Daten, da die Messdaten aus verschiedenen OpenWIS- sowie UDE-Stationen sowohl zeitlich als auch räumlich zusammenhängen können
- Unterstützung einer zentralisierten Datenanalyse, da die Messdaten aller Stationen zentralisiert in einer Datenbank abgespeichert werden

- Ausgabe eines nicht-binären Resultats (z.B. ein Wert zwischen 0 und 1), um angeben zu können, wie sehr sich die Messung von dem „normalen“ Muster unterscheidet
- Miteinbeziehung von Daten, bei denen bekannt ist, ob sie sich „normal“ verhalten oder nicht; d.h. Trainingsdaten sind vorhanden und können in überwachten Lernverfahren zum Einsatz kommen
- Robustheit, d.h. der Algorithmus kann mit Ausreißern in den Trainingsdaten umgehen

Abbildung 18 gibt einen Überblick über die gängigsten Methoden zur Outlier-Detection in Sensornetzwerken und stellt den Bezug zu den Anforderungen in diesem Projekt her.

Algorithmus	Multivar. Daten	Attr. korr.	spatial korr.	temporal korr.	zentral. Ansatz	Maß für Ausreißer	annotierte robust Daten
One-class SVM-based [Shahid et al., 2013], [Shahid et al., 2012]	✓	(✓)	✓	✓	✓		
OCPCC-based [Rassam et al., 2012]	✓		✓	✓			
MLPPLS-based [Garces and Sbarbaro, 2011]	✓				✓		✓ ✓
Dynamic Bayesian network-based [Dereszynski and Dietterich, 2011]			✓	✓	✓		
SOM and DWT-based [Siri-panadorn et al., 2010]	✓				✓		
Rule-based [Ramanathan et al., 2006], [Sharma et al., 2010], [Gupchup et al., 2008]				✓	✓		
HMM-based [Sharma et al., 2010]				✓	✓		✓
Robust multivariate statistical-based [Alameddine et al., 2010a]	✓			✓		(✓)	✓
LLSE-based [Sharma et al., 2010]			✓		✓		
ARIMA-based [Sharma et al., 2010]				✓	✓		
Bayesian belief networks-based [Janakiram et al., 2006]	✓	✓	✓	✓	✓		

Abbildung 18: Vergleich von verschiedenen State-of-the-Art Algorithmen zur Outlier-Detection mit Bezug zu den Anforderungen im Projekt. Klammern bedeuten, dass die Anforderungen nur teilweise erfüllt werden.

Jede Zeile repräsentiert einen State-of-the-Art Ansatz und dessen Bezug zu den Anforderungen im Projekt. Der Überblick in Abbildung 18 zeigt, dass kein Ansatz alle Anforderungen erfüllen konnte. Diese Recherche bildete wie bereits angedeutet die Grundlage für die Algorithmenentwicklung; dabei wurden alle in Frage kommenden Ansätze (also jene, die viele Anforderungen erfüllten) ausgetestet und evaluiert. Dabei zeigte sich, dass nur einige wenige über die erste Testphase hinauskamen, teilweise aufgrund von fehlenden

Implementierungsinformationen, teilweise aufgrund zu hoher Komplexität, die zu über die Massen großen Laufzeiten führten. Folgende Tabelle enthält einen Überblick über alle im Projekt entwickelten und implementierten algorithmischen Ansätze, die auf ihre Anwendbarkeit hin untersucht wurden; Tabelle 3 enthält zudem noch Informationen über die jeweiligen Vor- und Nachteile des Ansatzes bzw. die Kategorie des Ausreißers, die der Ansatz adressiert und somit erkennen kann.

Ansatz	Vor- / Nachteile	Ausreißer	Anwendbarkeit
In-Situ Analyse	+ einfach zu implementieren - benötigt mehrere Sensoren an einem Standort	Punktausreißer	ja
Majority-Voting	+ einfach und nachvollziehbar - benötigt Nachbarn	Punktausreißer	ja
Varianzanalyse	+ einfach zu implementieren - nicht zwingend alles Ausreißer	kontextuell (Verläufe konstanter Werte)	ja
Abschätzungs-basierte Methode	+ benötigt keine Nachbarn - benötigt korrelierte Stationen	kontextuell	teilweise
Netzwerkanalyse	+ erlaubt es Stationen miteinander in Beziehung zu setzen	kollektiv	teilweise
ARMA, ARIMA	+ Standardansatz - benötigt spezielle Eigenschaft in den Daten	kontextuell	nein, Sensordaten verfügen nicht über autoregressive Eigenschaft
Analyse der Saisonalität	+ Standardansatz - Datenverluste - nicht alle Sensoren unterliegen Saisonalität	-	nein, Daten sind nicht auf einfache Weise zerlegbar

Tabelle 3: Überblick über alle im Projekt entwickelten und implementierten algorithmischen Ansätze, die auf ihre Anwendbarkeit hin untersucht wurden

Erkannte Ausreißer führen zu einer Kennzeichnung in der Datenbank. Diese Kennzeichnungen hinsichtlich unzureichender Plausibilität zeichnen schlussendlich für die Darstellung in der graphischen Benutzeroberfläche verantwortlich; d.h. im Fall von UDEQI werden die Stationen mit den zugehörigen Fehlercodes versehen und entsprechend eingefärbt.

Im Folgenden werden drei Ansätze detaillierter vorgestellt, die im wissenschaftlichen Prototyp verwendet werden und sich aufgrund ihrer Einfachheit, Robustheit und Qualität bewährt haben.

In-Situ Ansatz

In diesem Ansatz werden lineare Regressionsmodelle dazu verwendet, um Ausreißer bei (zum Teil unterschiedlicher) Sensoren am gleichen Standort zu erkennen. Diese Regressionsmodelle sind dafür bekannt, zugrundeliegende kausale Abhängigkeiten zwischen Attributen erkennen zu können.

Das experimentelle Setup lautet wie folgt: jede Station S ist mit mehreren Sensoren ausgestattet, um verschiedene physikalische Größen wie z.B. Temperatur, Feuchtigkeit, etc. zu messen. Als Beispiel soll die Temperatur eines Sensors über die Temperaturwerte anderer Sensoren in der gleichen Station modelliert werden:

$$T_1 = f(T_2, T_3, \dots, T_k) + \epsilon$$

wobei T_1 die vorherzusagende Temperatur ist und $T_2 \dots T_k$ sind andere Temperatursensoren der gleichen Station

Abbildung 19 illustriert das Verfahren anhand einer konkreten Situation anhand der OpenWIS Station IP 10.171.141.101 (A9 nahe Bosruck) am 23.02.2016.

- T_2 (IRS31FBTDegrees CelsiusBT4) und
- T_3 (wireFBTDegrees CelsiusBT0) wurden benutzt, um
- T_1 (SensitFBTDegrees CelsiusBT8) zu modellieren und somit vorherzusagen.

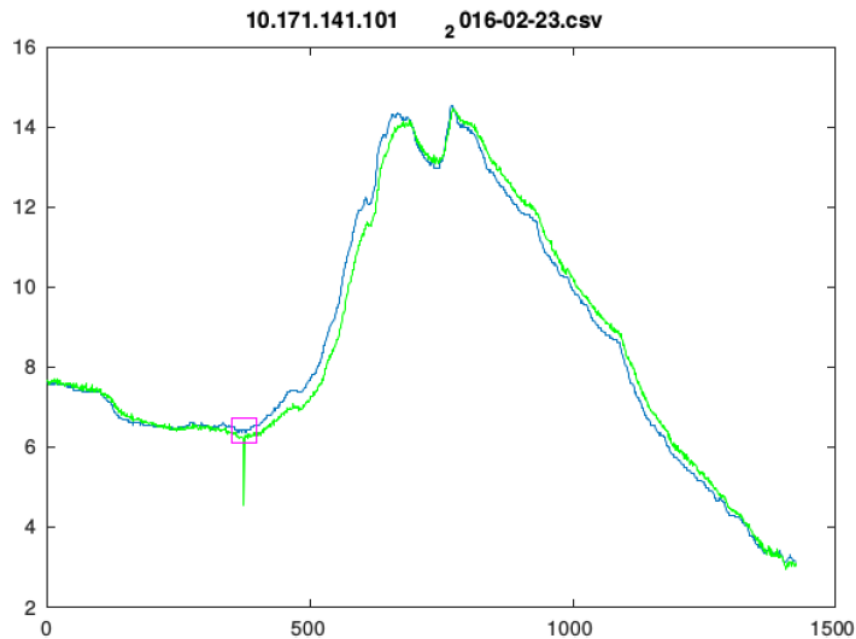


Abbildung 19: Anwendung linearer Regression zur Erkennung von Ausreißern. In grün: der Verlauf der gemessenen Sensorwerte; in blau der Sensorverlauf, wie er von der Regression ($T_2 + T_3$) vorhergesagt wird. Der Bereich in magenta zeigt die Stelle der Abweichung, in der der Ausreißer gefunden wurden.

Majority-Voting Ansatz

In diesem Ansatz werden gezielt benachbarte Stationen zur Erkennung von Ausreißern herangezogen. Ausreißer werden durch einen Mehrheitsbeschluss (75%) benachbarter, gleicher Sensoren erkannt. Im Folgenden einige visuelle Beispiele, wobei benachbarte Sensoren in einem Umkreis von 20km liegen.

Abbildung 20 vermittelt die Idee des Majority-Voting Ansatzes; das linke Bild zeigt den Verlauf der Fahrbahntemperatur (FBT) der UDE Station A23-2|013,955-RN-2, das rechte Bild zeigt denselben Verlauf (in rot) eingebettet in die Temperaturverläufe der benachbarten Stationen, die mehrheitlich ein ähnliches Verhalten aufweisen.

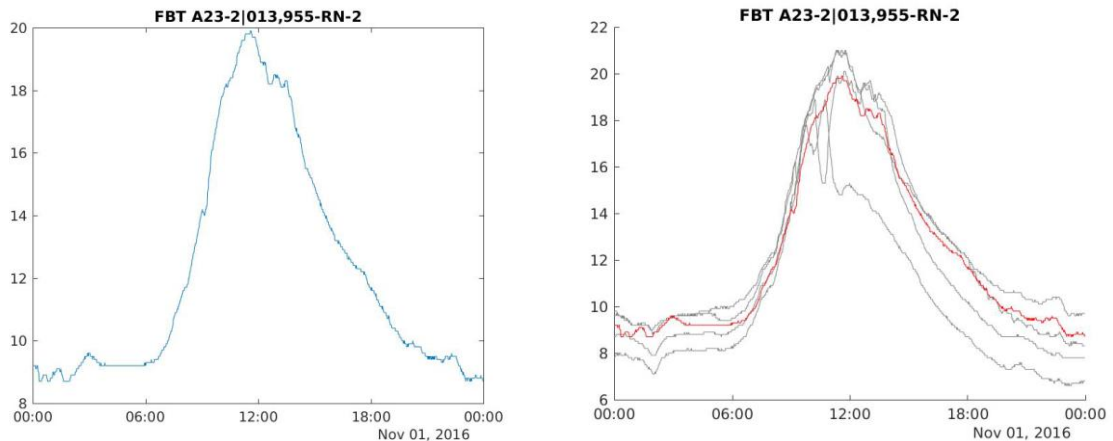


Abbildung 20: Links: Verlauf der Fahrbahntemperatur (FBT). Rechts: Verlauf der FBT (in rot) im Vergleich zu den FBT der benachbarten Stationen.

Abbildung 21 zeigt nun im linken Bild den FBT Verlauf mit in rot gekennzeichneten Ausreißern.

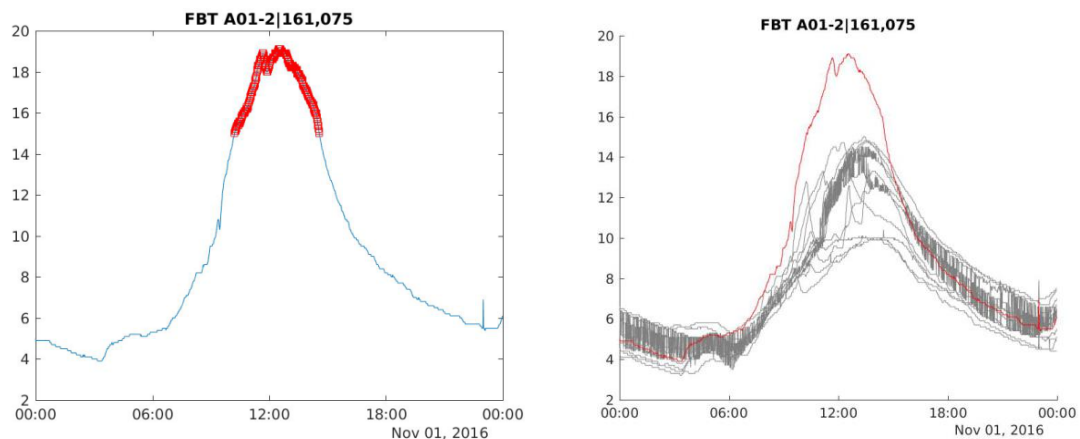


Abbildung 21: Links: Verlauf der Fahrbahntemperatur (FBT) und in rot die gekennzeichneten Ausreißer. Rechts: Verlauf der FBT (in rot) im Vergleich zu den FBT der benachbarten Stationen.

Das rechte Bild zeigt wiederum den Vergleich der FBT Verläufe mit benachbarten Stationen. Die erkannten Ausreißer im linken Bild stimmen mit dem zeitlichen Bereich überein, in dem sich die Verläufe unterscheiden.

Abbildung 22 zeigt ein weiteres Beispiel mit sehr vielen erkannten Ausreißern – in diesem Fall weicht die gemessene Wasserfilmdicke stark von den Verläufen der benachbarten Stationen ab.

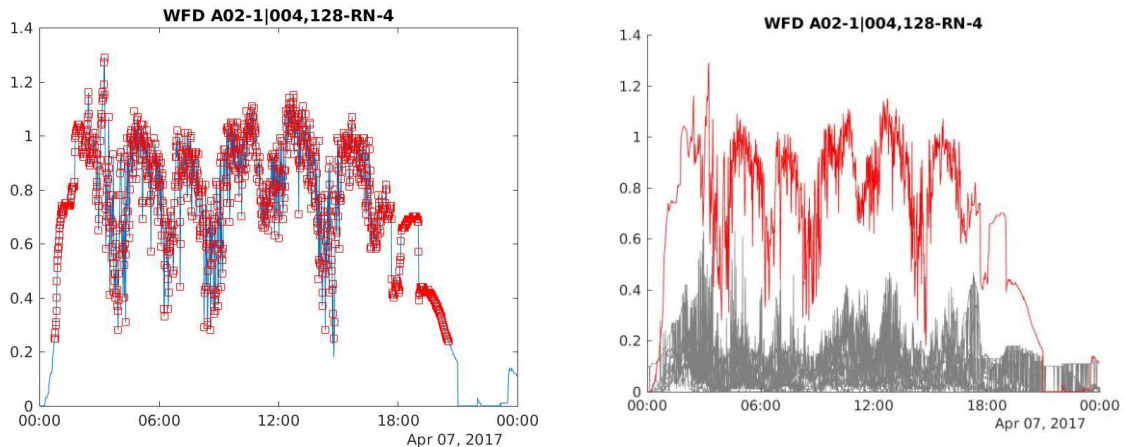


Abbildung 22: Links: Verlauf der Wasserfilmdicke (WFD) und in rot die gekennzeichneten Ausreißer. Rechts: Verlauf der WFD (in rot) im Vergleich zu den WFD der benachbarten Stationen.

Abschätzungs-basierte Methode

Da der Ansatz des Majority-Votings Stationen in der Nachbarschaft voraussetzt, wurden andere Ansätze angedacht darunter ein Ansatz von Sharma et al im Jahr 2010 [15], der auf einer statistischen Schätzung beruht. Diese Schätzung wird auf Basis von Sensorwerten einer hoch korrelierten Station (aus historischen Daten) errechnet; der Vorteil: diese Station muss zwar hoch korrelieren, muss sich jedoch nicht in der Nachbarschaft befinden. So kann durch den Vergleich von mehreren Schätzungen auf abweichendes Verhalten geschlossen werden. Im folgenden Beispiel wird für die Wasserfilmdicke (WFD) der UDE Station A02_1|004,128 jeweils eine Schätzung mit 2 hoch korrelierten Stationen, d.h. A02_1|002,358_RN_4 sowie A23_2|001,198_RN_2, berechnet und die Differenz zu den aktuellen Werten ermittelt (vgl. Abbildung 23).

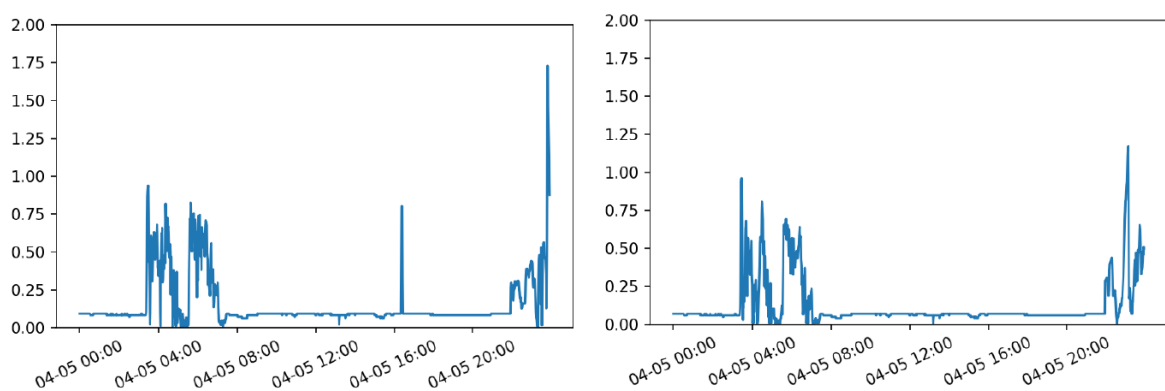


Abbildung 23: Differenz zwischen den Werten von A02_1|004,128 (WFD) und der Schätzung basierend auf A02_1|002,358_RN_4 (links) sowie A23_2|001,198_RN_2 (rechts) am 2017-04-05.

Die Bereiche mit erhöhten Differenzen in beiden Graphiken von Abbildung 23 gleichen einander und deuten auf Ausreißer in diesen Regionen hin.

2.3 Graphische Benutzeroberfläche

In Gesprächen mit ASFiNAG-Bereichsverantwortlichen wurden zu Beginn des Projekts Anwendungsfälle definiert, die die alltäglichen Anforderungen an die MitarbeiterInnen möglichst gut widerspiegeln sollen (vgl. Kapitel 4). Aus diesen Gesprächen ging hervor, dass es drei Ansichten mit unterschiedlichen Funktionalitäten geben sollte; eine „Einfache“, eine „Experten“ und eine „Admin Ansicht“.

Die „Einfache“ Ansicht (Abbildung 24) mit Standardvisualisierungen wie :

- einer Geo-Visualisierung mit einer Übersicht über alle Stationen
- Line-Charts, um den Verlauf von einzelnen Sensorwerten anzuzeigen
- einfache Tabellen, die beispielsweise defekte Sensoren auflistet

Die „Experten“ Ansicht erweitert die „Einfache“ Ansicht um eine Analysefunktionalität.

Die „Admin“ Ansicht ermöglicht zusätzlich zur „Experten“ Ansicht,

- die Veränderung von Stammdaten, der Stationen und einzelner Sensoren
- sie ermöglicht also einen Direktzugriff auf bereits gespeicherte Daten, die verändert werden können, und ist somit sorgfältig zu verwenden

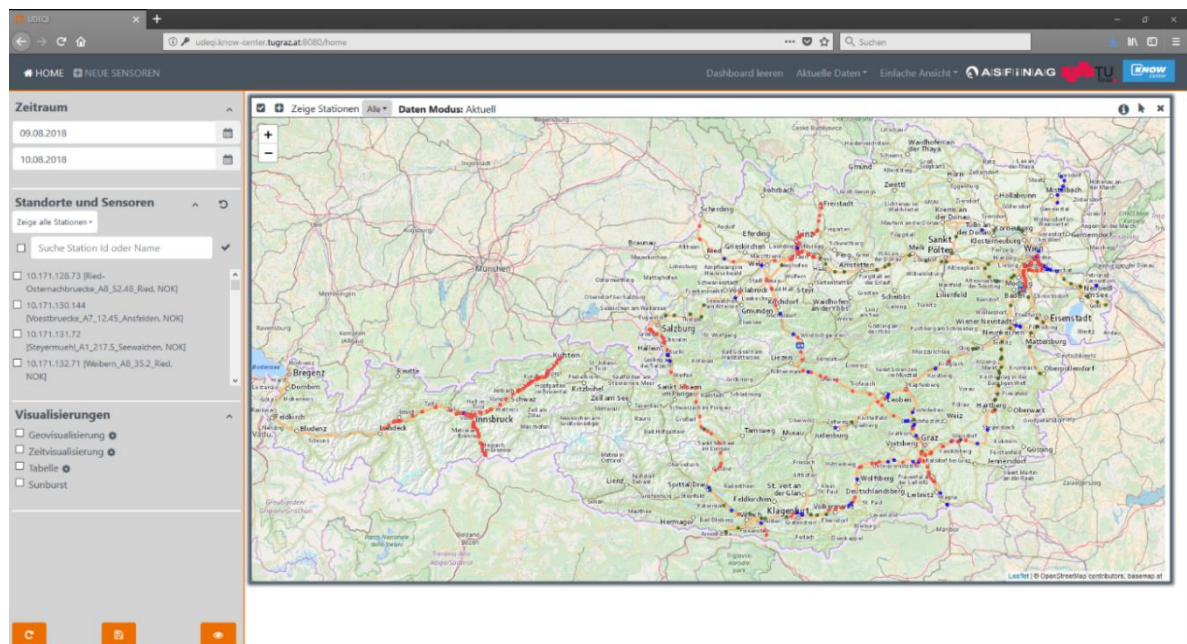


Abbildung 24: „Einfache“ Ansicht von UDEQI mit Anzeige des Zeitraums, der Standorte und Sensoren und der Visualisierungen im linken Bereich und der Geo-Visualisierung aller OpenWIS sowie UDE Stationen in Österreich. Die entsprechende Färbung der Station spiegelt deren Zustand wieder.

Im Folgenden werden die Visualisierungskomponenten der Dashboard-Anzeige vorgestellt sowie die Vorteile der jeweiligen Funktionalitäten dargelegt:

Die **Geo-Visualisierung** in Abbildung 25 gibt einen Überblick über alle gewählten Stationen. Je nach Stationstyp werden die Stationen als Kreise (OpenWIS) oder Rechtecke (UDE) dargestellt. Die Farbe der jeweiligen Station zeigt deren Status an (OK = grün, NOK = rot, Passiv = grau, keine Sensorinformationen = blau). Tooltips stellen Informationen über die IP-Adresse und den Namen der jeweiligen Station zur Verfügung. Durch Zoomen können Benutzer den für sie interessanten Bereich größer darstellen.

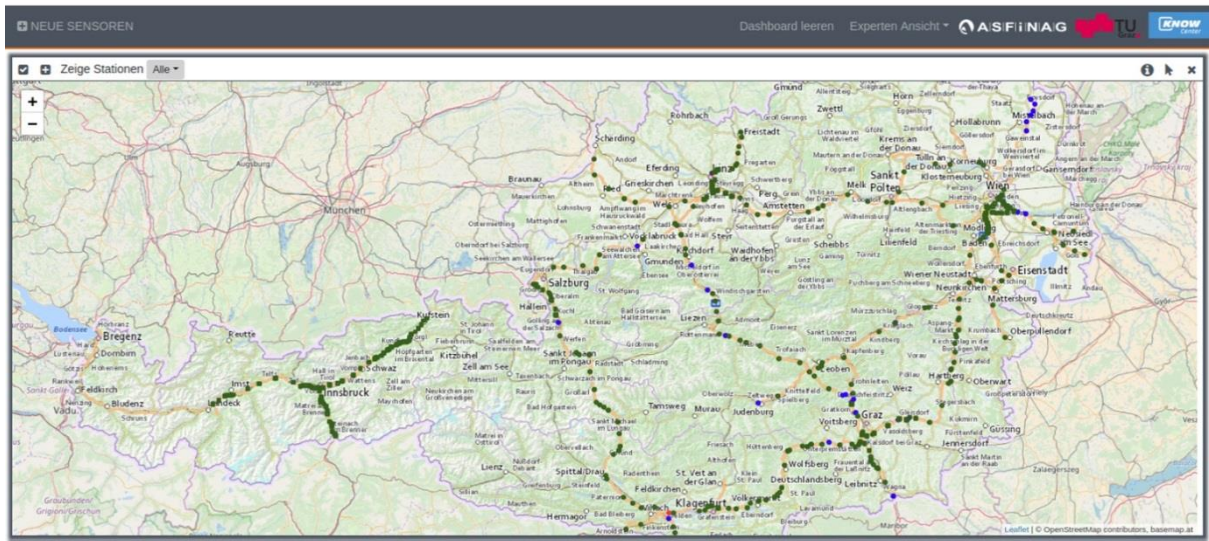


Abbildung 25: Die Geo-Visualisierungskomponente der Dashboard-Anzeige.

Durch Klicken auf eine der Stationen (außer jenen, die keine Sensorinformationen haben) erhalten BenutzerInnen detaillierte Informationen zu der Station in Form einer Sunburst Visualisierung (vgl. Abbildung 28). Des Weiteren können interessante Stationen markiert und für die weitere Analyse ausgewählt werden. Interessiert sich ein Benutzer nur für Stationen mit einem bestimmten Status z.B. NOK, so kann der entsprechende Filter gesetzt werden. Diese Visualisierung eignet sich dafür einen Überblick über alle Stationen und deren Status zu erhalten.

In Gesprächen mit den späteren BenutzerInnen des Prototyps wurde sehr oft der Wunsch nach einer Visualisierung von aufgezeichneten Messwerten anhand von LineCharts für manuelle Plausibilisierungsprüfungen geäußert, mögliche Visualisierungsformen wurden getestet und implementiert. Diese **LineCharts** (Abbildung 26) geben einen guten Überblick über den zeitlichen Verlauf von Sensorwerten: mehrere Sensoren können beispielsweise entweder in einem gemeinsamen oder in separaten Charts dargestellt werden, welche per Default synchronisiert werden.

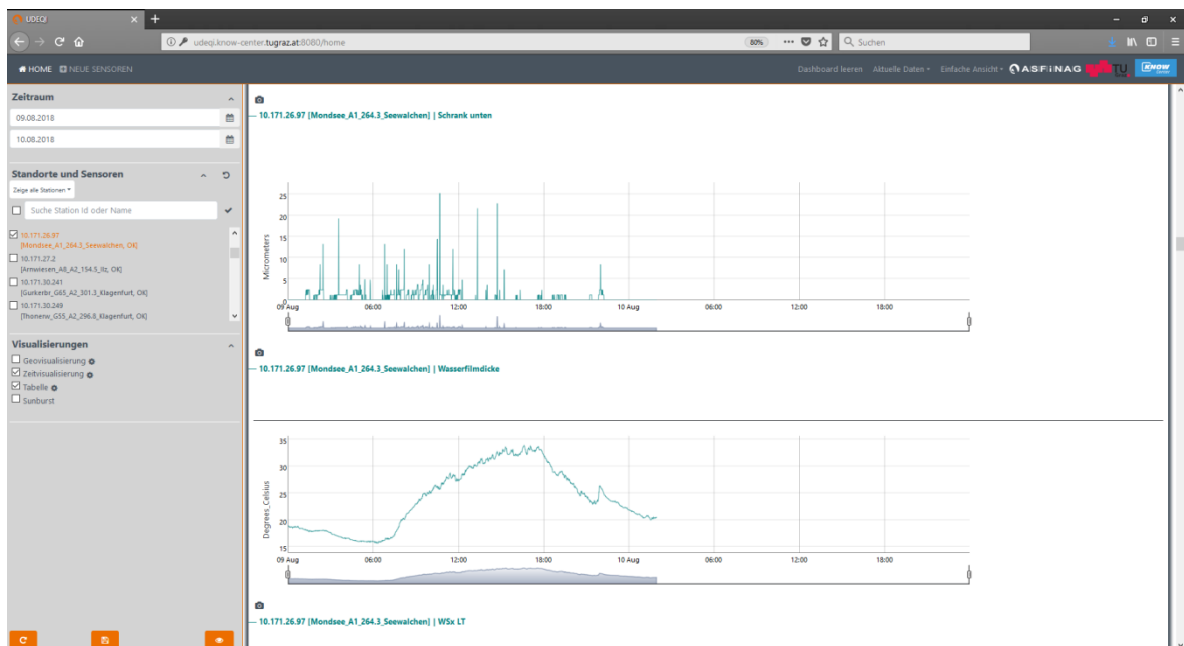


Abbildung 26: Ganglinien zweier Sensoren (Wasserfilmdicke, Lufttemperatur) mittels LineChart Visualisierungen.

Des Weiteren können BenutzerInnen im LineChart zoomen, um die Daten zu analysieren; bei Auffälligkeiten können entsprechenden Bereiche markiert und eine Annotation durchgeführt werden (vgl. violette Hinterlegung in Abbildung 27). Zudem werden sowohl Verfügbarkeit als auch Plausibilität dargestellt.

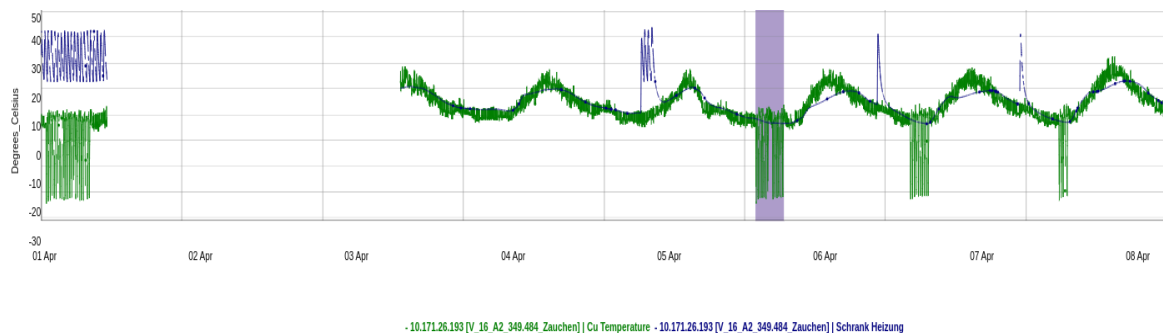


Abbildung 27: Manuelle Annotation von Sensorbereichen innerhalb eines LineCharts.

Mittels sogenannten „**Sunburst**“ Visualisierungen bekommen BenutzerInnen eine kompakte Übersicht über alle Sensoren einer Station; auf diese Weise können benachbarte Stationen einfach miteinander verglichen werden (Abbildung 28).

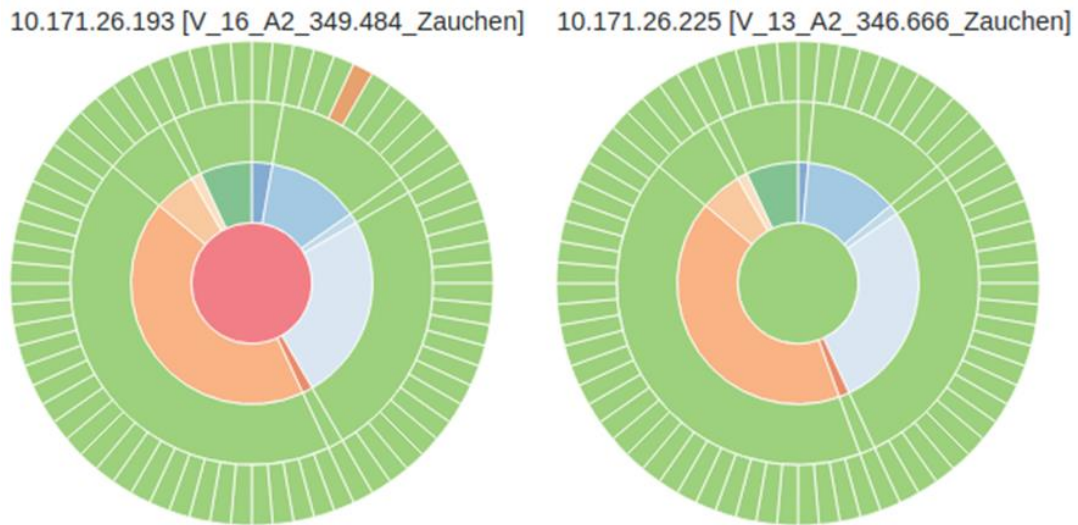


Abbildung 28: Einfacher Vergleich aller Sensoren zweier OpenWIS Stationen mittels Sunburst Visualisierungen.

Beschreibung der einzelnen Ebenen einer Sunburst Visualisierung:

- Erste (innerste) Ebene: Status der Station (OK = grün, NOK = rot, Passiv = grau)
- Zweite Ebene: Gruppierung der Sensoren nach Einheit (z.B. DegreesCelsius)
- Dritte Ebene: Gruppierung der Sensoren nach Status (aktiv vs. passiv)
- Vierte Ebene: alle Sensoren der Station. Sensoren, welche grün dargestellt werden, sind OK, Sensoren, welche orange dargestellt werden sind nicht plausibel und Sensoren, welche gelb dargestellt werden sind nicht verfügbar.

Durch Markierung einzelner Bereiche oder Sensoren kann der Benutzer oder die Benutzerin diese für die weitere Analyse auswählen. Tooltips auf den jeweiligen Bereichen beinhalten die oben beschriebenen Eigenschaften. Die Sunburst-Darstellung liefert eine kompakte Übersicht über alle Sensoren einer Station. BenutzerInnen bekommen einen schnellen Überblick, welche Sensoren nicht in Ordnung sind und ob es ein Problem der Verfügbarkeit oder Plausibilität handelt. Des Weiteren können benachbarte Stationen miteinander verglichen werden.

Des Weiteren gibt es in der Dashboard-Anzeige die Möglichkeit, den BenutzerInnen neben allgemeinen Informationen zu den Sensoren, die Plausibilität und Verfügbarkeit als auch statistische Informationen über den gewählten Zeitraum in einer **Tabelle** zur Verfügung zu stellen. Bei den Statistiken handelt es sich um den Mittelwert, den Median, den Modus (häufigster Wert), den Minimalwert, den Maximalwert, die Standardabweichung und die

Differenz zwischen minimalen und maximalen Wert. Diese Darstellung hilft den BenutzerInnen bei der Analyse der Sensorwerte.

2.4 Systemarchitektur

Abbildung 29 zeigt einen Überblick über die implementierte Systemarchitektur des wissenschaftlichen Prototyps. In Abschnitt 2.4.1 werden die einzelnen Module sowie ihr Zusammenwirken näher beschrieben. In Abschnitt 2.4.2 werden die Vorteile der gewählten Architektur zusammengefasst.

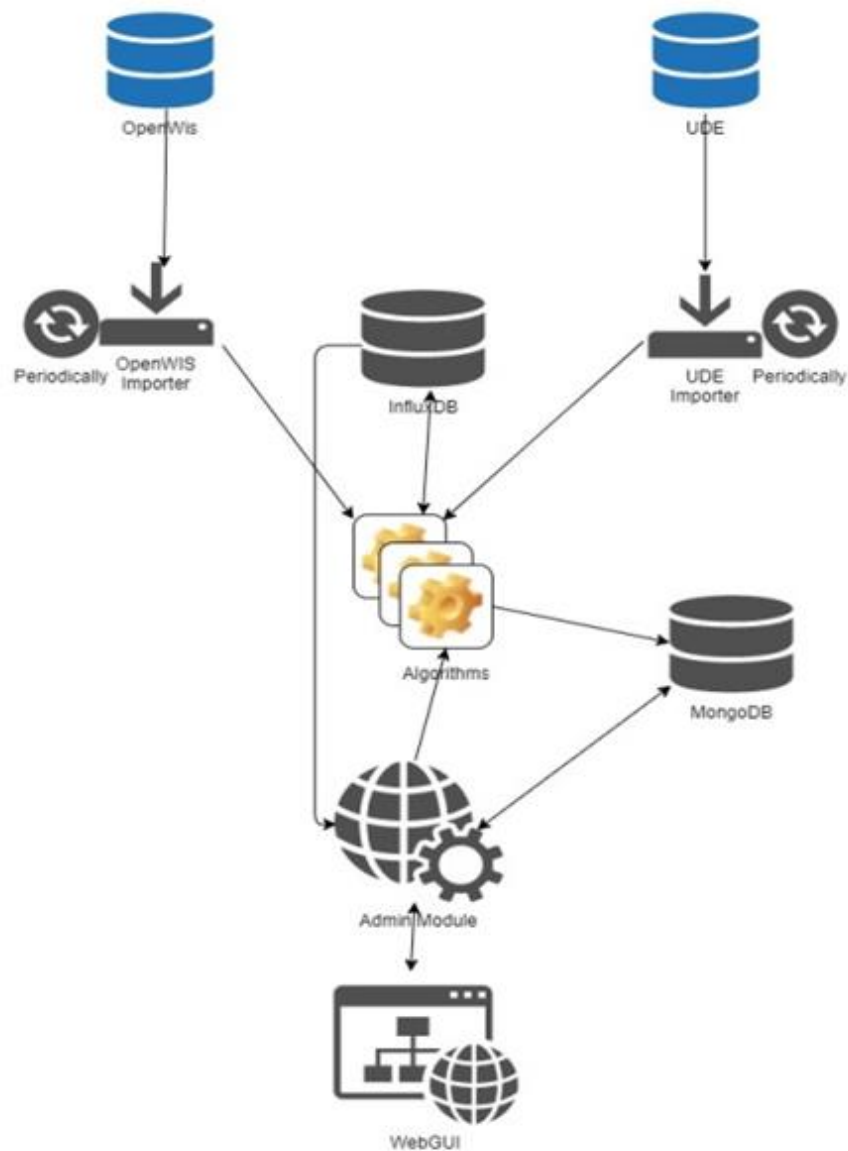


Abbildung 29: Überblick über die UDEQI Systemarchitektur.

2.4.1 Module & Workflow

- Es gibt zwei Import-Module, die in regelmäßigen Abständen die Daten aus den beiden Datenquellen (OpenWIS und UDE) abgreifen und für die weitere Verarbeitung aufbereiten.
- Die Algorithmen-Module prüfen die Daten auf Konsistenz und melden diese im Falle von Anomalien an die Backend-Datenbank.
- Die InfluxDB ist eine Datenbank, optimiert auf das Speichern und Auslesen von Zeitreihen. Im aktuellen Framework wird sie benutzt, um die vorverarbeiteten Sensordaten der einzelnen Sensorstationen zu speichern.
- Als Backend-Datenbank wird die MongoDB¹⁴ verwendet, um zusätzliche Informationen zu speichern. Dazu zählen unter anderem Informationen über den Zustand von Sensorstationen, von Algorithmen detektierte Anomalien und Konfigurationen der einzelnen Algorithmen-Modulen. Die MongoDB speichert also alles, wofür die InfluxDB nicht ausgelegt ist. Die MongoDB spiegelt auch den aktuellen Status des Systems wieder, der kontinuierlich von der graphischen Benutzeroberfläche abgefragt und angezeigt wird.
- Das Admin-Modul besteht aus einem web-basierten Frontend Modul zur Darstellung und Verwaltung der Sensorinformationen. Über eine Web-Oberfläche kann die aktuelle Sensorenlandschaft untersucht werden, sowie Änderungen an den verwendeten Detektionsalgorithmen vorgenommen werden. Sie dient auch der Darstellung und dem Export der aktuell in der InfluxDB gespeicherten Sensorinformation. Das Admin-Modul ist auch für die Bereitstellung der Resultate aus der Verfügbarkeits- sowie der Plausibilitätsanalyse zuständig. Der jeweilige Status einer Station ist im Admin-Modul abrufbar – folgende Tabelle 4 enthält die mögliche Zustände, die eine Station bzw. ein Sensor annehmen kann, sowie die entsprechende farbliche Darstellung in der web-basierten Benutzeroberfläche:

¹⁴ <https://www.mongodb.com/>

Farbcodierung		Beschreibung
Grün	→	Station ist OK
Blau	→	keine Sensordaten; Station sendet nicht
Grau	→	Station passiviert
Rot	→	Station nicht OK; Fehler
Orange	→	Sensordaten nicht plausibel
Gelb	→	Sensordaten nicht verfügbar

Tabelle 4: Mögliche Zustände, die eine Station bzw. ein Sensor annehmen kann mit der entsprechenden farbliche Darstellung in der web-basierten Benutzeroberfläche

Der farblichen Indikation eines Fehlers wird in der Benutzeroberfläche ein Fehlercode hinzugefügt, der beispielsweise Auskunft darüber gibt, welcher Plausibilitätsalgorithmus nun angeschlagen hat.

Datenfluss

Das Import-Modul liest die neuen Daten, die seit dem letzten Durchlauf entstanden sind, ein und bringt sie in ein geeignetes Format zur weiteren Verarbeitung. Entsprechende Algorithmen prüfen die Daten und melden etwaige Anomalien an die Backend-Datenbank. Die verarbeiteten Sensordaten werden in der InfluxDB für die weitere Verwendung abgelegt. Algorithmen vergleichen die neu hinzugekommenen Daten mit den bereits gespeicherten Daten, um entsprechend Anomalien zu erkennen. Das Admin-Modul dient dem Benutzer als einfaches Mittel zur Übersicht über den aktuellen Zustand der Sensorlandschaft. Im Fehlerfall können so Fehler einfach untersucht werden und entsprechend darauf reagiert werden.

2.4.2 Vorteile der gewählten Architektur

Ein entscheidendes Plus bei Verwendung der InfluxDB stellen zweifelsfrei der effiziente Datenimport sowie die Kompressionsrate dar. Diese Charakteristika, die in der Theorie angepriesen wurden, konnten in der Praxis bestätigt werden:

- Ein Import von OpenWIS-Daten von ~220 Stationen über die letzten 5 Jahre benötigte ungefähr 5 Stunden.
- Mit einer 10-fachen Kompressionsrate nahmen diese OpenWIS Daten ~5 GB Speicherplatz ein.

Die Wahl der InfluxDB bietet nun folgende Möglichkeiten an, mit einer zu erwartenden größer werdenden Datenmenge umzugehen und zwar:

- Die Veränderung der Datenbankgranularität, beispielsweise die Erstellung von einer Datenbank pro Jahr an Daten.
- Die Änderung der Aufbewahrungsstrategie, d.h. Daten, die beispielsweise älter als zwei Jahre sind, werden in eine andere Datenbank verschoben. Der Zugriff auf diese Daten wird damit natürlich auch verlangsamt.
- Der Umstieg auf die Enterprise Edition der InfluxDB, die kostenpflichtig ist, und somit eine Erhöhung der Skalierbarkeit.

Ein weiteres Merkmal der gewählten Architektur ist, dass alle Komponenten des UDEQI Systems verteilbar sowie skalierbar sind. Des Weiteren sind sie auf *Zero Downtime* ausgelegt, d.h. neue Algorithmen können sich im laufenden Betrieb anmelden, ohne dass dafür die Datenbank heruntergefahren werden müsste.

Ein letztes, nicht unwichtiges Merkmal der gewählten Architektur ist, dass alle Komponenten mit Open-Source Bibliotheken implementiert worden sind.

3 REFERENZTESTFELD

Im ersten Schritt wurden die Anforderungen und Umsetzung des Referenztestfeldes abgestimmt, danach wurden von Seiten der ASFiNAG Hersteller von bereits im ASFiNAG-Netz verbauter Sensorik eingeladen sich an einem geplanten Realtest ab Sommer 2017 zu beteiligen. Während der Verhandlungen mit den Sensorikherstellern konnte keine Einigung erzielt werden.

Durch den notwendigen Klärungsprozess zur Finanzierung und administrativen Betreuung des Referenztestfeldes-Umfelddatenerfassung von Seiten der ASFiNAG wurde das Bearbeitungsende des Referenztestfeldes zunächst auf Dezember 2017 geändert. Eine, nach Absage der Teilnahme durch die Sensorikhersteller, notwendige ASFiNAG-interne Finanzierung des Referenztestfeldes konnte trotz erster positiver Signale nicht aufgestellt werden. Daher wurde die Möglichkeit eines „virtuellen“ Referenztestfeldes evaluiert und für den weiteren Projektverlauf umgesetzt. In diesem „virtuellen“ Referenztestfeld wurde neu verbaute oder ausgetauschte Sensorik an möglichst wenigen Standorten im Netz der ASFiNAG als Referenzgröße herangezogen und mit bestehender bereits länger im Einsatz befindlicher Sensorik verglichen (vgl. Abschnitt 2.2).

In Ergänzung des virtuellen Referenztestfeldes wurde ein ausgewählter im ASFiNAG-Autobahnnetz oft eingesetzter Sensortyp auf Abweichungen durch den Alterungsprozess getestet. Hierzu wurde ein Lufft WS300-UMB All-in-One-Wettersensor durch die ASFiNAG gekauft und parallel zu einem bestehenden Sensor des gleichen Typs an der A1 (Streckenkilometer 246,52; Richtung Walserberg, vgl. Abbildung 30) im Dezember 2017 verbaut. Der bestehende Sensor wurde am 01.08.2011 an der Station verbaut und war somit bis zu diesem Zeitpunkt bereits 6,37 Jahre in Betrieb ohne eine vom Hersteller empfohlene Re-Kalibrierung nach 2 Jahren. Dieser bestehende Sensor wird nachfolgend mit dem neu verbauten und kalibrierten Sensor statistisch für den Zeitraum 01.01.2018 bis 30.06.2018 beispielhaft verglichen.



Abbildung 30: OpenWIS-Station auf der A1 an Streckenkilometer 246,52 in Richtung Walsberg, die Testsensorik ist in Rot hervorgehoben

In der Analyse der beiden Sensoren, wird davon ausgegangen, dass der neu verbaute Sensor durch die zuvor durchgeführte aktuelle Kalibration, den Herstellerspezifikationen entsprechend korrekte Werte für die Parameter Lufttemperatur, Luftdruck, Relative Feuchtigkeit und Taupunkttemperatur ermittelt. Durch den Vergleich mit den ermittelten Werten des bereits seit 6,37 Jahren betriebenen Sensors soll der Einfluss der Re-Kalibration auf die Messwertermittlung dieses Sensortyps aufgezeigt werden.

Dazu wurden die ermittelten Werte für die jeweiligen Messparameter der beiden Sensoren mittels eines t-Tests verglichen (vgl.

Tabelle 5). Der Test weist unterschiedliche Ergebnisse für die jeweiligen Parameter auf, z.B. unterscheiden sich die Werte für Lufttemperatur nicht statistisch signifikant, hingegen unterscheiden sich Luftdruck, Relative Feuchtigkeit und Taupunkttemperatur statistisch signifikant.

Parameter	t-Wert	DOF	p-Wert	Stat. signifikant
Lufttemperatur	-0.881	477.750	0.3783	Nein
Luftdruck	62.907	362.060	< 2.2e-16	Ja
Relative Feuchtigkeit	22.288	472.080	< 2.2e-16	Ja
Taupunkttemperatur	10.476	476.460	< 2.2e-16	Ja

Tabelle 5: Vergleich der t-Test-Ergebnisse für den Vergleich der unterschiedlichen ermittelten Parameter beider Sensoren

Dieses Ergebnis bestätigen auch die Visualisierungen der Messergebnisse (vgl. Abbildungen 31 bis 34). Mit Ausnahme des Parameters Luftdruck befinden sich die Messwerte beider Sensoren grob im Rahmen der jeweiligen vom Hersteller angegebenen Genauigkeiten.

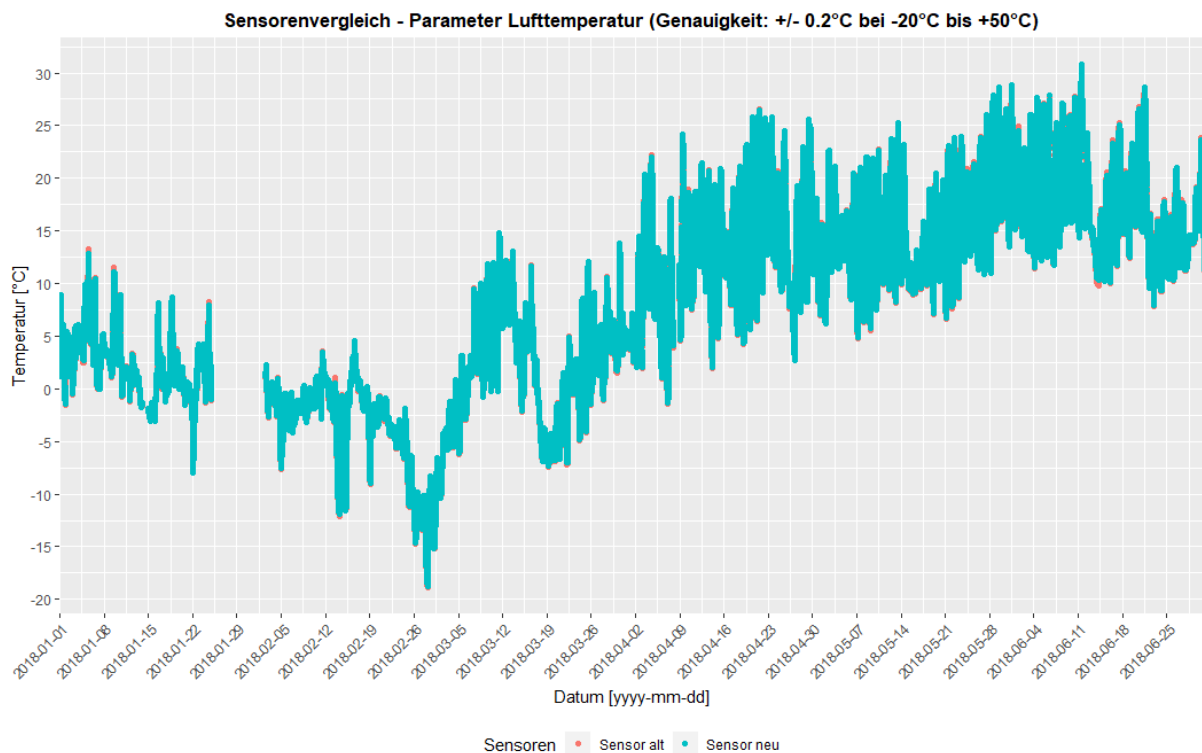


Abbildung 31: Sensorenvergleich - Parameter Lufttemperatur (Genauigkeit: +/- 0.2°C bei -20°C bis +50°C) für den Zeitraum 01.01.2018 bis 30.06.2018

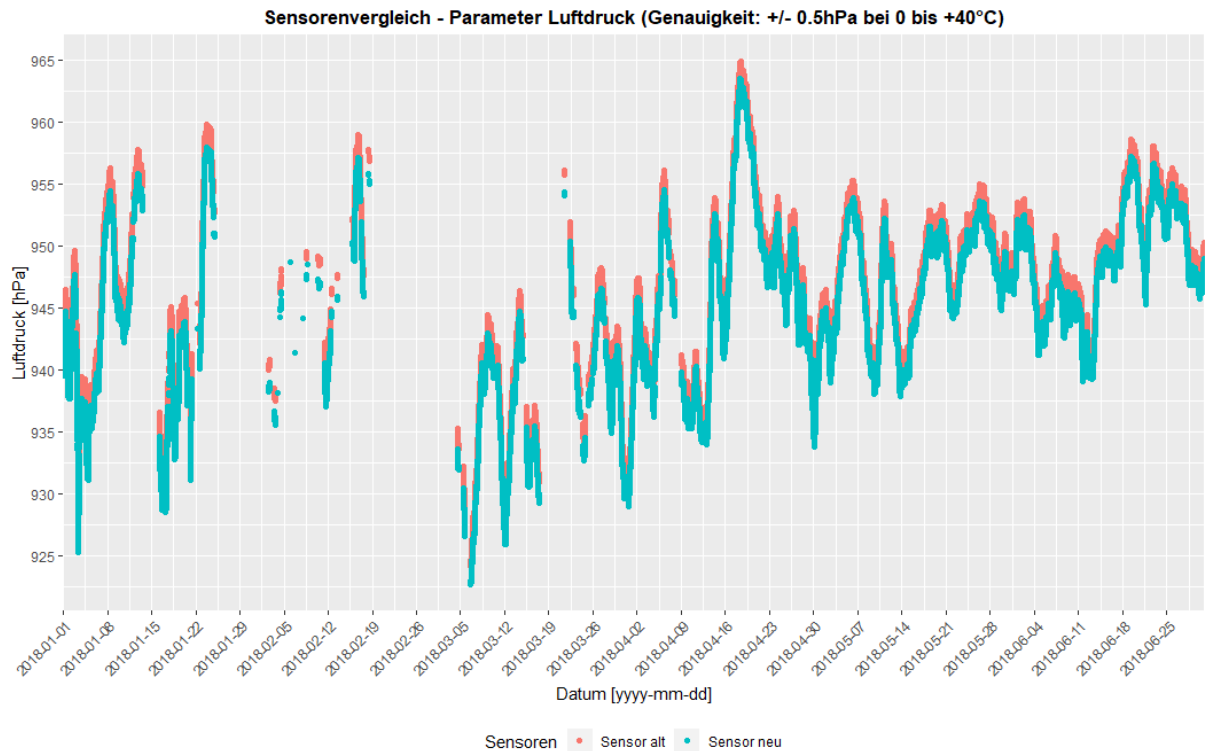


Abbildung 32: Sensorenvergleich - Parameter Luftdruck (Genauigkeit: +/- 0.5hPa bei 0 bis +40°C) für den Zeitraum 01.01.2018 bis 30.06.2018, exklusive von Messwerten außerhalb der Genauigkeit (bei Temperaturen von unter 0°C)

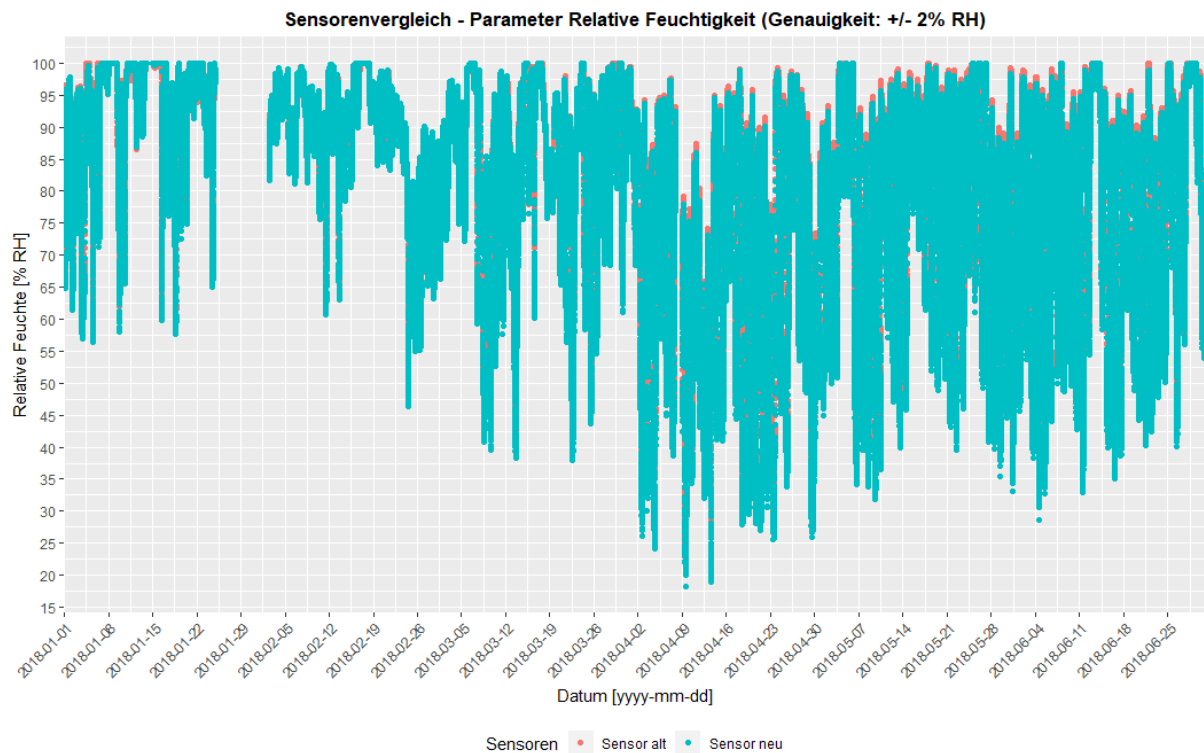


Abbildung 33: Sensorenvergleich - Parameter Relative Feuchtigkeit (Genauigkeit: +/- 2% RH) für den Zeitraum 01.01.2018 bis 30.06.2018

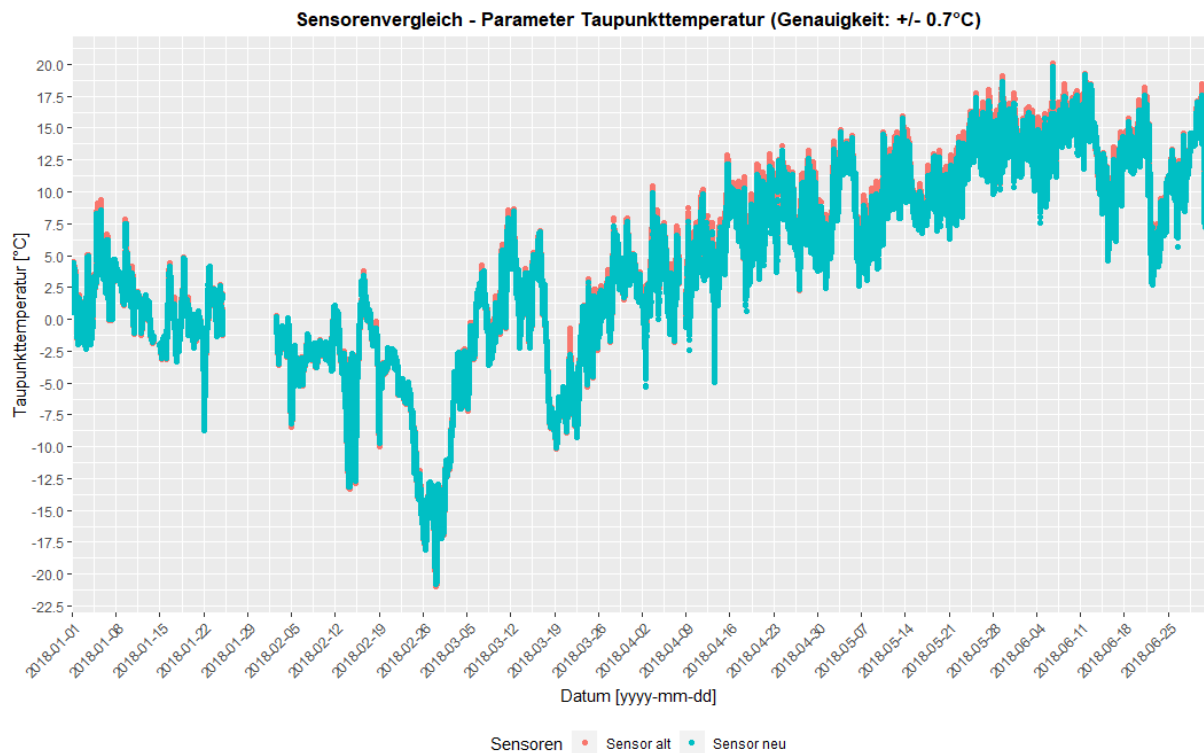


Abbildung 34: Sensorenvergleich - Parameter Taupunkttemperatur (Genauigkeit: +/- 0.7°C) für den Zeitraum 01.01.2018 bis 30.06.2018

Im nächsten Schritt sollte dieser Umstand durch die ASFiNAG anhand der UDEQI-Daten mit den zugehörigen Plausibilitäten und Qualitätskriterien, weiter beobachtet und gegebenenfalls Prozesse zur Sicherstellung der Datenqualität durch geeignete Sensorwartung vorangetrieben werden.

4 VALIDIERUNG DER PROTOTYPISCHEN ERGEBNISSE ANHAND EINES REALEN ANWENDUNGSSZENARIOS

Der wissenschaftliche Prototyp wurde gemeinsam mit der ASFiNAG anhand eines umfangreichen Anwendungsszenarios evaluiert. In diesem Kapitel wird dieses Szenario anhand von Screenshots detailliert und nachvollziehbar beschrieben, um (i) die Funktionalität, (ii) die Komplexität sowie (iii) die Benutzbarkeit des UDEQI Systems zu demonstrieren. Dafür wurde ein möglichst realistisches Szenario ausgewählt.

Die Ausgangssituation stellt sich wie folgt dar:

- Die OpenWIS Station 10.171.26.193 (nahe Velden), als auch deren Sensoren „Schrank Heizung“ und „Cu Temperature“ sollen passiviert werden; für das Szenario ist die Passivierung der Indikator, dass die Station bzw. die Sensoren nicht OK sind und werden dementsprechend „rot“ markiert.
- Als Detail: Um Stationen und Sensoren zu passivieren, gibt es 2 Möglichkeiten, die nur im Experten-Modus verfügbar sind.
 - Stationen und Sensoren: Icon neben dem Stations- bzw. Sensornamen (im „Home“- und „Daten aktualisieren“-Tab)
 - Sensoren: Tab „Daten aktualisieren“ → Auswahl des Sensors → Feld „Status“ ändern

Ein ASFiNAG Mitarbeiter oder eine ASFiNAG-Mitarbeiterin (im Folgenden als BenutzerIn bezeichnet) möchte sich einen Überblick über alle Stationen verschaffen. Er oder sie möchte überprüfen, ob alle Stationen und die zugehörigen Sensoren ordnungsgemäß funktionieren. Sollte es Stationen bzw. Sensoren geben, die nicht in Ordnung sind, möchte der Benutzer oder die Benutzerin diese detailliert analysieren.

Start des Anwendungsszenarios:

Um einen Überblick über alle Stationen zu erhalten, wählt der Benutzer oder die Benutzerin den Tab „Home“ und lässt sich alle Stationen in der Geo-Visualisierung anzeigen (vgl. Abbildung 35).

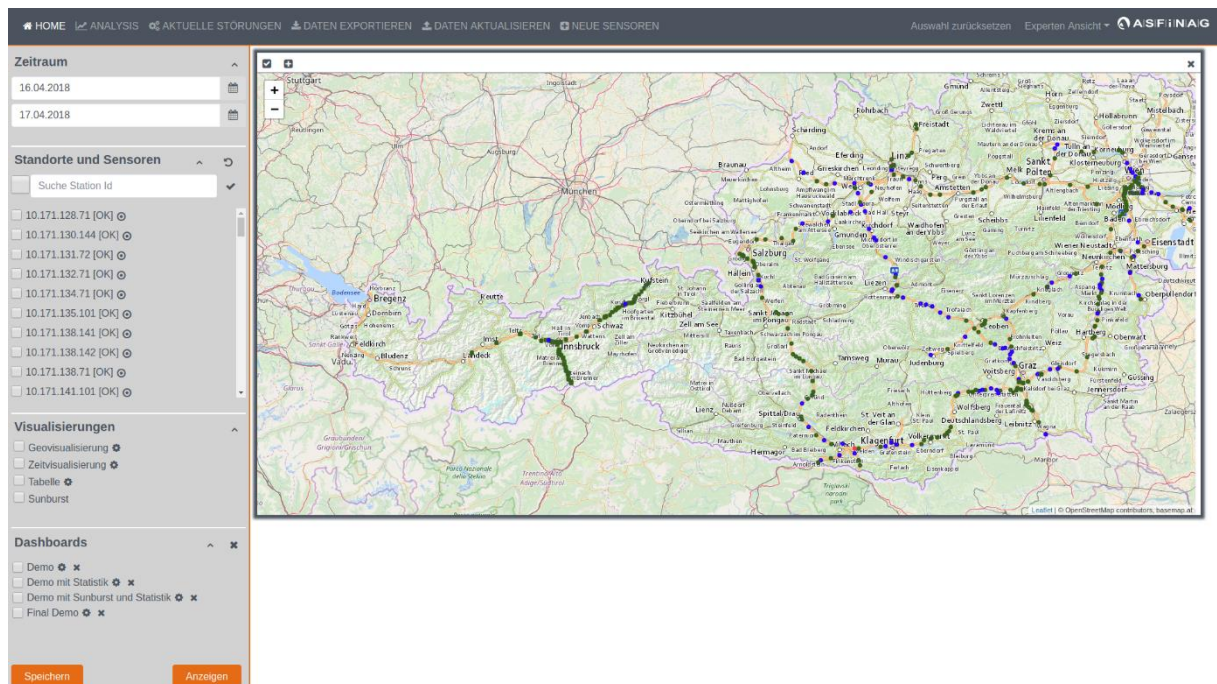


Abbildung 35: Übersicht aller OpenWIS und UDE Stationen. Die jeweilige Färbung codiert den aktuellen „Zustand“ einer Station (OK = grün, NOK = rot).

Dabei sieht der Benutzer oder die Benutzerin, dass eine Station in Kärnten nahe Velden „rot“ markiert ist, ein Zeichen dafür, dass etwas nicht in Ordnung ist. Um detailliertere Informationen zu erhalten klickt der Benutzer oder die Benutzerin auf diese Station und erhält die Sunburst-Darstellung der Station (vgl. Abbildung 36).

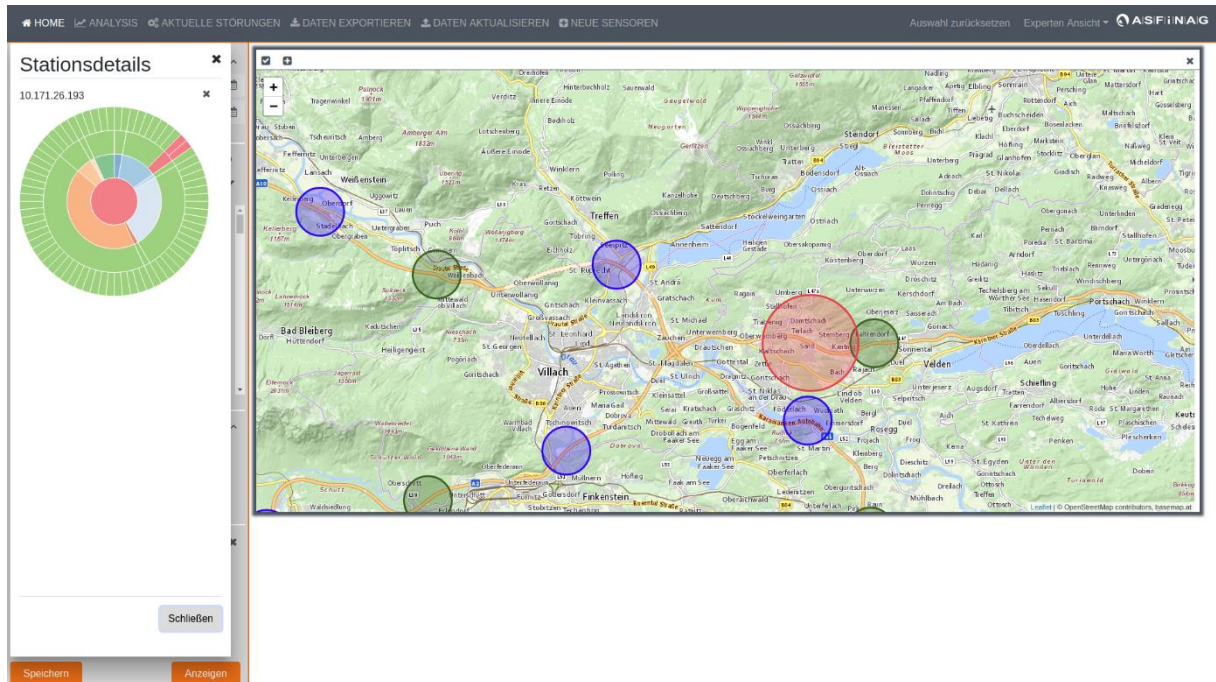


Abbildung 36: Übersicht aller Sensoren der Station nahe Velden mittels der Sunburst Visualisierung.

Die Sunburst-Visualisierung im linken oberen Abschnitt von Abbildung 36 zeigt zwei Sensoren der Station als rot (nicht OK) markiert an. Da beide Werte der Einheit „Degrees_Celsius“ angehören, markiert der Benutzer oder die Benutzerin alle Sensoren der Station mit dieser Einheit, um die Werte miteinander zu vergleichen (vgl. Abbildung 37).

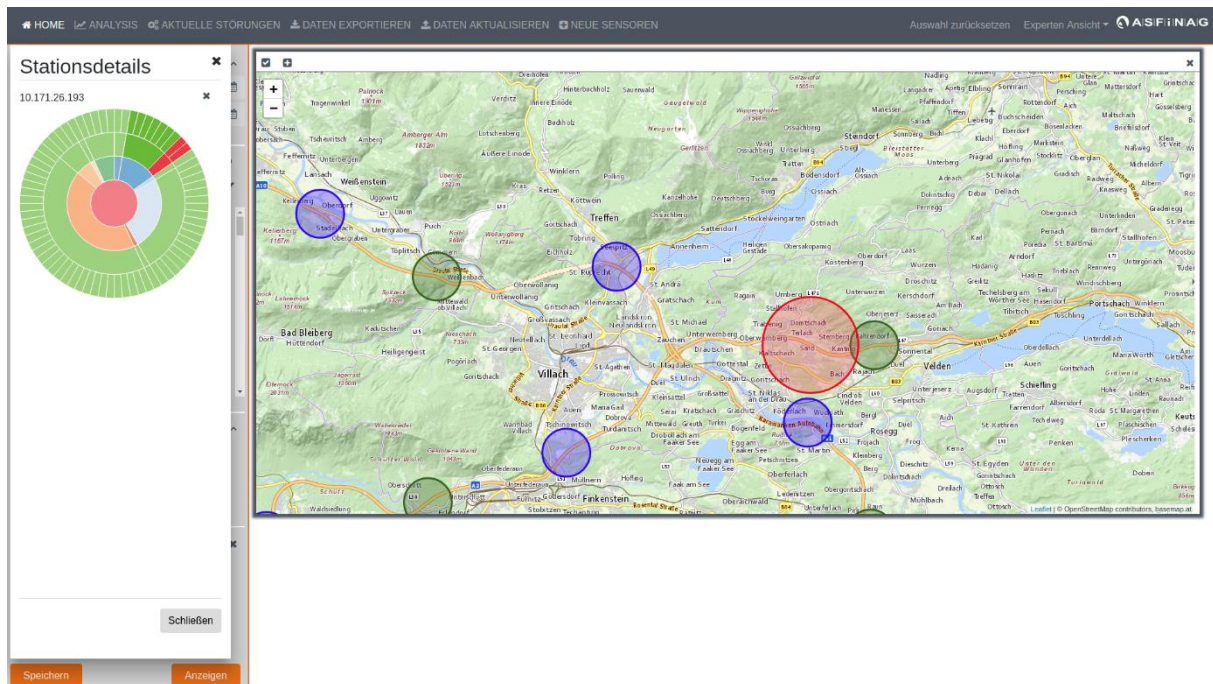


Abbildung 37: Auswahl aller Sensoren der Station nahe Velden mit der Einheit „Degree_Celsius“.

Im nächsten Schritt schließt der Benutzer oder die Benutzerin die Sunburst-Visualisierung und wendet diese Auswahl, d.h. alle Sensoren mit der Einheit „Degree_Celsius“ links an (Icon oberhalb der Geo-Visualisierung). Der Benutzer oder die Benutzerin interessiert sich nur für die Werte einer speziellen Woche und wählt den Zeitraum von 04.04.2018-11.04.2018 aus (vgl. Abbildung 38). Um einzelne Werte darzustellen, wählt der Benutzer oder die Benutzerin die Zeitvisualisierung mit der Einstellung des Zusammenfügens der Sensoren einer Station aus (vgl. Abbildung 38). Diese Funktionalität erlaubt es dem Benutzer oder der Benutzerin, die Werte aller ausgewählten Sensoren in einer Auswertung angezeigt zu bekommen; diese können dann leichter miteinander verglichen werden. Wird auf die Einstellung des Zusammenfügens verzichtet, werden die Werte eines jeden Sensors in einer separaten Auswertung dargestellt.

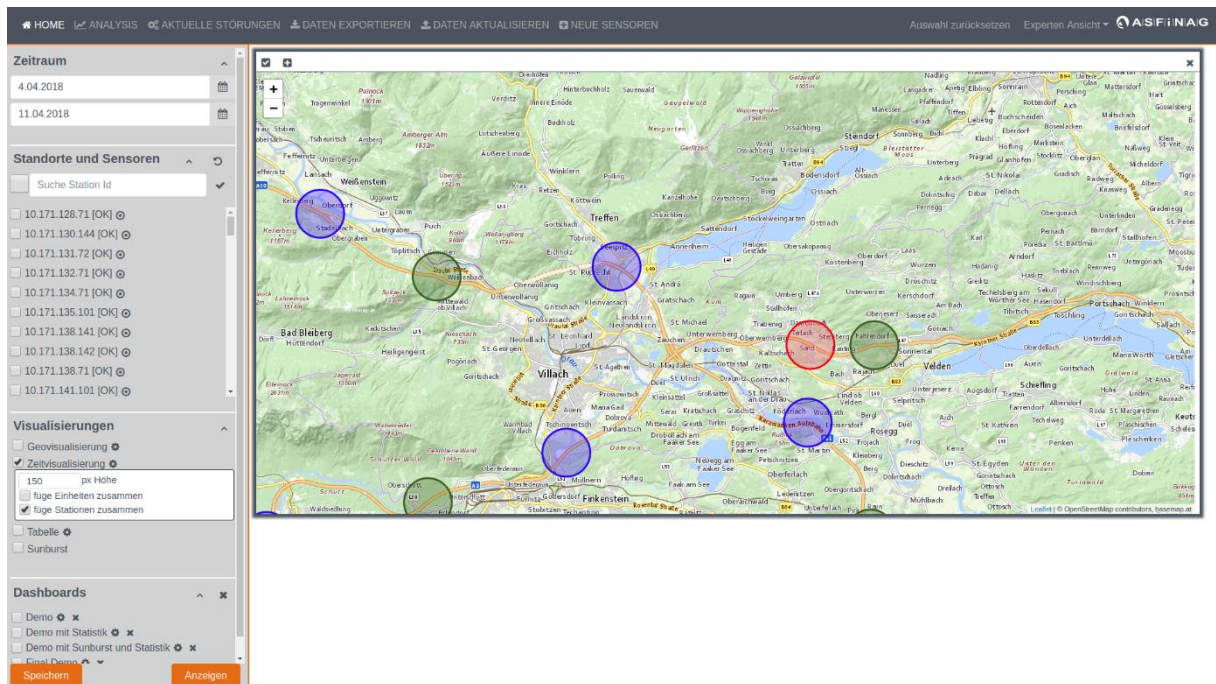


Abbildung 38: Auswahl des gewünschten Zeitraums und der gewünschten Visualisierung.

Das Ergebnis dieser Auswertung wird im unteren Bereich von Abbildung 39 angezeigt. Die Zeitvisualisierung zeigt dem Benutzer oder der Benutzerin, dass zwei Sensoren „Schrank Heizung“ und „Cu Temperature“ Einzelausreißer haben.

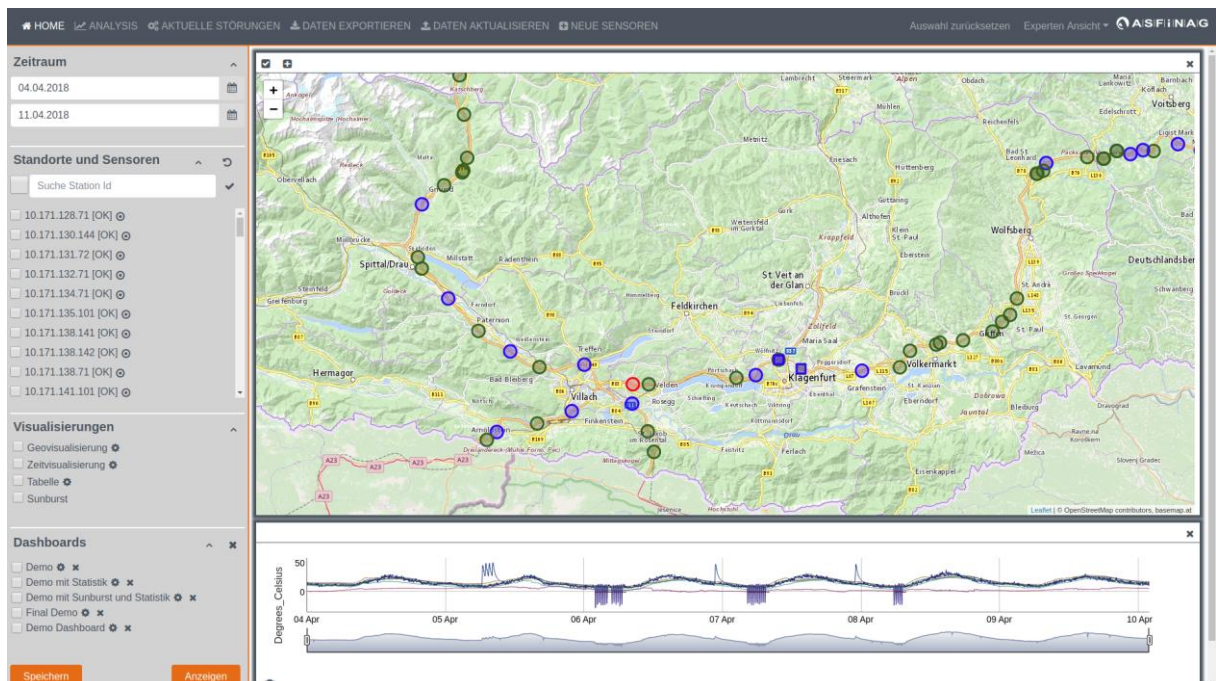


Abbildung 39: Zeitliche Darstellung der gewählten Sensoren der Station nahe Velden.

Um die Werte genauer analysieren zu können hat der Benutzer die Möglichkeit mittels des Range-Selectors (unterhalb der Darstellung) oder mittels der Kombination Shift + Maus (Mousedown + Mousemove + Mouseup) den gewünschten Zeitraum auszuwählen (vgl. Abbildung 40 und Abbildung 41).

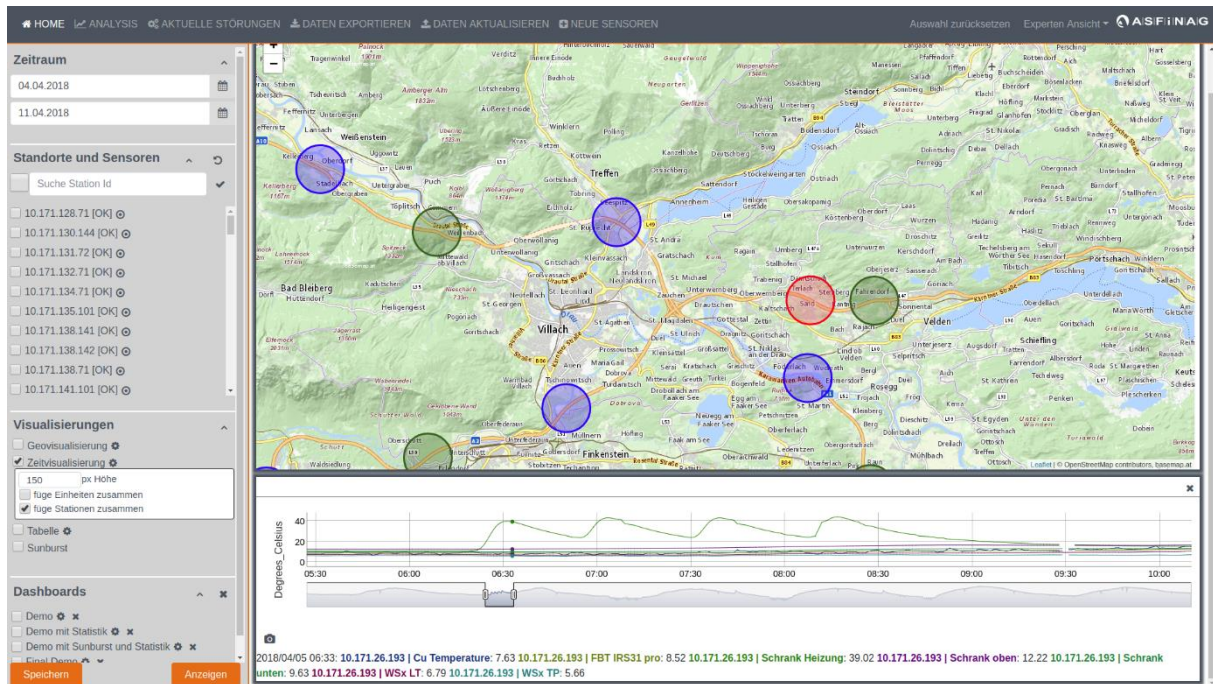


Abbildung 40: Zoomen in einem Bereich um die Daten genauer zu analysieren.

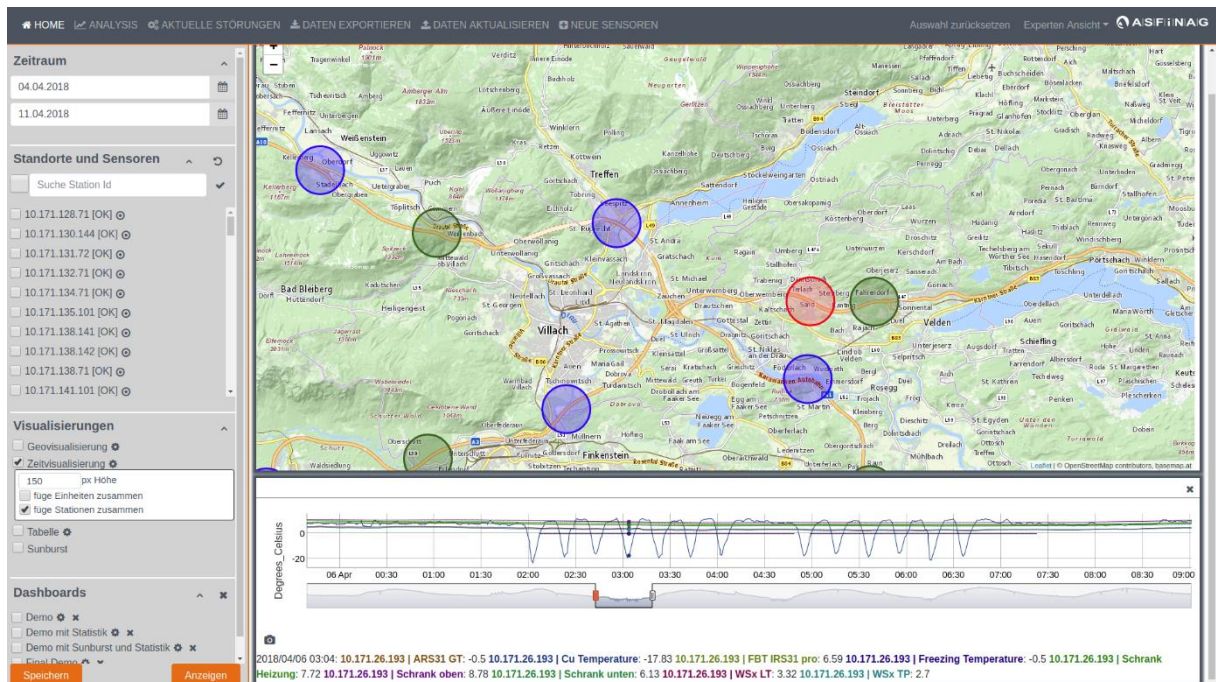


Abbildung 41: Zoomen in einem Bereich um die Daten genauer zu analysieren.

Möchte sich der Benutzer oder die Benutzerin einen Bereich für eine etwaige spätere Analyse merken, so kann dieser mittels Ctrl + Maus (Mousedown + Mousemove + Mouseup) markiert werden (vgl. Abbildung 42). Der Benutzer oder die Benutzerin kann diesen Bereich textuell beschreiben; er oder sie kann also seine Bedenken oder Ideen für eine spätere Analyse, etc. mit dem ausgewählten Bereich verknüpfen und abspeichern.

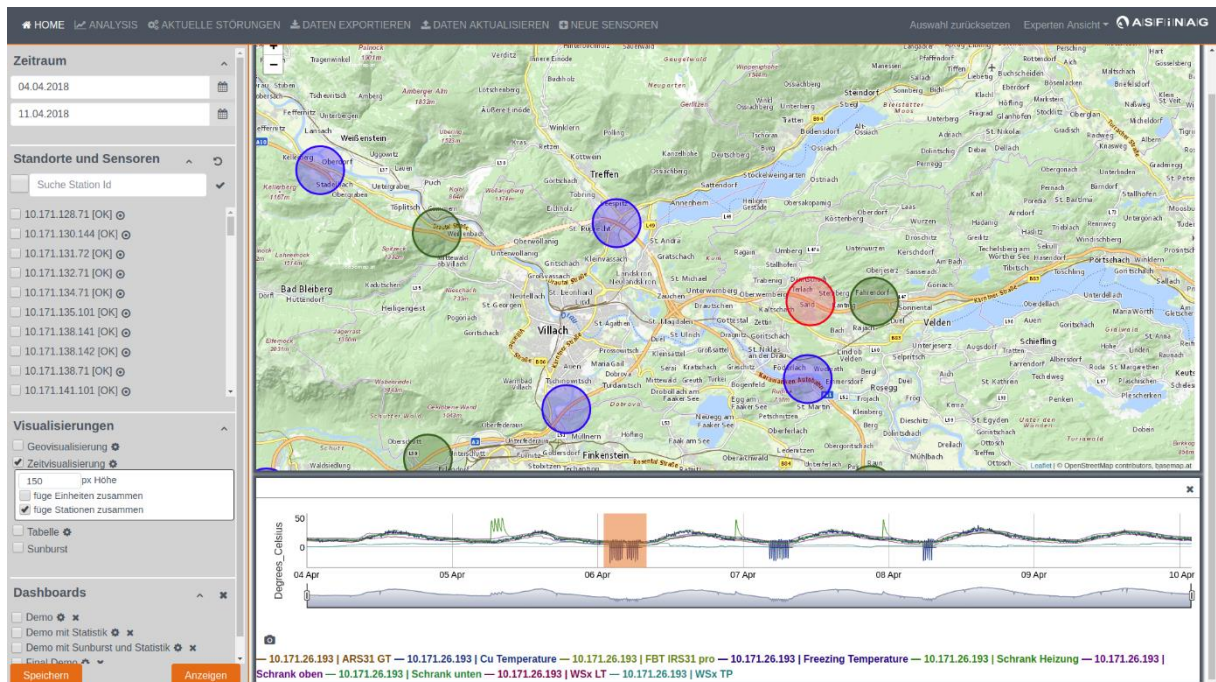


Abbildung 42: Markierung (Annotation) eines gewünschten Bereichs (in orange) für eine spätere Analyse.

Der Benutzer oder die Benutzerin hat auch die Möglichkeit, Auswertungen, z.B. Zeitvisualisierungen, für eine Berichterlegung als PNG zu exportieren (vgl. Abbildung 43).

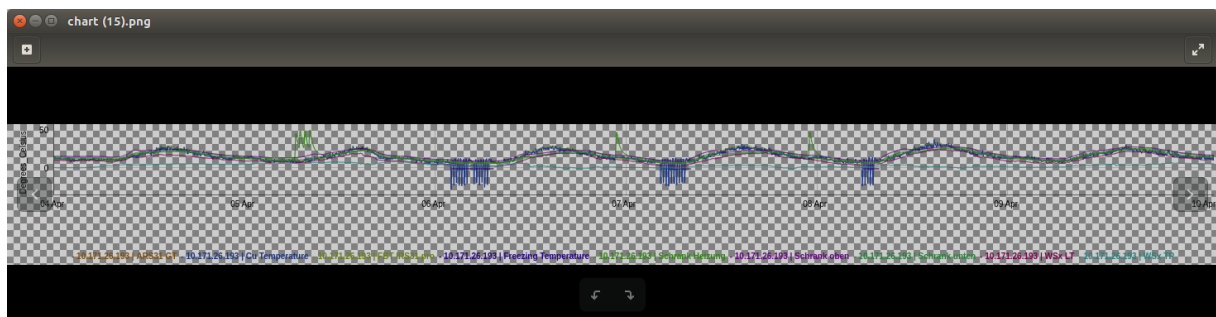


Abbildung 43: Auswertung zur Verwendung in einem Bericht als PNG exportieren.

Nachdem der Benutzer oder die Benutzerin alle Sensoren der Station nahe Velden mit der Einheit „Degree_Celsius“ analysiert hat, richtet er seine Konzentration auf jene beiden Sensoren, welche Ausreißer aufweisen – also nicht ok sind und dementsprechend „rot“ markiert sind.

Dazu öffnet der Benutzer oder die Benutzerin die Station wieder in der Geo-Visualisierung und markiert nur mehr jene Sensoren, welche „rot“ dargestellt sind (vgl. Abbildung 44).

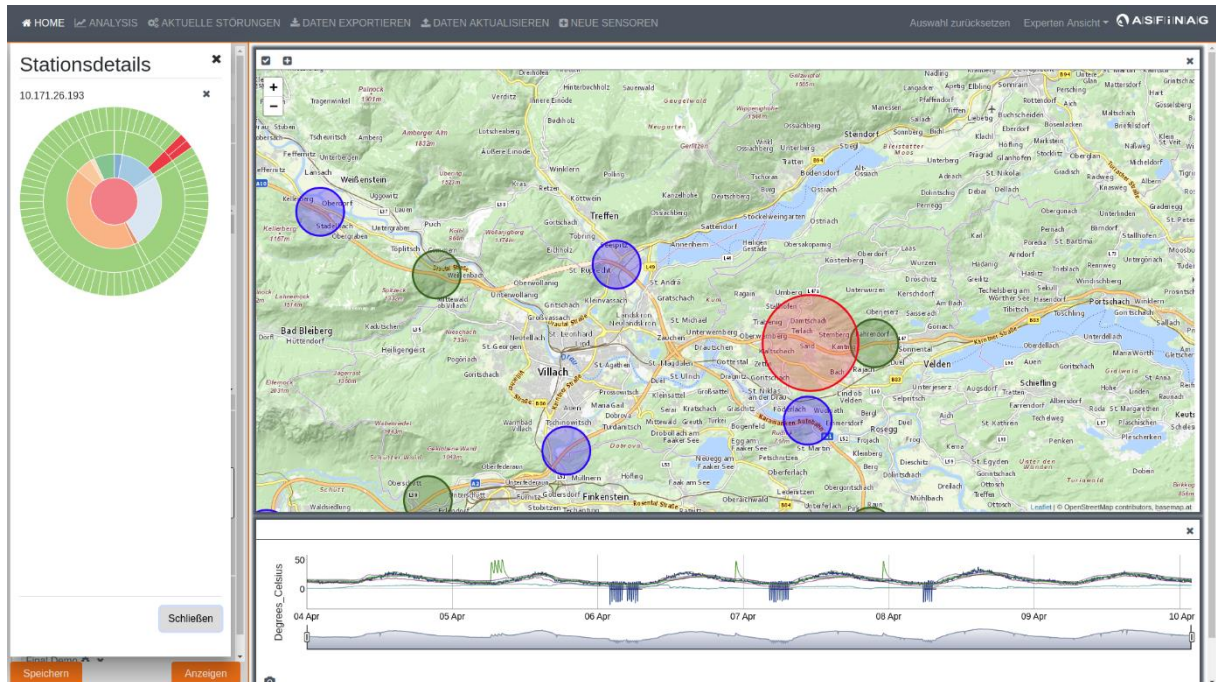


Abbildung 44: Auswahl der Sensoren, welche nicht in Ordnung sind („rot“ dargestellt)

Der Benutzer oder die Benutzerin möchte nun statistische Informationen über diese beiden Sensoren erhalten. Dazu wählt er die Tabelle als Darstellungsform und erhält detaillierte Informationen zu den Statistiken (vgl. Abbildung 45).

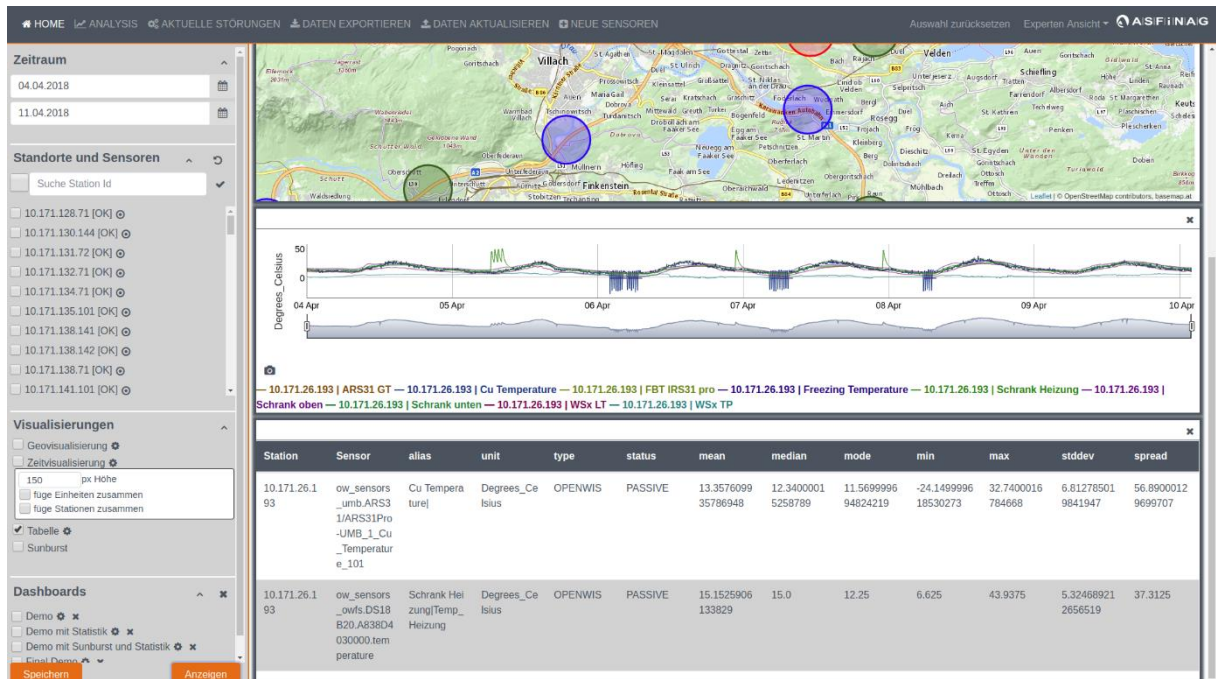


Abbildung 45: Statistische Informationen zu den beiden Sensoren, die "rot" markiert waren.

Bevor der Benutzer oder die Benutzerin eine Entscheidung trifft, möchte er die Daten mit einem Kollegen besprechen. Um die Daten zu exportieren und somit an seine oder ihre KollegInnen schicken zu können, ist es nötig, den Tab zu wechseln. Zuvor jedoch, speichert er die aktuelle Dashboard Ansicht (vgl. Abbildung 46).

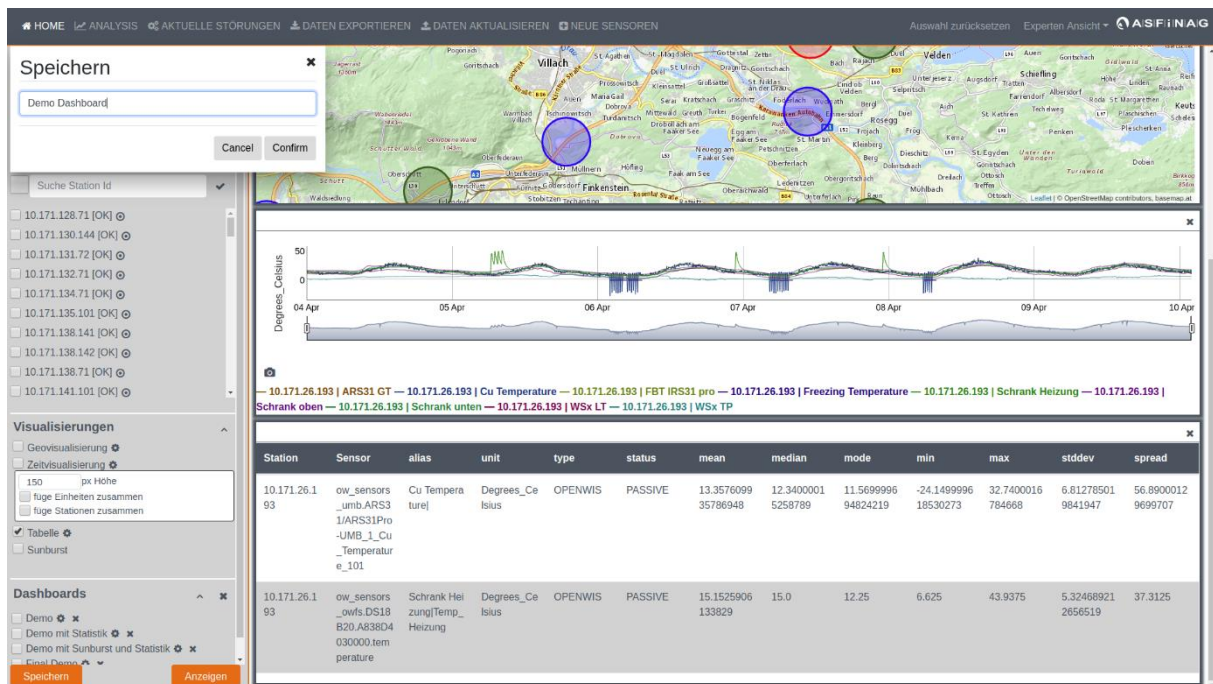


Abbildung 46: Speichern der aktuellen Dashboard Ansicht zu späteren Wiederverwendung.

Beim Exportieren behält der Benutzer oder die Benutzerin nun die zuvor getätigte Auswahl und wählt sowohl Daten und Statistiken für den Export (vgl. Abbildung 47).

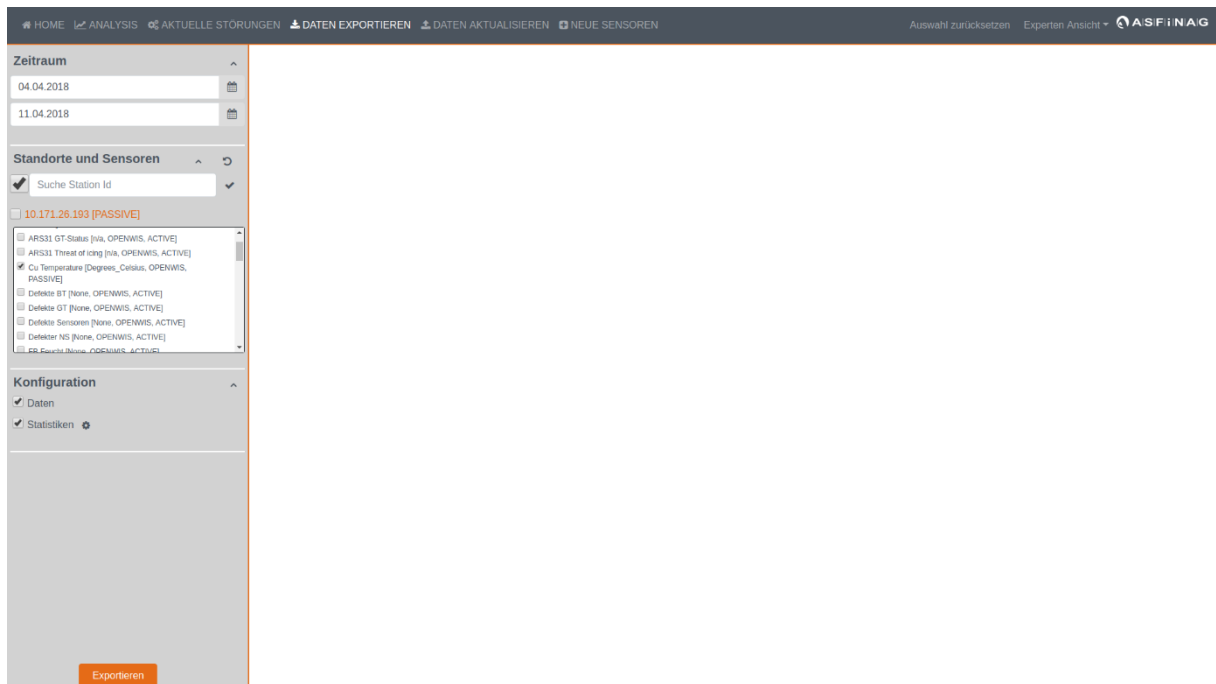


Abbildung 47: Exportieren des Dashboards (Daten und Statistiken) als CSV.

In der Folge werden zwei CSV File generiert und heruntergeladen – jene beiden Files enthalten die Daten bzw. Statistiken. Diese Informationen kann der Benutzer oder die Benutzerin jetzt gemeinsam mit dem zuvor generierten PNG an einen Kollegen schicken, um die Daten zu besprechen und zu analysieren.

Nach detaillierter Analyse der Daten bestätigt der Kollege oder die Kollegin, dass der Sensor „Schrack Heizung“ in Ordnung ist und wieder aktiviert werden kann, daher wechselt der Benutzer zum Tab „Daten aktualisieren“ und setzt den Sensor „Schrack Heizung“ auf aktiv (vgl. Abbildung 48).

Abbildung 48: Daten Update - Status ändern

Anschließend wechselt der Benutzer oder die Benutzerin wieder zum „Home“ Tab und stellt sein zuvor gespeichertes Dashboard wieder her, um weitere Analysen durchzuführen (vgl. Abbildung 49).

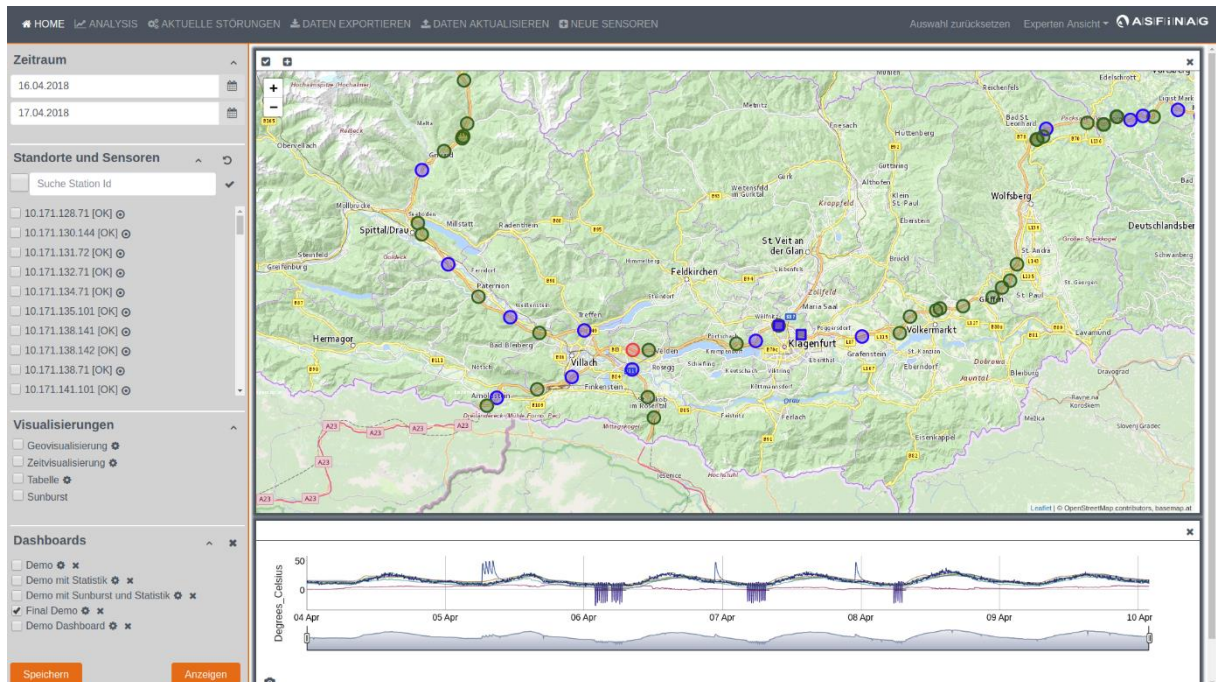


Abbildung 49: Wiederherstellen des zuvor gespeicherten Dashboards

Dafür öffnet der Benutzer oder die Benutzerin die zuvor analysierte Station nahe Velden und eine benachbarte Station (10.171.26.225). Nachdem beide Stationen denselben Sensorwert „Cu Temperature“ haben, wählt der Benutzer oder die Benutzerin beide aus (vgl. Abbildung 50).

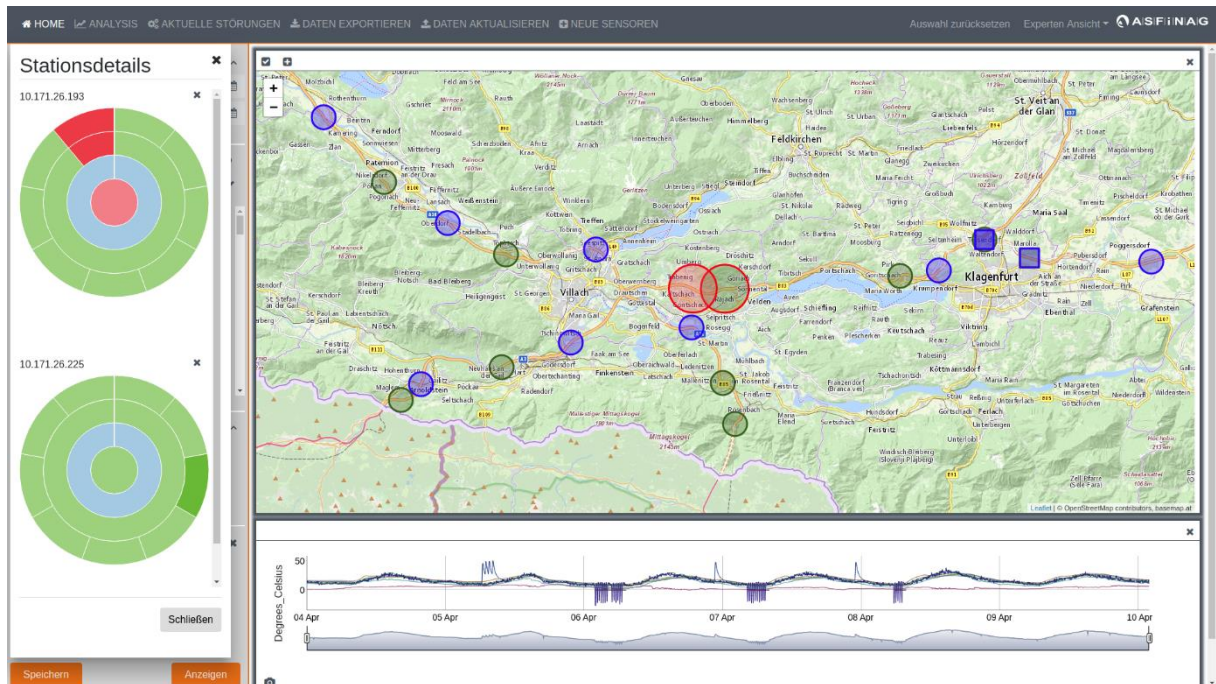


Abbildung 50: Auswahl desselben Sensors (Cu_Temperature) zweier benachbarter Stationen.

Anschließend wählt der Benutzer oder die Benutzerin die Zeitvisualisierung unter Zusammenfügung der Einheiten aus (vgl. Abbildung 51).

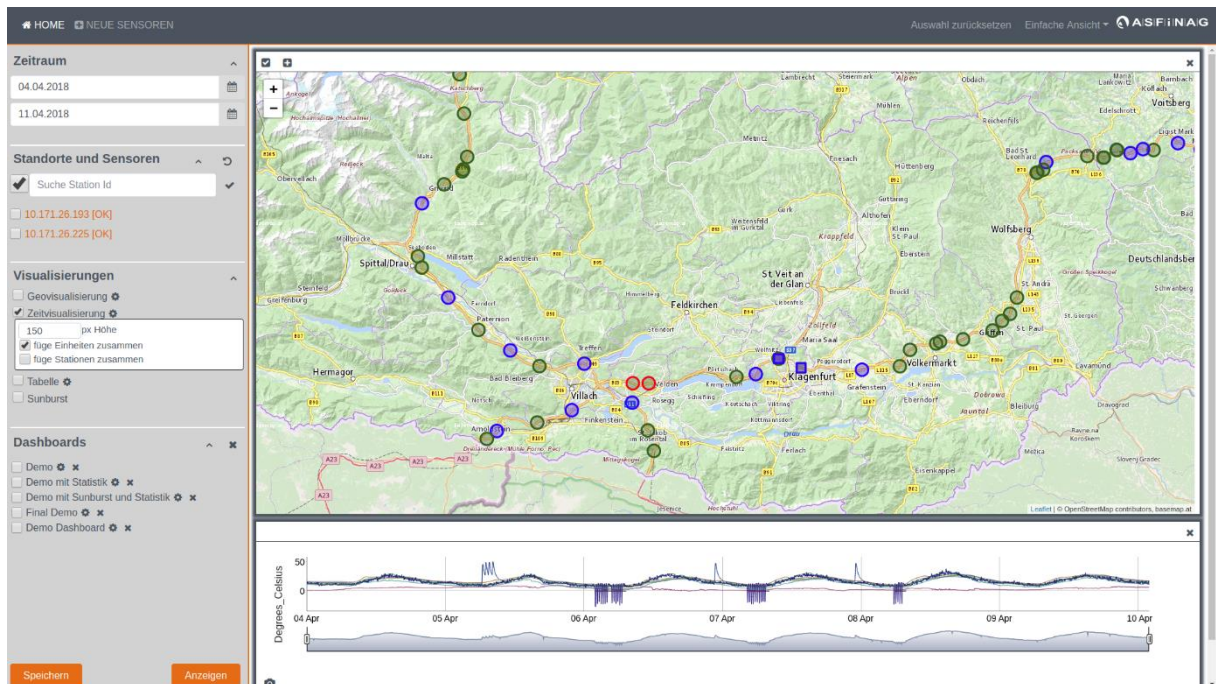


Abbildung 51: Konfiguration um Sensoren von benachbarten Stationen in einer Auswertung gemeinsam anzeigen zu können.

Klickt der Benutzer oder die Benutzerin auf „Anzeigen“ so werden die beiden Sensoren in einem Line Chart dargestellt (vgl. Abbildung 52 und Abbildung 53).

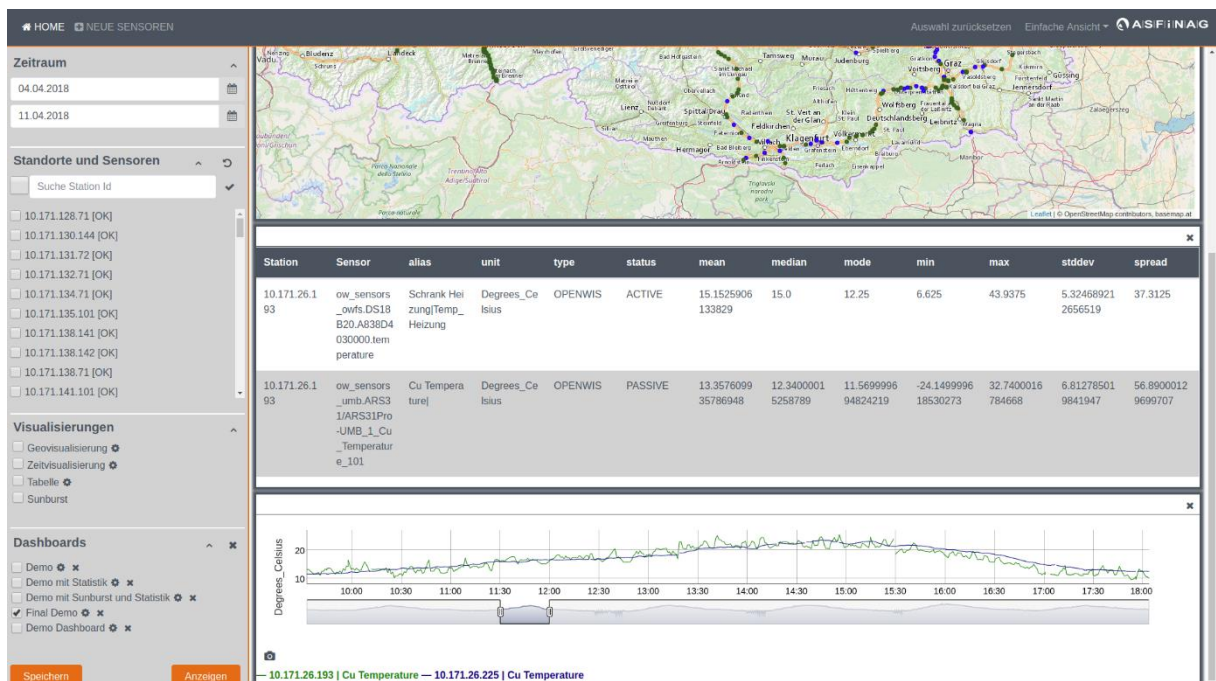


Abbildung 52: Vergleich derselben Sensoren von zwei benachbarten Stationen – zoomen zeigt die größere Fluktuation eines Sensors.

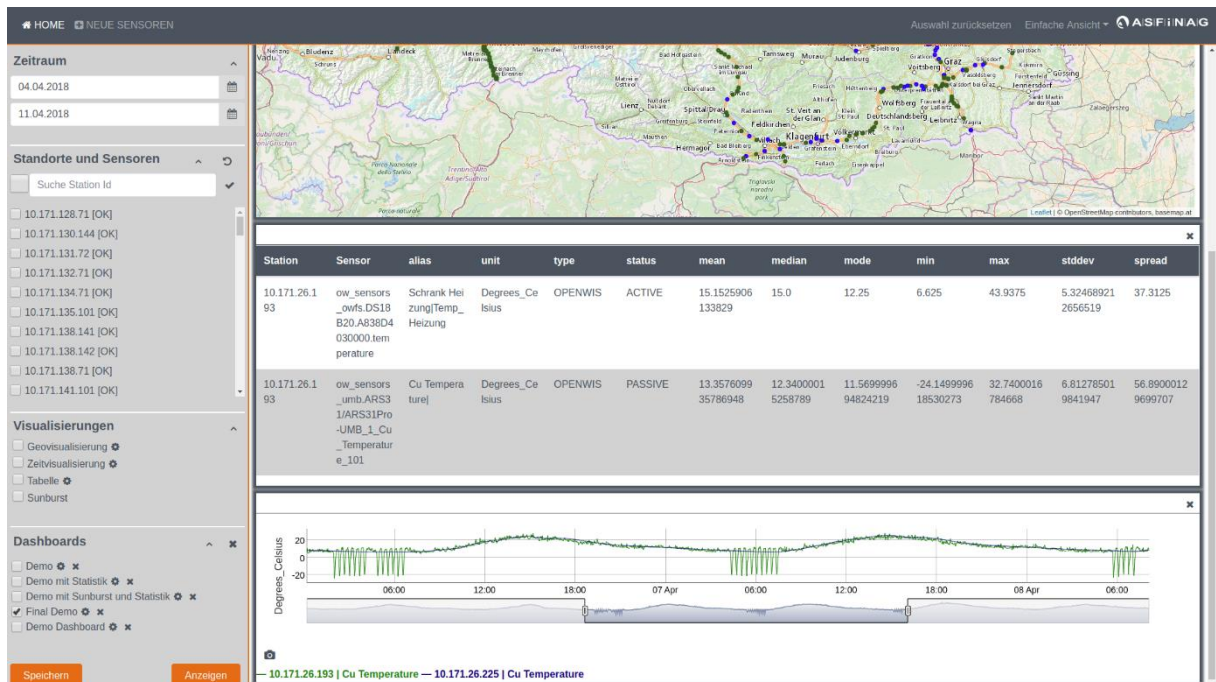


Abbildung 53: Vergleich derselben Sensoren von zwei benachbarten Stationen – zoomen zeigt eine detailliertere Darstellung von Ausreißern.

Der visuelle Vergleich mittels der Zeitvisualisierung ermöglicht folgende Aussagen:

- Im Vergleich zu Station 10.171.26.225 hat der Sensor in Station 10.171.26.193 eine vergleichsweise große Fluktuation. (vgl. Abbildung 52)
- Die zuvor identifizierten Ausreißer kommen nur in Station 10.171.26.193 vor. (vgl. Abbildung 53)

Anschließend wechselt der Benutzer oder die Benutzerin in den Analyse Tab, um die Zeitreihen genauer zu analysieren. Dafür wählt er wieder „Cu Temperature“ der beiden Stationen 10.171.26.193 und 10.171.26.225, um diese in einem Line Chart darzustellen. Hier kann der Benutzer oder die Benutzerin Annotationen definieren: Ctrl + Bereich markieren. (vgl. Abbildung 54)

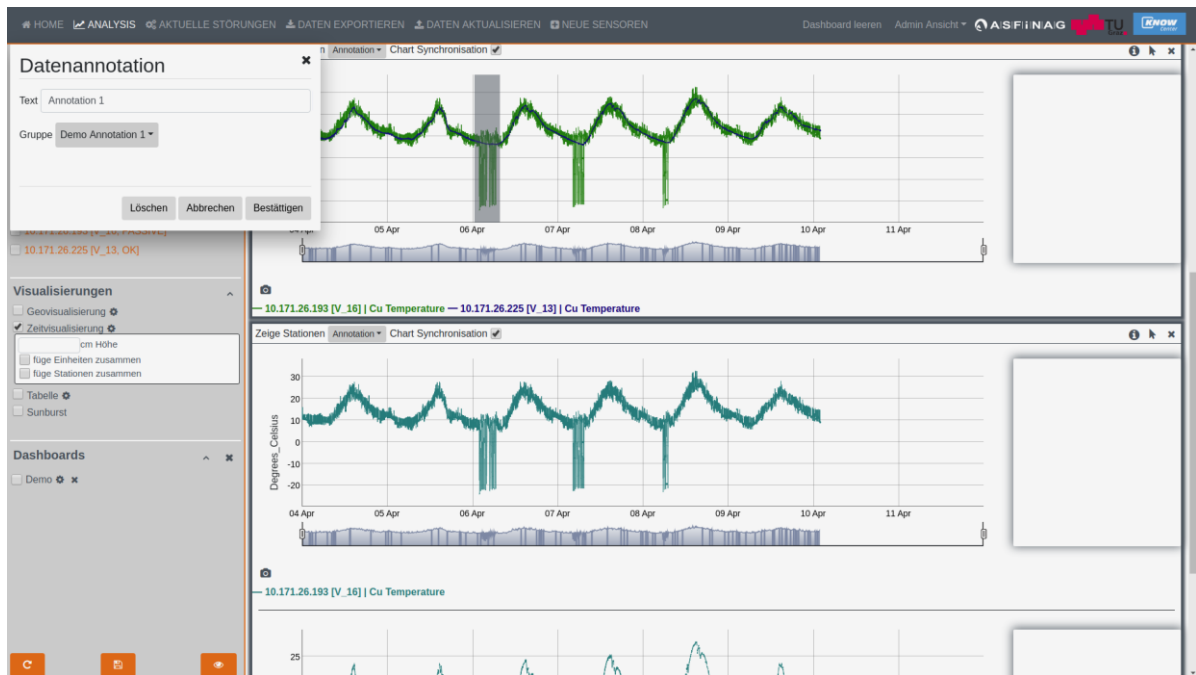


Abbildung 54: Beschreibung (Annotation) eines gewünschten Bereichs.

Das Anlegen der Annotation mit Text und Error Code (je nachdem, ob der Fehler auf Verfügbarkeit bzw. Plausibilität zurückzuführen ist) funktioniert mit Ctrl + Shift auf den markierten Bereich. (vgl. Abbildung 55)

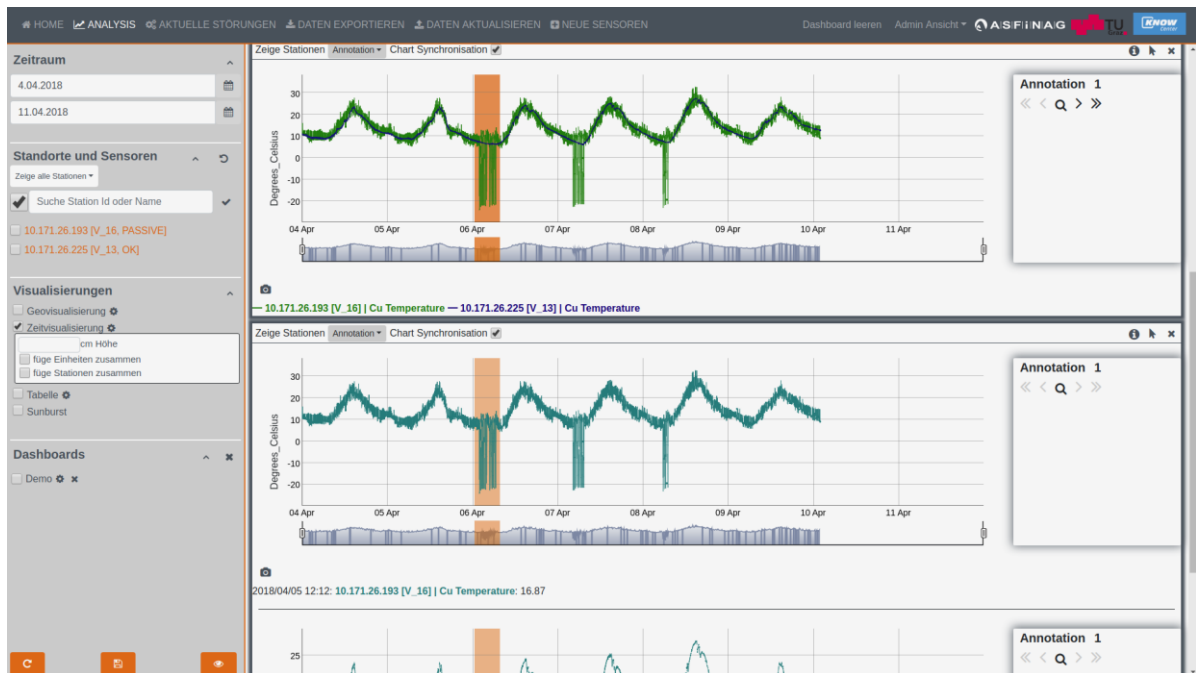


Abbildung 55: Beschreibung (Annotation) eines gewünschten Bereichs.

Im rechten Bereich erscheint ein zusätzliches Fenster, das es den BenutzerInnen ermöglicht, zwischen den Annotationen zu navigieren. Dieses „Springen“ zwischen einzelnen Annotationen erleichtert ein Vergleichen von mehreren Ausreißern.



Abbildung 56: Navigieren zwischen den einzelnen Annotationen erleichtert den Vergleich und die Analyse von Ausreißern.

So können beispielsweise gemeinsame Fehlergruppen bzgl. Verfügbarkeit oder Plausibilität sorgfältiger analysiert werden und somit eine verbesserte Ursachenforschung betrieben werden. Unterstützt werden diese Prozesse durch eben diese benutzerdefinierten Annotationen, die bei einem Export ebenfalls mitexportiert werden.

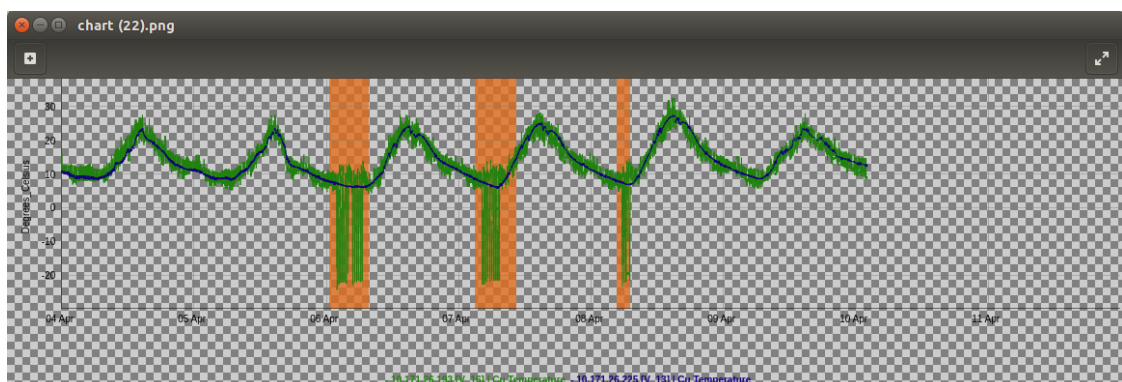


Abbildung 57: Benutzerdefinierte Annotationen werden bei einem Export ebenfalls mitexportiert.

5 ZUSAMMENFASSUNG

Die im Rahmen von Verkehrsbeeinflussungsanlagen (VBA) auf Autobahnen und Schnellstraßen ermittelten Umfelddaten stellen Eingangsgrößen für situationsabhängige automatische Schaltvorschläge dar. Mittels dynamischen Warnhinweisen, Geschwindigkeitsbeschränkungen oder Spurzuweisungen werden die Verkehrsteilnehmer vor potenziell gefährlichen Umfeldbedingungen (wie z.B. Nässe und Schneefall) gewarnt. Diesen negativen Einflüssen wird durch die Anzeige einer angepassten zulässigen Höchstgeschwindigkeit und ggf. von Warnhinweisen mittels Wechselverkehrszeichen (WVZ) begegnet. Witterungsbedingte Schaltungen können nur dann einen positiven Einfluss auf die Verkehrssicherheit erzielen, wenn die Anzeigen der VBA durch die Verkehrsteilnehmer befolgt werden. Hierfür ist es wiederum erforderlich, dass die stationär erfassten Umfelddaten eine hohe Qualität aufweisen und die aktuellen Umfeldbedingungen zeitnah im System abgebildet werden. Nur so kann in Abhängigkeit der vorliegenden Verkehrs- und Umfeldsituationen eine sinnvolle und wirksame Beeinflussung des Verkehrsablaufs stattfinden.

Aktuell wird im Streckennetz der ASFiNAG die Qualität der Sensoren zur Umfelddatenerfassung weder systematisch noch vergleichend untersucht und Fehler häufig nicht oder erst spät bzw. zufällig erkannt. Aufgrund der Wichtigkeit der Umfelddatenerfassung und des für die Praxis zu erwartenden Nutzens bestand das Ziel von UDEQI einerseits darin, eine Verfügbarkeitsaussage pro Sensor und für alle Sensoren gemeinsam zu haben, sowie andererseits eine Performance-Kennzahl pro Sensor zu entwickeln. Dabei wurde eine Plausibilisierung zwischen den unterschiedlichen Sensoren sowie ein Längsabgleich (entlang der Strecke, zwei benachbarte Stationen, Datenfusion) angestrebt. Dies wurde in einem lauffähigen, performanten Prototyp umgesetzt, in dem die Werte der einzelnen Sensoren (je nach Sensortyp + Verfügbarkeitswerte) sowie deren Qualität (Plausibilisierung) bewertet werden. Die Daten werden in Echtzeit abgerufen und bewertet, fehlerhafte oder auffällige Sensoren werden anschließend in einer eigenen Auswertung ausgegeben. Zusätzlich werden die ermittelten Minutendaten in einem Monitoringtool und die userspezifischen Auswertungen auf einem Graphical User Interface (GUI) angezeigt (Langzeitdaten, Verfügbarkeit und Performance).

Die Schwierigkeiten des Forschungsvorhabens bestanden v.a. darin, dass sowohl in den geltenden nationalen als auch in internationalen Richtlinien bisher keine Methoden zur Prüfung der Genauigkeit von Umfelddaten definiert sind.

Den Input von UDEQI stellen die bestehende Sensorinfrastruktur sowie schon verfügbare Sensordatenbanken der ASFiNAG dar. Während die Datenbanken nur einmal in das UDEQI-System importiert wurden, stellt der Sensorinput einen kontinuierlichen Prozess dar. Der oder die BenutzerIn interagiert mit der UDEQI-Web-GUI und wird dadurch in den täglichen Arbeitsabläufen unterstützt. Wesentliche Operationen, wie beispielsweise schnell einen Überblick über den aktuellen Zustand des Sensornetzwerks zu bekommen, werden zur Verfügung gestellt. Zudem besteht die Möglichkeit, Auswertungen für spätere Analysen, Diskussionen oder auch Berichte zu erstellen. Dazu wird nicht direkt auf die Datenbank zugegriffen, sondern ausschließlich auf das UDEQI-Analysemodul. Dieses stellt das zentrale Modul dar und ist verantwortlich für die Implementation der heuristischen, statistischen und maschinellen Lernverfahren, die Aussagen über die Verfügbarkeit sowie die Plausibilität der Sensorik zulassen. Zudem verwaltet diese Komponente den Zugriff auf die Datenbank, sodass die Datenbankzugriffe eingeschränkt bzw. mit einer Rechteverwaltung hinterlegt werden können. Erfahrungsgemäß wiederholen sich gleiche oder sehr ähnliche Zugriffsmuster häufig. Daher ist eine Caching-Komponente vorgesehen, in der häufig wiederkehrende Auswertungen zwischengespeichert werden können.

Die Softwareentwicklung erfolgte iterativ und agil, basierend auf tagtäglichen Anwendungsszenarien der ASFiNAG. Bereits zu Beginn des Datenflusses werden die Algorithmen zur Bestimmung der Verfügbarkeit angewendet, d.h. beim Verarbeiten und Einlesen der Sensordaten in die InfluxDB, der ausgewählten Zeitreihendatenbank. Die Heuristiken zur Erkennung von Verfügbarkeit führen dann zu Warnmeldungen, die weiterführende Aktionen triggern.

Während im Allgemeinen die Outlier-Detection einen Vorverarbeitungsschritt darstellt, ist es in diesem Projekt eine der zentralen Forschungsfragen. Im Kontext des Projektes wurden jeweils jene Verfahren ausgesucht, die sich einerseits auf die bestehenden Zeitserien und verfügbaren Daten anpassen lassen sowie jene Verfahren, die auch in einem Online-Kontext eingesetzt werden können. Sind alle einzelnen Messwerte auf ihre Plausibilität hin kontrolliert, kann eine logisch/physikalische Überprüfung folgen. Da einige Messgrößen voneinander meteorologisch abhängig sind, können nicht plausible Messwerte durch Vergleiche untereinander identifiziert werden. Die Langzeitprüfungen liefern Hinweise auf systematische Fehler. Für eine Langzeit-Plausibilitätsprüfung werden Ergebnisse der jeweils gleichen Sensortypen einer Messstelle mit denen der beiden Nachbarstellen verglichen. Des Weiteren werden Messwerte gleicher Messgrößen aber unterschiedlicher Sensoren korreliert und der Korrelationskoeffizient als dimensionsloses Maß für den Grad des linearen Zusammenhangs

berechnet. Erkannte Ausreißer führen zu einer Kennzeichnung in der Datenbank. Diese Kennzeichnungen hinsichtlich unzureichender Plausibilität zeichnen schlussendlich für die Darstellung in der graphischen Benutzeroberfläche verantwortlich; d.h. im Fall von UDEQI werden die Stationen mit den zugehörigen Fehlercodes versehen und entsprechend eingefärbt.

Dies ermöglicht es den Verantwortlichen der ASFiNAG die aufgetretenen Probleme weiter zu beobachten und Prozesse zur Sicherstellung der Datenqualität durch geeignete Sensorwartung voranzutreiben oder Prozesse durch Erfahrungen mit einzelnen Sensoren hinsichtlich Kosteneffizienz zu adaptieren.

6 AUSBLICK

Aufgrund des Projektziels einer prototypischen Umsetzung eines qualitätssichernden und plausibilisierenden Systems für Sensormessdaten wurden zwar eine Vielzahl von Meinungen, Erfahrungen, Wünschen und Anregungen zusammengetragen, jedoch konnten nicht alle Ideen im Prototypen umgesetzt werden. In einem zusätzlichen Schritt wurden ähnliche Funktionalitäten zusammengefasst, kategorisiert und auf Umsetzbarkeit geprüft (vgl. Abbildung 58). Wenn eine Umsetzung der jeweiligen Funktionalität im Kontext des Forschungsvorhabens hinsichtlich des zeitlichen und finanziellen Budgets möglich war, wurde diese Funktionalität zusätzlich zur Grundfunktionalität im Prototypen umgesetzt. Im Falle von nicht umgesetzten Funktionalitäten wurde die Grundlage für eine zukünftige, potentielle Entwicklung in einem Umsetzungsprojekt für ein Produktivsystem auf Basis des UDEQI-Prototypen vorgesehen.

Funktionalität	Vorgesehen für die Zukunft	Im Projekt UDEQI	Beschreibung/Detailgrad
Stammdaten	Ja	nein	Bearbeitung von Stammdaten einzubauen braucht wieder Zeit am Client + am Server → Nachfolgeprojekt
Statistiken	Ja	Ja	einfache Statistiken zu Sensoren/Sensoren implementiert; keine Filterung bzgl. Hersteller, etc. → Nachfolgeprojekt
Export	Ja	Ja	eigener Exportreiter
Referenztestfeld	Ja	nein	wäre dann ein eigener Reiter mit den Stationen, die das Referenztestfeld umfassen
BüS Tickets	Ja	nein	Anbindung an BüS System Asfinag notwendig → Nachfolgeprojekt
Mobile Stationen	Ja	nein	wäre dann ein eigener Reiter mit den Mobilien Stationen
Annotation von Daten	Ja	exemplarisch	Feedbackmechanismen für Asfinag Mitarbeiter bei LineCharts
Regeln	Ja	exemplarisch	Zusammenhang zwischen WFD und NI wird bei ausgewählten Sensoren aufgezeigt; ev. sogar externes Dokument und nicht in der GUI → etwaige Implementierung im Nachfolgeprojekt
Qualitätssicherung	Ja	Exemplarisch	Betrifft die Flags aus der letzten Telko; vorgesehen ist es in Udeqi; ein wirkliches Bereitstellen, um diese Daten (meteorologisch ok, GMA ok, etc) in die DB eintragen zu können, geht ins Nachfolgeprojekt

Abbildung 58: Übersicht vorgeschlagene Funktionalitäten und deren Status innerhalb UDEQI

In einem größeren Kontext wäre ein Umsetzungsprojekt für ein Produktivsystem auf Basis des UDEQI-Prototypen dahingehend für die ASFiNAG interessant, als dass dieses eine Schnittstellenfunktion für unterschiedliche bestehende Sensoriksysteme und Wartungsprozesse der Sensorik übernehmen könnte (vgl. Abbildung 59). Diese Lösung würde im Vergleich zur gegenwärtigen Situation, z.B. unterschiedliche Versionen von Wartungslisten oder allgemein gänzlich fehlenden Informationen durch einen zentralen skalierbaren Zugang per Website unterbinden und damit zu einer Vereinfachung in der Abwicklung von Wartungsprozessen beitragen.

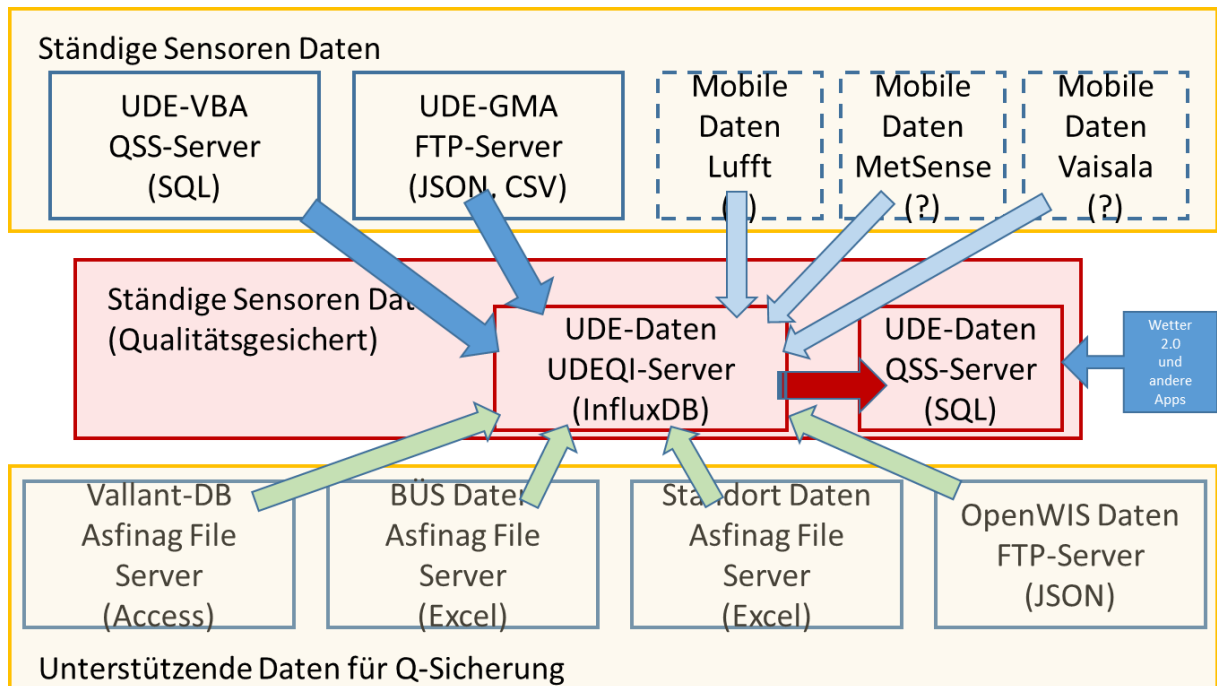


Abbildung 59: Mögliche Schnittstellenfunktion eines zukünftigen UDEQI-Produktsystems

7 REFERENZEN

- [1] BASt (2017). Anlagen zur Verkehrsbeeinflussung auf Bundesfernstraßen, https://www.bast.de/BASt_2017/DE/Verkehrstechnik/Fachthemen/v5-verkehrsbeeinflussungsanlagen.html?nn=1817946, abgerufen am 19.09.2018 um 12:15 Uhr
- [2] ASFiNAG (2018). Verkehrsbeeinflussungsanlagen – Für mehr Sicherheit, <https://www.ASFInAG.at/verkehrssicherheit/verkehrsmanagement/verkehrssteuerung/>, abgerufen am 19.09.2018 um 12:17 Uhr
- [3] Rascher, A.; Grošanić, S.; Busch, F.: Umfelddatenerfassung in Streckenbeeinflussungsanlagen, Testfeld „Eching Ost“ des Bundes, Abschlussbericht 9. Testphase, Herausgegeben vom Bundesministerium für Verkehr, Bau und Stadtentwicklung, 2014
- [4] MARZ: Merkblatt für die Ausstattung von Verkehrsrechnerzentralen und Unterzentralen, Herausgegeben von der Bundesanstalt für Straßenwesen (BASt), Bergisch Gladbach, 1999
- [5] TLS: Technische Lieferbedingungen für Streckenstationen, Herausgegeben von der Bundesanstalt für Straßenwesen (BASt), Bergisch Gladbach, 2012
- [6] FGSV: Sensor-Anschlussmöglichkeiten im Testfeld des AK 3.2.1 Umfelddatenerfassung in VBA, FGSV, AK 3.2.1, 2011
- [7] FGSV: Merkblatt für Nässeerkennung in Streckenbeeinflussungsanlagen, 2002
- [8] Scharsching, H.: Witterungsbedingungen – Verkehrsverhalten – Verkehrssteuerung, Bundesministerium für wirtschaftliche Angelegenheiten / Straßenforschung Heft 363, Wien, 1998
- [9] Mangold, M., Träger, K., Lindenbach, A.: Wirksamkeit von Streckenbeeinflussungsanlagen unter besonderer Berücksichtigung der Umfelddatenerfassung, Forschung Straßenbau und Straßenverkehrstechnik Heft 729, Bonn, 1996
- [10] FGSV: Hinweise zur Erfassung und Nutzung von Umfelddaten in Streckenbeeinflussungsanlagen, W1, Köln, 2010

- [11] Petraczek, O.: Experimentelle Ermittlung der Wasserfilmdicken auf Fahrbahnen unter natürlichen Bedingungen, München, Aus: Forschung Straßenbau und Straßenverkehrstechnik, 1975
- [12] Lang, J.: Analyse und Simulation des Feuchtekontinuums auf Straßenoberflächen, Technische Berichte über Ingenieurhydrologie, Institut für Wasserbau an der Technischen Universität Darmstadt, Darmstadt, 1991
- [13] ASFiNAG – BMG (2016). Planungshandbuch der ASFiNAG - Technische Infrastruktur Freiland - OpenWIS Dokumentation
- [14] Zhang, Y., Meratnia, N., and Havinga, P. (2010). Outlier detection techniques for wireless sensor networks: A survey. *Commun. Surveys Tuts.*, 12(2):159–170.
- [15] Sharma, A. B., Golubchik, L., and Govindan, R. (2010). Sensor faults: Detection methods and prevalence in real-world datasets. *ACM Trans. Sen.Netw.*, 6(3):23:1–23:39.
- [16] Alameddine, I., Kenney, A., Gosnell, J., and Reckhow, H. (2010a). Robust multivariate outlier detection methods for environmental data. *Journal of Environmental Engineering*, 136(11):1299–1304.
- [17] Dereszynski, E. W. and Dietterich, T. G. (2011). Spatiotemporal models for data-anomaly detection in dynamic environmental monitoring campaigns. *ACM Trans. Sen. Netw.*, 8(1):3:1–3:36.
- [18] Garces, H. and Sbarbaro, D. (2011). Outliers detection in environmental monitoring databases. *Eng. Appl. Artif. Intell.*, 24(2):341–349.
- [19] Gupchup, J., Sharma, A., Terzis, A., Burns, A., and Szalay, A. (2008). The perils of detecting measurement faults in environmental monitoring networks. In *DCOSS*.
- [20] Janakiram, D., Reddy, V. A., and Kumar, A. V. U. P. (2006). Outlier detection in wireless sensor networks using bayesian belief networks. In *2006 1st International Conference on Communication Systems Software Middleware*, pages 1–6.

- [21] Ramanathan, N., Balzano, L., Burt, M., Estrin, D., Harmon, T., Harvey, C., Jay, J., Kohler, E., Rothenberg, S., and Srivastava, M. (2006). Rapid Deployment with Confidence: Calibration and Fault Detection in Environmental Sensor Networks. Technical report.
- [22] Rassam, M. A., Zainal, A., and Maarof, M. A. (2012). One-class principal component classifier for anomaly detection in wireless sensor network. In 2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN), pages 271–276.
- [23] Shahid, N., Naqvi, I. H., and Qaisar, S. B. (2012). Quarter-sphere svm: Attribute and spatio-temporal correlations based outlier & event detection in wireless sensor networks. In WCNC, pages 2048–2053. IEEE.
- [24] Shahid, N., Naqvi, I. H., and Qaisar, S. B. (2013). One-class support vector machines: analysis of outlier detection for wireless sensor networks in harsh environments. Artificial Intelligence Review, 43:515–563.
- [25] Siripanadorn, S., Hattagam, W., and Teaumroong, N. (2010). Anomaly detection using self-organizing map and wavelets in wireless sensor networks. In Proceedings of the 10th WSEAS International Conference on Applied Computer Science, ACS’10, pages 291–297, Stevens Point, Wisconsin, USA. World Scientific and Engineering Academy and Society (WSEAS).