

Star-Plot-Karten für die visuelle Untersuchung von mobilen Sensordatensätzen

Star Plot Maps for a Visual Investigation of Mobile Sensor Datasets

Christian Röger

Universität Augsburg · christian.roeger@geo.uni-augsburg.de

Zusammenfassung: Im Bereich der Messsysteme zur Aufnahme von Umweltparametern rücken mobile Lösungen immer weiter in den Fokus. Dabei werden häufig große, heterogene Datensätze mit vielen Datenfeldern generiert. Da die Darstellung multivariater Datensätze Wissenschaftler häufig vor Herausforderungen stellt, testet diese Arbeit den Nutzen von Star-Plot-Karten zur Visualisierung mobiler Sensordatensätze. In einer Fallstudie auf dem Campus der Universität Augsburg wurden Datensätze generiert, die aufgenommenen Daten prozessiert und am Ende in einer Star-Plot-Karte des Testgebietes visualisiert. Die Ergebnisse zeigen, dass Star-Plot-Karten bei einer passenden Datengrundlage eine gute Möglichkeit darstellen, multivariate Sensordatensätze zu visualisieren.

Schlüsselwörter: Visual Geo Data Mining (VGDM), Star-Plot-Karten (SPM), Star-Plots, Visualisierung, Sensordaten

Abstract: *Mobile solutions for measuring environmental parameters gain popularity. Resulting datasets tend to be big and heterogeneous, which challenges researchers when investigating them. We state that using star plot maps can be a solution for visualizing multi-variate sensor datasets. In our case study, we generate a dataset on the campus of the University of Augsburg, Germany. After processing the data, we create a star plot map of the study area showing multiple parameters. Findings indicate a good suitability of star plot maps for visualizing multi-variate sensor datasets.*

Keywords: *Visual Geo Data Mining (VGDM), Star Plot Maps (SPM), Star Plots, Visualization, Sensors*

1 Über die Relevanz mobiler Messsysteme zur Luftqualitätsbestimmung

Mobile Messsysteme zur Erstellung von Sensordatensätzen haben in den vergangenen Jahren stark an Bedeutung gewonnen (Hart & Martinez, 2006). Die Visualisierung solcher Daten ist eine Herausforderung, denn häufig müssen viele Variablen gleichzeitig dargestellt werden, um eine ganzheitliche Analyse solcher Datensätze möglich zu machen.

Es gibt einige Möglichkeiten, multivariate Datensätze zu visualisieren. Diese Arbeit evaluiert die Nutzung von Star-Plot-Karten, um die Qualität eines selbst erstellten Sensordatensatzes zu beschreiben.

Um Phänomene wie Veränderungen im Straßenverkehr (zum Beispiel mehr Autos oder Staus) sowie die daraus resultierende Veränderung der Schadstoffkonzentration in der Luft zu erforschen, sind Datensätze notwendig. Beispiele hierfür sind Floating Car Daten (Fabriiis et al., 2008) und Sensordaten (Hart & Martinez, 2006). Dabei spielt die Art des verwendeten Messsystems eine genauso große Rolle wie die Wahl der für die jeweilige Analyse

benötigten Messsensorik. Gegenüber herkömmlichen, stationären Messsystemen bieten mobile Lösungen Vorteile, wie kein fest installiertes Equipment, was in der Regel zu geringeren Wartungs- sowie Installationskosten führt (Fabritiis et al., 2008).

Es finden sich viele Beispiele für mobile Lösungen zur Verkehrsdatenanalyse und Umweltmonitoring. Krampe et al. (2013) nutzen die Methode der erweiterten Floating-Car-Daten (xFCD), um in der Modellregion Salzburg mithilfe eines mobilen Sensornetzes Verkehrsanalysen zu betreiben. Dutta et al. (2009) stellen ein auf Citizen Science (Bundesministerium für Bildung und Forschung, 2019) basierendes Konzept vor, mit dem Anwender über ein mobiles Messgerät Daten aufnehmen und mithilfe einer Webplattform analysieren können. Das Portal Luftdaten.info (<https://luftdaten.info>) stellt Anwendern eine Plattform bereit, selbst aufgezeichnete Umweltdaten mit anderen Nutzern auf Open-Source-Basis zu teilen. Das Projekt SmartAQnet setzt sich zum Ziel, in der Modellregion Augsburg flächendeckende, aber möglicherweise ungenaue low-cost Sensor Daten mit genauen stationären Messdaten anzureichern, um die Vorteile beider Messarten zu kombinieren (Budde et al., 2017).

Mobile Sensornetzwerke stehen einigen Herausforderungen gegenüber. Nach Hart & Martinez (2006) muss auf folgende Faktoren geachtet werden:

- Die Stromversorgung,
- die Benutzerfreundlichkeit,
- die Standardisierung der Daten,
- die Datenqualität,
- die Sicherheit,
- die Methodik zur Datenverarbeitung und -interpretation,
- sowie die Entwicklung neuer Sensorik.

Diese Fallstudie adressiert die Entwicklung von Methodik zur Datenanalyse. Viele der in der Literatur beschriebenen Ansätze zur Implementierung eines mobilen Sensorsystems stützen sich auf dreidimensionale Darstellungen (Geographische Breite, Geographische Länge, Sensor-Wert). Die vorliegende Arbeit leistet ihren Beitrag zum Diskurs, indem sie Darstellungsmethoden für multivariate Datensätze nutzt. Dafür wird zunächst die Visualisierungsmethode der Star-Plot-Karten beleuchtet. In einer anschließenden Fallstudie wird mithilfe eines mobilen Messsystems ein Datensatz erstellt. Die daraus resultierenden Daten werden im Anschluss prozessiert, um Qualitätsfaktoren des Datensatzes zu bestimmen. Abschließend wird eine Star-Plot-Karte erstellt, analysiert und diskutiert.

2 Star-Plot-Karten für die Visualisierung multivariater Datensätze

Das Visualisieren mehrdimensionaler Datensätze stellt eine weit verbreitete Fragestellung dar. Ein- sowie zweidimensionale Datensätze können durch einfache Diagramme beziehungsweise Scatterplots dargestellt werden. Eine dritte Dimension kann zusätzlich über eine Farb- abstufung gezeigt werden. Die Darstellung vier- oder höherdimensionaler Datensätze erfordert hingegen die Benutzung komplexerer Plotting-Symbole (Chambers et al., 1983).

Chambers et al. (1983) nennen als Beispiel die Nutzung von Wetterkarten-Symbolen, die sowohl die aktuelle Wolkenbedeckung (durch Schattierung des Symbols) als auch die Windrichtung (durch die Ausrichtung des Symbols) und die Windgeschwindigkeit (durch die Anzahl und Art von horizontalen Linien auf dem Symbol) zeigen (Forkel 2012). Weitere Möglichkeiten stellen die Verwendung von parallelen Koordinaten (PCP) dar (Inselberg, 1985; Hauser et al., 2002) dar. Lind et al. (2009) schlagen darüber hinaus vor, PCPs in einer n:n Beziehung darzustellen, um weiterführende relationale Beziehungen zu zeigen. Die Hyperbox nach Alpern et al. (1991) ist eine weitere Möglichkeit zur Visualisierung von multivariaten Datensätzen. Als eine weitere nützliche Darstellungsmethode bringen Chambers et al. (1983) die Verwendung von Star-Plots ins Spiel. Hierbei handelt es sich um Symbole, die mehrere Werte anhand von Sternsegmenten zeigen, die um einen Mittelpunkt angeordnet sind.

Nach Chambers et al. (1983) geschieht die Generierung von Star-Plots nach folgendem Schema:

- Die Daten müssen positive Werte aufweisen. Die Werte sollen auf ein Intervall von c zu l skaliert werden, wobei c den Abstand vom kürzesten zum längsten Stern-Wert zum längsten definiert.
- Die Sternsegmente werden gleichverteilt um einen Punkt angeordnet.
- Die jeweils höchsten Werte eines Sterns werden miteinander verbunden.

Damit können die in Zeilen und Reihen angeordneten Sterne visuell miteinander verglichen werden (Chambers et al., 1983). Ansätze, die Methode der Star-Plots zu verbessern, finden sich in der Literatur. Klippel et al. (2009) diskutieren über die Anordnung der Sternsegmente, um individuelle Sterne besser miteinander vergleichen und unterscheiden zu können. Eine weitere Arbeit von Klippel et al. (2009b) spricht die Einfärbung von Sternsegmenten an, um den Einfluss von stark differenzierten Segment-Radien einzudämmen für eine bessere Interpretierbarkeit der Plots.

3 Visualisierung von Qualitätsfaktoren eines multivariaten Datensatzes

Es wird eine Fallstudie an der Universität Augsburg durchgeführt. Als Ziel des Vorhabens soll eine Star-Plot-Karte des Testgebietes erstellt werden, die Qualitätsparameter einer Messung zeigt. Als Maß dafür soll die Standardabweichung einer Messung innerhalb eines Gebietes dienen. Es wird davon ausgegangen, dass diese kleinräumig nicht signifikant streuen sollte. Abbildung 1 gibt Aufschluss über den Ablauf der Fallstudie. Die Arbeitsschritte sind in drei Phasen aufgeteilt.

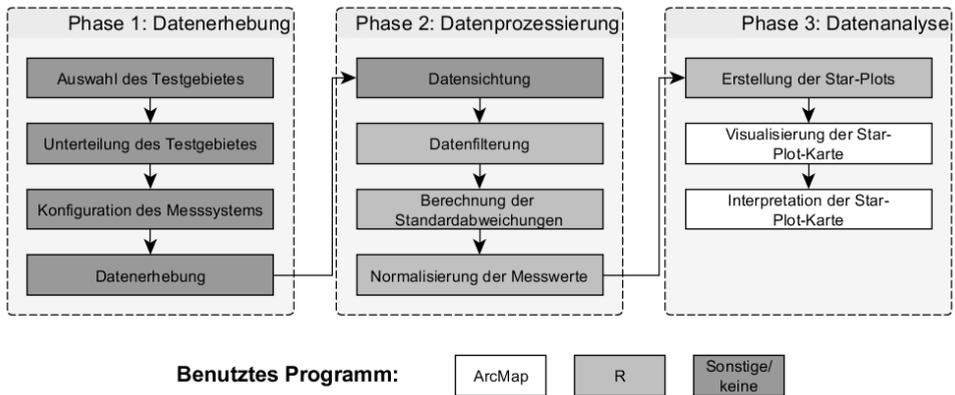


Abb. 1: Drei Phasen der Fallstudie: Datenerhebung, Datenprozessierung und Datenanalyse

Abbildung 1 zeigt den Arbeitsablauf in drei Phasen. In Phase 1 wird zunächst das Testgebiet definiert und in Regionen unterteilt. Die Mittelpunkte dieser Regionen dienen später als Position der Star-Plots. Anschließend wird das benutzte Messsystem ausgewählt. Die darauffolgende Erhebung resultiert in einem Datensatz, der im Anschluss in Phase 2 verarbeitet wird. Dafür werden zunächst die erhobenen Daten gesichtet. In der anschließenden Filterung werden offensichtliche Extremwerte ausgeschlossen. Danach werden die Standardabweichungen der Messwerte berechnet. Um die Daten sinngemäß darzustellen, müssen die Parameter normalisiert werden. Die darauffolgende dritte Phase besteht aus der Erstellung der Star Plots, der Visualisierung der Star-Plot-Karte sowie der Interpretation der Analyse-Ergebnisse.

3.1 Erhebung der Daten

Als Testgebiet dient der Campus der Universität Augsburg. Star-Plot-Karten erfordern die Präsenz einer Untergliederung des Gesamtgebietes. Deshalb wird über dem ausgewählten Areal ein Grid (vier Spalten, drei Zeilen) aufgespannt. Die Datenaufnahme erfolgt mit einem selbst gebauten mobilen Messsystem. Das Arduino-basierte Setup wird schematisch in Abbildung 2 dargestellt.

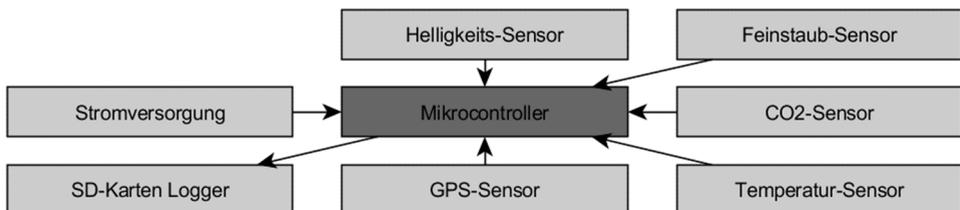


Abb. 2: Schematische Darstellung eines Sensorsystems zur mobilen Aufnahme von Umweltdaten

Nach Abbildung 2 besteht das System aus mehreren Einheiten, die unterschiedlichen Funktionen zugeordnet sind. Der Mikrocontroller (Arduino Nano, Vita 2014) sorgt für die Steuerung und das Auslesen der angeschlossenen Geräte. Als GPS-Sensor dient das neo-6m System von u-blox (u-blox 2011). Das Sensor Array zum Bestimmen von Umweltdaten besteht aus dem CO₂-Sensor mh-z19b (Winsen n.d.), dem Helligkeits-Sensor BH1750 (ROHM, 2011), dem Feinstaub-Messgerät SDS021 (Nova Fitness Co. n. d.) sowie dem Temperatursensor DHT22 (Liu n.d.). Im regelmäßigen Abstand von drei Sekunden werden die Daten der Umwelt-Sensorik sowie des Positionierungssystems ausgelesen und über einen SD-Karten-Logger in eine csv-formatierte Datei geschrieben. Tabelle 1 gibt einen Überblick über alle aufgenommenen Parameter.

Tabelle 1: Aufgenommene Parameter der mobilen Sensorbox

Parameter	Erklärung	Wert
id	Eindeutige ID eines Datenpunktes	1
date	Datum	010219
time	Zeit	145322
lat	Geographische Breite in Dezimalgrad	48.336093
lon	Geographische Länge in Dezimalgrad	10.898552
alt	Höhe in Metern	498.0
satellites	Anzahl der sichtbaren Satelliten	9
hdop	Horizontale Dilution of Precision	1.1
co2	Kohlenstoffdioxid in Parts per Million (ppm)	580
lux	Helligkeit in lux	291
pm10	Feinstaubpartikel ($\varnothing < 10 \mu\text{m}$) in $\mu\text{g pro m}^3$ Luft	14
pm25	Feinstaubpartikel ($\varnothing < 2.5 \mu\text{m}$) in $\mu\text{g pro m}^3$ Luft	8
temp	Temperatur in Grad Celsius	3.2
hum	Luftfeuchtigkeit in Prozent	82.9

Tabelle 1 zeigt die 14 aufgezeichneten Parameter. Die ID lässt jeden Datenpunkt eindeutig zuordnen. Während die Felder 'date' und 'time' den genauen Zeitpunkt der Messung aufzeigen, verorten die Parameter 'lat', 'lon' und 'alt' einen Messpunkt im Raum. 'satellites' und 'hdop' beschreiben die Genauigkeit der Positionsbestimmung über die Anzahl der zum Messzeitpunkt sichtbaren Satelliten sowie die horizontale Streubreite der Messwerte (ESRI n.d.). Zusätzlich zu den raumzeitlichen Informationen werden auch Sensormesswerte aufgezeichnet. Das Feld 'co2' gibt den Kohlenstoffdioxidgehalt der Luft an in parts per million (ppm). 'lux' ist ein Maß für die Umgebungshelligkeit rund um das Messsystem. Die Felder 'pm10' sowie 'pm25' geben Aufschluss über den Feinstaubgehalt der Luft aufgeteilt in Partikel mit einem Durchmesser kleiner als 10 μm (pm10) sowie kleiner als 2.5 μm (pm2.5). Komplettiert wird der Datensatz durch Informationen über die aktuelle Temperatur ('temp') in Grad Celsius und die aktuelle Luftfeuchtigkeit ('hum') in Prozent.

Mithilfe des beschriebenen Setups wird ein Datensatz erstellt. Eine Begehung des Testgebietes von 120 Minuten Dauer resultiert in einem Datensatz aus 1108 aufgenommenen Datenpunkten. Die Route ist so gewählt, dass die Punkte am Ende möglichst gut über das gesamte Gebiet verteilt sind. Abbildung 3 zeigt das Testgebiet, die abgelaufene Trajektorie sowie die räumliche Unterteilung des Gebietes.

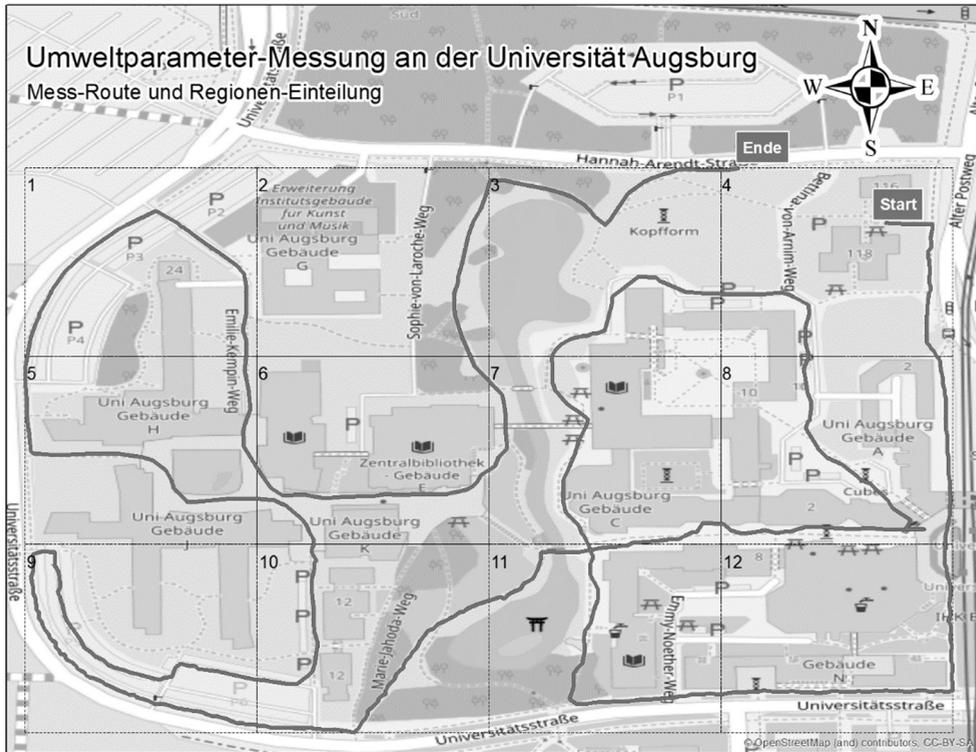


Abb. 3: Karte mit abgelaufener Trajektorie sowie räumlicher Unterteilung des Testgebietes

3.2 Vorbereitung der Daten

Punkte, die durch Positionierungsfehler deutlich außerhalb des Gebietes liegen, werden zunächst ausgefiltert. Die Star-Plots sollen am Ende einen Aufschluss über verschiedene Qualitätsmerkmale geben. Als qualitative Kriterien werden in dieser Arbeit Positionierungsfaktoren (hdop, Anzahl der sichtbaren Satelliten), die Verteilung der Punkte (Anzahl der Datenpunkte pro Region) sowie die Streuung der Messwerte der Umweltsensoren (Standardabweichungen der CO₂-, PM₁₀-, PM_{2.5}-, Helligkeits-, Temperatur- und Luftfeuchtigkeitswerte) ausgewählt. Dabei soll gelten, dass 'gute' Werte einen möglichst kleinen Ausschlag im Star-Plot bewirken. Abbildung 4 zeigt den Aufbau eines Star-Plots mit Beispielwerten.

Abbildung 4 zeigt die benutzten Felder in einer Stern-Form angeordnet. Im gezeigten Beispiel sind deutlich erkennbare Ausschläge auf den Segmenten 'points', 'SD_co2', 'SD_lux', 'SD_temp' sowie 'SD_hum' wahrnehmbar. Das würde bedeuten, dass die dazugehörigen Sensorwerte in der jeweiligen Region streuen und die Anzahl der Datenpunkte niedriger ist als in anderen Regionen des Testgebietes.

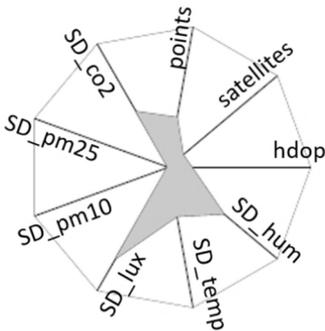


Abb. 4: Star-Plot mit Beispielwerten für die Felder `hdop`, `satellites`, `points`, `SD_co2`, `SD_pm25`, `SD_pm10`, `SD_lux`, `SD_temp` und `SD_hum`

Im Zuge der Daten-Vorverarbeitung werden zunächst die Datenpunkte der Sensorik mit den Regionen des Testgebietes verschnitten. Die Felder `'hdop'` und `'satellites'` sollen die jeweiligen Mittelwerte pro Region zeigen. Mithilfe folgender R Formel kann der Durchschnitt eines Arrays bestimmt werden:

$$\text{mittelwert} \leftarrow \text{average}(x)$$

Für die Felder der Umweltsensorik (`co2`, `pm10`, `pm25`, `lux`, `temp`, `hum'`) werden die Standardabweichungen benötigt. Dafür wird folgende Formel verwendet:

$$\text{standardabweichung} \leftarrow \text{sd}(x)$$

Um Daten innerhalb eines Star-Plots anzuzeigen, ist eine Normalisierung der Werte erforderlich. Um die jeweiligen Werte auf eine Skala von 0 bis 1 zu normalisieren, kann in R folgende Formel verwendet werden:

$$\text{werte_normalisiert} \leftarrow \text{apply}(\text{werte}, 2, \text{function}(x) \ x / (\max(x)))$$

mit den jeweiligen Werten in einem Array, einer 2 für die Anwendung einer Formel auf alle Zeilen des Arrays und der Funktion $x/\max(x)$. Die Daten sollen bei einem ‚guten‘ Wert (zum Beispiel ‚viele Satelliten‘ oder ‚niedrige Standardabweichung der CO₂-Messung‘) einen niedrigen Ausschlag des Star-Plots bewirken und bei ‚schlechten‘ Werten (zum Beispiel ‚wenige Satelliten‘ oder ‚hohe Standardabweichung der CO₂-Messung‘) einen hohen Ausschlag zeigen. Für die Standardabweichungen der Umweltsensorik ist das bereits der Fall. Die Daten der Felder für die Anzahl der Datenpunkte sowie der Anzahl der Satelliten müssen jedoch noch invertiert werden. Dafür lässt sich in R folgende Formel anwenden:

$$\text{werte_invertiert} \leftarrow \text{apply}(\text{werte}, 2, \text{function}(x) \ (1 - x))$$

Mit den Werten der Punktzahl pro Region und der Anzahl der sichtbaren Satelliten in einem Array, einer 2 für die Anwendung einer Formel auf alle Zeilen des Arrays und der Funktion $(1 - x)$. Für die Zusammenführung der standardisierten Werte und der invertierten Daten lässt sich die R-Funktion `cbind` verwenden:

$$\text{werte_gesamt} \leftarrow \text{cbind}(\text{werte_normalisiert}, \text{werte_invertiert})$$

Das Array *werte_gesamt* stellt nun die Datenbasis für die Berechnung der Star-Plots dar. Tabelle 2 zeigt alle berechneten Werte pro Region in einer Übersicht.

Tabelle 2: Berechnete Werte aller Felder für jede Region des Testgebietes als Datengrundlage für die Kalkulation der Star-Plots

Region	hdop (avg)	sats (avg)	points (abs)	co2 (sd)	pm10 (sd)	pm25 (sd)	lux (sd)	temp (sd)	hum (sd)
1	0.15	0	0.46	0.38	0.01	0	0.06	0.72	0.6
2	0.14	0	0.72	0.15	0	0	0.12	0.46	0.27
3	0.17	0.14	0.36	0.45	0	0.01	0.74	0.35	0.51
4	1	0.37	0.42	0.9	1	1	1	0.5	0.93
5	0.15	0	0.39	0.84	0.01	0.01	0.12	0.73	0.45
6	0.14	0	0.48	0.8	0.01	0	0.15	0.53	0.29
7	0.19	0.16	0.36	0.95	0.01	0.01	0.75	0.47	0.57
8	0.24	0.24	0	0.68	0.08	0.03	0.94	0.65	0.73
9	0.15	0.06	0.02	0.59	0.02	0.01	0.17	1	1
10	0.16	0.07	0.38	1	0.01	0.01	0.27	0.58	0.79
11	0.27	0.3	0.42	0.57	0.24	0.12	0.35	0.39	0.95
12	0.39	0.44	0.47	0.25	0.01	0	0.38	0	0.66

Die Tabelle zeigt alle berechneten Werte, die für den finalen Plot verwendet werden sollen. Für die Generierung von Star-Plots stellt die R-Bibliothek *graphics* die Funktion *stars()* bereit. Für den Feldversuch dieser Arbeit wird folgende Formel verwendet:

```
star_plots ← stars(werte_gesamt, locations = mittelpunkte_regionen,
scale = FALSE)
```

mit den darzustellenden Werten *werte_gesamt* und den Positionen der Mittelpunkte der Regionen, um die Star-Plots richtig anzuordnen und zu verorten. Des Weiteren wird der boolsche Wert *scale* auf *FALSE* gesetzt, um dem Algorithmus mitzuteilen, dass eine Normalisierung bereits stattfand.

3.3 Qualitätsparameter der Messung als Star-Plot-Karte

Die Phase der Datenprozessierung resultiert in einem Bild mit einem Star-Plot pro Region. Die jeweiligen Star-Plots sind richtig angeordnet und im Raum verortet, sodass sie in Arc-Map importiert und georeferenziert werden können. Abbildung 5 zeigt das Ergebnis der Fallstudie dargestellt als Star-Plot-Karte.

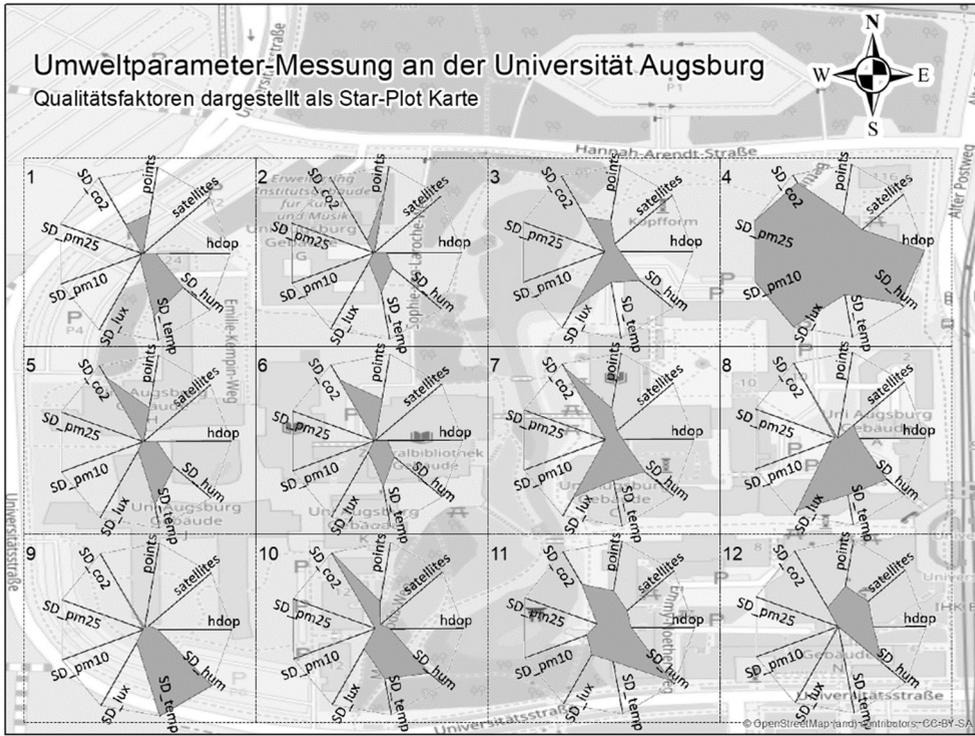


Abb. 5: Karte mit den berechneten Star-Plots für jede Region innerhalb des Testgebietes

Abbildung 5 zeigt die 12 Regionen des Testgebietes, die dazugehörigen Star-Plots sowie eine Hintergrundkarte zur besseren Orientierung. Was zunächst auffällt ist, dass der Plot in Region 4 die größten Ausschläge der Star-Plot Segmente zeigt. Ein Blick auf Abbildung 3 offenbart, dass dort der Startpunkt der Messung liegt. Der vergleichsweise schlechte hdop-Wert lässt sich also darauf zurückführen, dass der Sensor noch nicht so lange lief. Die vergleichsweise hohen Standardabweichungen lassen sich ebenfalls darauf zurückführen, dass die Sensorik zum Anfang hin möglicherweise noch nicht so präzise arbeitet. Des Weiteren kreuzt zu einem späteren Zeitpunkt die Trajektorie südwestlich erneut die Region 4. In der Zwischenzeit haben sich möglicherweise die Werte verändert, was ebenfalls zu erhöhten Standardabweichungen führen kann.

Allgemein lässt sich festhalten, dass die Standardabweichungen des Feinstaubensensors (SD_pm10 und SD_pm25), abgesehen von Region 4, durchgehend niedrig erscheinen. Das deutet darauf hin, dass die Werte innerhalb der Regionen bezüglich der Streuung eine gute Qualität aufweisen. Selbiges lässt sich für die Positionierungsqualität (hdop, satellites) sagen, die ausgenommen von Region 4 durchweg im positiven Bereich zu liegen scheint. Die Anzahl der Punkte pro Region streut jedoch. Auffällig ist, dass Region 2 die niedrigste Punktdichte aller Regionen aufweist. Die Werte SD_lux, SD_temp sowie SD_hum lassen sich schwerer beurteilen als die anderen Variablen, da sowohl die Helligkeit, als auch die Temperatur sowie die Luftfeuchtigkeit im Zuge der Messung kontinuierlich ab- beziehungsweise

zugenommen hat. Die Standardabweichung der CO₂-Messung scheint große Unterschiede aufzuweisen von Region zu Region.

4 Diskussion

Die Fallstudie zeigt, dass Star-Plot-Karten ein probates Mittel sein können, um multivariate Sensordatensätze zu visualisieren. Es lassen sich auf einen Blick mehrere Datenreihen einzelner Variablen miteinander vergleichen, was die Gewinnung eines schnellen Gesamteindrucks des Datensatzes ermöglicht. Gerade wenn es darum geht, verschiedene Qualitätsparameter einer Messung zu vergleichen, scheinen die Star-Plot-Karten eine schnelle und Übersichtliche Möglichkeit zu bieten, Datensätze visuell zu beschreiben. Nichtsdestotrotz gibt es diskutabile Faktoren, die im Nachfolgenden genauer beleuchtet werden.

Die Wahl des Testgebietes sowie dessen räumliche Unterteilung spielen eine zentrale Rolle für das finale Resultat. Nachdem bei zu großen Teilregionen der Generalisierungsfaktor stark ansteigt, bei zu vielen Teilregionen die Karte jedoch unübersichtlich werden könnte, muss ein guter Mittelweg gefunden werden. Hierfür wäre es denkbar, zunächst eine Mindestgröße der Star-Plots zu definieren, um die Lesbarkeit der Karte zu gewährleisten. Dazu sollte beachtet werden, dass keine Regionen mit signifikant weniger Datenpunkten als der Durchschnitt innerhalb des Testgebietes entstehen. Die Unterteilung der Teilregionen in der Fallstudie dieser Arbeit scheint praktikabel, denn die Star-Plots wirken gut lesbar und weisen eine ausgewogene Anzahl an Datenpunkten in jeder Region auf.

Die Wahl und Vorverarbeitung des verwendeten Datensatzes muss kritisch hinterfragt werden. Zum einen finden sich in der Literatur noch weitere Qualitätsmerkmale einer Messung als die Standardabweichung eines Datenfeldes. Des Weiteren muss evaluiert werden, inwiefern eine vorherige Filterung des Datensatzes Notwendig ist. Gerade am Beispiel der Region 4 in Abbildung 5 sieht man, dass einige Extremwerte das Ergebnis deutlich verzerren können. Ein weiterer wichtiger Punkt stellt die Normalisierung der Daten dar. Hier muss hinterfragt werden, ob die verwendete Methode eine gute Grundlage für die Erstellung der Star-Plots darstellt. Nach Chambers et al. (1983) sollen die Werte eines Sternsegmentes auf eine Skala von c bis 1 skaliert werden. In der Fallstudie dieser Arbeit wird c als 0 definiert. Es sollten weitere Skalierungen evaluiert werden.

Auch die Wahl der darzustellenden Datenfelder ist wichtig. Nach Klippel et al. (2009) spielen sowohl die Anzahl der Sternsegmente als auch deren Anordnung untereinander eine Rolle, wenn eine gute Vergleichbarkeit erreicht werden soll. Demnach führen gleiche Datensätze bei variierenden Anordnungen der Segmente zu komplett unterschiedlichen Ergebnissen. Abbildung 5 zeigt in den Regionen 1, 2, 3, 5, 6, 7 und 10 ähnliche Star-Plot Formen. Unterschiede zu den Regionen 4, 8, 9, 11 und 12 sind sofort zu erkennen und erklären. Es ist weiterhin zu evaluieren, inwiefern eine andere Anordnung oder eine Hinzunahme/Reduzierung von Datenfeldern die Lesbarkeit der Star-Plot-Karte positiv oder negativ beeinflusst. Des Weiteren sollte evaluiert werden, ob durch die Verwendung von Farben (nach Klippel et al., 2009b) für Sternsegmente eine verbesserte Lesbarkeit der Karte erreicht werden kann.

5 Fazit und Ausblick

Die vorliegende Arbeit testet die Verwendung von Star-Plot-Karten für die Visualisierung multivariater Sensordatensätze. In einer Fallstudie werden dafür zunächst ein Datensatz erstellt, die aufgenommenen Daten in R vorverarbeitet, für jede Teilregion des Testgebietes ein Star-Plot erstellt und daraus anschließend in ArcMap eine Star-Plot-Karte erstellt.

Die Verwendung von Star-Plot-Karten stellt eine probate Möglichkeit dar, mehrere Datenfelder eines Sensordatensatzes schnell und aussagekräftig zu visualisieren. Es müssen jedoch einige Punkte beachtet werden, um ein aussagekräftiges Ergebnis zu bekommen. Zum einen ist die Wahl und die Untergliederung des Testgebietes wichtig. Es findet allgemein eine Generalisierung der Daten statt, wenn komplette Teilregionen in ein Star-Plot zusammengefasst werden. Deshalb ist es wichtig, die Untergliederung so zu gestalten, dass die Teilgebiete weder zu groß, noch zu klein (aufgrund der Übersichtlichkeit des Ergebnisses) gewählt werden. Des Weiteren ist die Normalisierung beziehungsweise Skalierung der Daten wichtig. Auf einer Skala von c bis 1 sollte der Wert c so gewählt werden, dass unterschiedliche Datenfelder gut miteinander vergleichbar sind. Es muss sich darüber hinaus Gedanken gemacht werden, ob Datenfelder invertiert werden sollten. Verschiedene Felder können unterschiedliche „Richtungen“ aufzeigen („je höher der Wert, desto besser“ versus „je niedriger der Wert, desto besser“). Abschließend muss auf die Anzahl und Anordnung der Sternsegmente geachtet werden, da gleiche Datensätze bei unterschiedlichen Reihenfolgen der Segmente unterschiedliche Ergebnisse produzieren können.

Die zukünftige Arbeit wird sich auf die Optimierung der Verwendung von Star-Plot-Karten fokussieren. Dafür sollen die Vorverarbeitung der Daten, die Untergliederung des Testgebietes, die Verwendung weiterer Qualitätsparameter sowie die Anordnung einzelner Star-Plots und deren Segmente untersucht werden. Im weiteren Verlauf sollen auch weitere der in dieser Arbeit vorgestellten Methoden zur Visualisierung von multivariaten Datensätzen überprüft werden.

Literaturverzeichnis

- Alpern, B., & Carter, L. (1991). *The hyperbox*.
- Budde, M., Riedel, T., Beigl, M., Schäfer, K., Emeis, S., Cyrus, J., and others. (2017). *SmartAQnet: remote and in-situ sensing of urban air quality*.
- Bundesministerium für Bildung und Forschung (2019). *Citizen Science*. Retrieved Jan 01, 2019, from <https://www.bmbf.de/de/citizen-science-wissenschaft-erreicht-die-mitte-der-gesellschaft-225.html>.
- Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). *Graphical methods for data analysis*. Belmont, Calif.: Wadsworth.
- Claessen, J. H., & van Wijk, J. J. (2011). Flexible linked axes for multivariate data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2310–2316.
- Dutta, P., Aoki, P. M., Kumar, N., Mainwaring, A., Myers, C., Willett, W., & Woodruff, A. (2009). *Common sense: participatory urban sensing using a network of handheld air quality monitors*.

- ESRI (n. d.). *HDOP*. Retrieved Jan 01, 2019, from <https://support.esri.com/de/other-resources/gis-dictionary/term/358112bd-b61c-4081-9679-4fca9e3eb926>.
- Fabritiis, C., Ragona, R., & Valenti, G. (2008). *Traffic estimation and prediction based on real time floating car data*.
- Forkel, M. (2012). *Infoblatt Wetterkarten*. Retrieved Jan 01, 2019, from https://www2.klett.de/sixcms/list.php?page=infothek_artikel&extra=TERRA-Online%20/%20Gymnasium&artikel_id=107946&inhalt=klett71prod_1.c.271689.de.
- Hart, J. K., & Martinez, K. (2006). Environmental sensor networks: A revolution in the earth system science? *Earth-Science Reviews*, 78(3-4), 177–191.
- Hauser, H., Ledermann, F., & Doleisch, H. (2002). *Angular brushing of extended parallel coordinates*.
- Inselberg, A. (1985). The plane with parallel coordinates. *The visual computer*, 1(2), 69–91.
- Klippel, A., Hardisty, F., & Weaver, C. (2009). Star plots: How shape characteristics influence classification tasks. *Cartography and Geographic Information Science*, 36(2), 149–163.
- Klippel, A., Hardisty, F., Li, R., & Weaver, C. (2009b). Colour-enhanced star plot glyphs: Can salient shape characteristics be overcome? *Cartographica: The International Journal for Geographic Information and Geovisualization*, 44(3), 217–231.
- Krampe, S., Leitinger, S., Pucher, G., & Rehrl, K. (2013). *FCD Modellregion Salzburg: Einsatz und Nutzen von Extended Floating Car Data im Bundesland Salzburg*.
- Lind, M., Johansson, J., & Cooper, M. (2009). *Many-to-many relational parallel coordinates displays*.
- Liu, T. (n. d.). *Digital-output relative humidity & temperature sensor/module DHT22*. Retrieved Jan 01, 2019, from <https://www.sparkfun.com/datasheets/Sensors/Temperature/DHT22.pdf>.
- Nova Fitness Co. (n. d.). *Laser PM2.5 Sensor specification*. Retrieved Jan 01, 2019, from https://cdn.sparkfun.com/assets/parts/1/2/2/7/5/SDS021_laser_PM2.5_sensor_specification-V1.0.pdf.
- ROHM (2011). *Ambient Light Sensor IC*. Retrieved Jan 01, 2019, from <https://www.mouser.com/ds/2/348/bh1750fvi-e-186247.pdf>.
- Tominski, C., Abello, J., & Schumann, H. (2004). *Axes-based visualizations with radial layouts*.
- u-blox (2011). *Neo-6*. Retrieved Jan 01, 2019, from [https://www.u-blox.com/sites/default/files/products/documents/NEO-6_DataSheet_\(GPS.G6-HW-09005\).pdf](https://www.u-blox.com/sites/default/files/products/documents/NEO-6_DataSheet_(GPS.G6-HW-09005).pdf).
- Vita, E. (2014). *Arduino Nano*. Retrieved Jan 01, 2019 from https://www.arduino.cc/en/uploads/Main/Arduino_Nano-Rev3.2-SCH.pdf.
- Winsen (n. d.). *Intelligent Infrared CO₂ Module (Model: MH-Z19B)*. Retrieved Jan 01, 2019, from https://www.winsen-sensor.com/d/files/infrared-gas-sensor/mh-z19b-co2-ver1_0.pdf.