

Martin-Luther-Universität Halle-Wittenberg  
Naturwissenschaftliche Fakultät III  
Institut für Geowissenschaften und Geographie



# **Anwendung von mobilen Wetterdaten zur Analyse und Einschätzung von Mikroklimaten**

*Application of mobile weather data for the analysis and assessment of  
microclimates*

Masterarbeit im Studiengang Geographie der Martin-Luther-Universität  
Halle-Wittenberg zur Erlangung des akademischen Grades Master of  
Science

Verfasst von:  
Christopher Günter Lorenz  
Matrikelnummer: 219226310  
geboren am 29.06.1991 in Bremen

Erstgutachter:  
Dr. Detlef Thürkow – Martin-Luther-Universität Halle-Wittenberg

Zweitgutachter:  
Dr. Mike Teucher – Martin-Luther-Universität Halle-Wittenberg

Datum der Abgabe: 08.08.2022

## Abstract

Die Verfügbarkeit mikroklimatischer Datensätze gilt als erforderlich, wenn die Entwicklung und Populationsdynamik von Pflanzen- und Tierarten im Kontext des anthropogenen, globalen Klimawandels untersucht werden soll. Zu diesem Zweck werden aktuelle Ansätze aufgezeigt und diskutiert, die Anwendung finden, um derlei Datensätze zu generieren. Der Fokus dieser Arbeit liegt dabei auf mobile Wetterstationsdaten; Feldmessungen, die daraufhin untersucht werden sollen, ob sie einen Mehrwert für die Betrachtung, Analyse und Einschätzung von mikroklimatischen Bedingungen leisten können. Um diese Leitfrage zu beantworten, werden die Wetterdaten anhand programmierter R-Skripte und eines R-Paketes bezüglich ihrer Datenqualität validiert und mit Stationen des Deutschen Wetterdienstes verglichen. Diese Datenqualitätsüberprüfung ergibt, dass circa 95 % aller Lufttemperaturwerte, 97 % aller Niederschlagswerte und 91 % aller Werte der relativen Luftfeuchtigkeit als valide gekennzeichnet werden. Der Stationsvergleich hingegen zeigt auf, dass die untersuchten Klimavariablen *Temperatur* und *relative Luftfeuchtigkeit* starke Ähnlichkeiten mit den Kontrollstationen aufweisen. Die Überprüfung der *Niederschlagswerte* gestaltet sich aufgrund der hohen räumlichen Variabilität dieser Variable aufgrund der Distanzen zwischen den Stationen als ungenau. Diese Auswertung der untersuchten Wetterstationsdaten ergibt, dass die eingesetzten Sensoren jedoch in der Lage sind, lokale Klimaausprägungen mit Distanzen von mindestens einem Meter zum Boden zu erfassen. Deswegen können sie für aufbauende Forschungsansätze in Kombination mit weiteren Datensätzen angewendet werden.

---

The availability of microclimatic datasets is considered necessary when trends of ecological processes like population dynamics of species are to be studied in the context of climate change. To this end, current approaches are highlighted and discussed, while the focus will be on mobile weather station data. Can in situ datasets add to the observation, analysis, and assessment of microclimates? To answer this question, the data will be validated concerning its quality using R scripts and an R package. Furthermore, they will be compared with data from the German Weather Service. This data quality check shows that 95 % of all tested temperature values are classified as valid. Meanwhile, 97 % of the precipitation and 91 % of the relative humidity measurements are valid. The station comparison shows that the examined climate variables *temperature* and *relative humidity* show strong similarities with the control stations. The verification of the *precipitation* values turns out to be inaccurate due to the high spatial variability caused by the distances between the stations. This evaluation of the investigated weather station data shows that the sensors are able to detect local climate characteristics when used properly and can be used for further research.

## Vorwort und Danksagung

Die vorliegende Masterarbeit „Anwendung von mobilen Wetterdaten zur Analyse und Einschätzung von Mikroklimaten“ beschäftigt sich mit der Forschungsfrage: *Eignen sich mobile Wetterstationen wie die Arable-Mark-2, um einen Mehrwert für die Betrachtung, Analyse und Einschätzung von Mikroklimaten zu erzeugen? Wie gut ist die Datenqualität solcher Messsensoren und wie stark weichen die erhobenen In situ Daten von den makroklimatischen Wetterdaten nationaler Wetterstationen ab?*

Diese Masterarbeit stellt den Abschluss meines Masterstudienganges *Master Geographie* an der Martin-Luther-Universität Halle-Wittenberg dar; einem Studium, das außergewöhnlicher kaum sein konnte. Die COVID-19 Pandemie prägte über fast drei Jahre nicht nur das Studium, sondern auch diese Masterarbeit, deren Bearbeitungszeitraum sich mehr als einmal nach hinten verschob.

Gemeinsam mit meinen Betreuern, Dr. Detlef Thürkow (Martin-Luther-Universität Halle-Wittenberg) und Dr. Mike Teucher (Martin-Luther-Universität Halle-Wittenberg), entwickelte ich den Forschungsrahmen dieser Arbeit. Ihnen gebührt mein ganzer Dank. Ohne ihr Zutun wäre der Inhalt der vorliegenden Seiten ein anderer. Ferner möchte ich Herrn Philipp Alb (Martin-Luther-Universität Halle-Wittenberg) für die Bereitstellung der Daten und die Beantwortung von Datenbank-spezifischen Fragen danken. Besonders hervorzuheben ist zudem Herr Christian Hohmann (GFZ-Potsdam), der mir den Einstieg in die Materie der Datenqualitätsüberprüfung von Wetterdaten erst ermöglichte. Ihm verdanke ich erste Gedankenanstöße, die mir aufzeigten, in welche Richtung die Arbeit methodisch gehen könnte. Untermauert wurden diese weiterführend von Dr. rer. Nat. Gerd Schmidt (Martin-Luther-Universität Halle-Wittenberg), der mir zu Beginn der Arbeit half, die methodischen Grundbausteine zu sortieren. Für die Richtlinien und Literatur, die das Fundament der methodischen Auseinandersetzung dieser Arbeit darstellen, muss ich mich an dieser Stelle bei den Damen und Herren des Deutschen Wetterdienstes und der Weltorganisation für Meteorologie bedanken, die mir diese zur Verfügung stellten.

Abschließend möchte ich mich bei meiner Familie und meinen Freunden bedanken, die mir zu jederzeit mit Rat und Tat, aufbauenden Worten und im Fall meiner kleinen Tochter mit Bauklötzen neben der Tastatur zur Seite gestanden haben.

Christopher Günter Lorenz

Halle (Saale), 2. August 2022

# Inhaltsverzeichnis

<b>Abstract</b>	<b>ii</b>
<b>Vorwort und Danksagung</b>	<b>iii</b>
<b>Abbildungsverzeichnis</b>	<b>v</b>
<b>Tabellenverzeichnis</b>	<b>vi</b>
<b>Abkürzungsverzeichnis</b>	<b>vii</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Vorstellung des Themas . . . . .	1
1.2 Ziel der Masterarbeit und verwendete Methoden . . . . .	2
1.3 Aufbau der Masterarbeit . . . . .	3
<b>2 Theoretische Grundlagen</b>	<b>4</b>
2.1 Thematische Eingrenzung . . . . .	4
2.2 Begriffsdefinitionen . . . . .	6
2.3 Forschungsstand . . . . .	9
2.3.1 Entwicklung des wissenschaftlichen Interesses an das Mikroklima . . . . .	9
2.3.2 Rückgriff: Historische Entwicklung der Mikroklimatologie . . . . .	11
2.3.3 Forschungsansätze zur Erforschung von Mikroklimaten . . . . .	12
2.3.4 Zusammenfassung und Ausblick . . . . .	16
2.4 Leitfrage und Relevanz der Arbeit . . . . .	17
<b>3 Methodik</b>	<b>19</b>
3.1 Datengrundlage . . . . .	19
3.2 Datenqualitätsüberprüfung . . . . .	20
3.2.1 Fehlwertüberprüfung . . . . .	22
3.2.2 Technische Überprüfung: Sensor Range Test . . . . .	23
3.2.3 Klimatische Überprüfung: Plausible Values . . . . .	25
3.2.4 Innere Konsistenzprüfung . . . . .	27
3.2.5 Zeitliche Konsistenzprüfung . . . . .	28
3.2.6 Erstellung des allgemeinen Qualitätsschlüssel . . . . .	30
3.2.7 Erstellung des R-Pakets . . . . .	32
3.2.8 Erstellung und Aufbau der R-Skripte . . . . .	32
3.3 Stationsspezifischer Datenvergleich (Arable/DWD) . . . . .	36

3.3.1	Aufbau des Skriptes für den datenspezifischen Stationsvergleich . . . . .	37
<b>4</b>	<b>Ergebnisse</b>	<b>41</b>
4.1	Datenqualitätsüberprüfung und -flagging . . . . .	41
4.2	Stationspezifischer Datenvergleich (Arable/DWD) . . . . .	44
4.2.1	Vergleich der Arable-Station Erdeborn mit der DWD-Station Querfurt . . . . .	45
4.2.2	Vergleich der Arable-Stationen Bentzin mit der DWD-Station Greifswald . . . . .	54
<b>5</b>	<b>Diskussion</b>	<b>62</b>
5.1	Auswahl der angewandten Methoden . . . . .	63
5.2	Ergebnisdiskussion . . . . .	65
5.3	Hypothesenüberprüfung . . . . .	69
5.4	Literaturdiskussion: Relevanz von in situ Daten . . . . .	72
<b>6</b>	<b>Fazit</b>	<b>75</b>
	<b>Anhang</b>	<b>90</b>
	Erläuterung zu den Liniendiagrammen . . . . .	90
	Stationsvergleich: Erdeborn-Querfurt . . . . .	90
	Erläuterung zu den Histogrammen . . . . .	92
	Stationsvergleich: Erdeborn-Querfurt . . . . .	92
	Stationsvergleich: Bentzin-Greifswald . . . . .	92
	Regression Bentzin . . . . .	102
	<b>Selbstständigkeitserklärung</b>	<b>105</b>

## Abbildungsverzeichnis

1	Darstellung der Web of Science Suchen als Liniendiagramm (Eigene Darstellung). . . . .	11
2	Vergleich der Boxplots der Stationen <i>Arable Erdeborn</i> und <i>DWD Querfurt</i> für die Lufttemperaturwerte (Eigene Darstellung). . . . .	47
3	Vergleich der täglichen Durchschnittstemperaturwerte der Stationen <i>Arable Erdeborn</i> und <i>DWD Querfurt</i> anhand eines Liniendiagramms (Eigene Darstellung). . . . .	47
4	Einfaches lineares Regressionsmodell der Temperaturwerte der Stationen <i>Arable Erdeborn</i> und <i>DWD Querfurt</i> (Eigene Darstellung). . . . .	48
5	Vergleich der Boxplots der Stationen <i>Arable Erdeborn</i> und <i>DWD Querfurt</i> für die Niederschlagswerte (Eigene Darstellung). . . . .	49
6	Einfaches lineares Regressionsmodell der Niederschlagswerte der Stationen <i>Arable Erdeborn</i> und <i>DWD Querfurt</i> (Eigene Darstellung). . . . .	50
7	Vergleich der täglichen Summenwerte des Niederschlages der Stationen <i>Arable Erdeborn</i> und <i>DWD Querfurt</i> anhand eines Liniendiagramms (Eigene Darstellung). . . . .	50
8	Vergleich der Boxplots der Stationen <i>Arable Erdeborn</i> und <i>DWD Querfurt</i> für die relativen Luftfeuchtwerte (Eigene Darstellung). . . . .	51
9	Vergleich der täglichen Durchschnittswerte der relativen Luftfeuchtigkeit der Stationen <i>Arable Erdeborn</i> und <i>DWD Querfurt</i> anhand eines Liniendiagramms (Eigene Darstellung). . . . .	52
10	Einfaches lineares Regressionsmodell der relativen Luftfeuchtwerte der Stationen <i>Arable Erdeborn</i> und <i>DWD Querfurt</i> (Eigene Darstellung). . . . .	53
11	Gemeinsame Darstellung der Boxplots der Stationen <i>Arable Bentzin</i> und <i>DWD Greifswald</i> der Klimavariablen <i>Lufttemperatur</i> , <i>Niederschlag</i> und <i>Relative Luftfeuchtigkeit</i> (Eigene Darstellung). . . . .	55
12	Vergleich der täglichen Durchschnittstemperaturwerte der Stationen <i>Arable Bentzin</i> und <i>DWD Greifswald</i> anhand eines Liniendiagramms (Eigene Darstellung). . . . .	56
13	Vergleich der täglichen Niederschlagssummenwerte der Stationen <i>Arable Bentzin</i> und <i>DWD Greifswald</i> anhand eines Liniendiagramms (Eigene Darstellung). . . . .	57
14	Vergleich der täglichen Durchschnittswerte der relativen Luftfeuchtigkeit für die Stationen <i>Arable Bentzin</i> und <i>DWD Greifswald</i> anhand eines Liniendiagramms (Eigene Darstellung). . . . .	57
15	7 tägiger Vergleich der Stundenwerte der Temperatur für Erdeborn und Querfurt (Eigene Darstellung). . . . .	90
16	7 tägiger Vergleich der Stundenwerte des Niederschlages für Erdeborn und Querfurt (Eigene Darstellung). . . . .	91

17	7 tägiger Vergleich der Stundenwerte der relativen Luftfeuchtigkeit für Erdeborn und Querfurt (Eigene Darstellung). . . . .	91
18	Histogramm der Lufttemperatur für Erdeborn (Arable) (Eigene Darstellung). . . . .	93
19	Histogramm der Lufttemperatur für Querfurt (DWD) (Eigene Darstellung). . . . .	93
20	Histogramm des Niederschlages für Erdeborn (Arable) (Eigene Darstellung). . . . .	94
21	Histogramm des Niederschlages für Querfurt (DWD) (Eigene Darstellung). . . . .	94
22	Histogramm der relativen Luftfeuchtigkeit für Erdeborn (Arable) (Eigene Darstellung). . . . .	95
23	Histogramm der relativen Luftfeuchtigkeit für Querfurt (DWD) (Eigene Darstellung). . . . .	95
24	Histogramme: Lufttemperatur Bentzin (Arable) und Greifswald (DWD) ab 04.05.2022 (Eigene Darstellung). . . . .	96
25	Histogramme: Lufttemperatur Bentzin (Arable) und Greifswald (DWD) ab 10.05.2022 (Eigene Darstellung). . . . .	97
26	Histogramme: Niederschlag Bentzin (Arable) und Greifswald (DWD) ab 04.05.2022 (Eigene Darstellung). . . . .	98
27	Histogramme: Niederschlag Bentzin (Arable) und Greifswald (DWD) ab 10.05.2022 (Eigene Darstellung). . . . .	99
28	Histogramme: Rel. Luftfeuchtigkeit Bentzin (Arable) und Greifswald (DWD) ab 04.05.2022 (Eigene Darstellung). . . . .	100
29	Histogramme: Rel. Luftfeuchtigkeit Bentzin (Arable) und Greifswald (DWD) ab 10.05.2022 (Eigene Darstellung) . . . . .	101
30	Lineare Regressionen der Lufttemperaturwerte für Bentzin (Arable) und Greifswald (DWD) (Eigene Darstellung). . . . .	102
31	Lineare Regressionen der Niederschlagswerte für Bentzin (Arable) und Greifswald (DWD) (Eigene Darstellung). . . . .	103
32	Lineare Regressionen der relativen Luftfeuchtigkeitswerte für Bentzin (Arable) und Greifswald (DWD) (Eigene Darstellung). . . . .	104

## Tabellenverzeichnis

1	Informationen zu der Sensorik des Arable-Mark-2 (Arable-Labs 2021, S. 1). . . . .	20
2	Ergebnisse der Datenqualität (Eigene Darstellung). . . . .	42
3	Ergebnisse der einzelnen Datenqualitätstests: Lufttemperatur (Eigene Darstellung). .	43
4	Ergebnisse der einzelnen Datenqualitätstests: Niederschlag (Eigene Darstellung). . .	44
5	Ergebnisse der einzelnen Datenqualitätstests: Rel. Luftfeuchtigkeit (Eigene Darstel- lung). . . . .	44
6	Statistischer Überblick über die Klimavariablen <i>Lufttemperatur</i> , <i>Niederschlag</i> und <i>re-</i> <i>lative Luftfeuchtigkeit</i> für die Stationen <i>Arable Erdeborn</i> und <i>DWD Querfurt</i> (Eigene Darstellung). . . . .	45
7	Regressionsanalyse zwischen den Stationen <i>Arable Bentzin</i> und <i>DWD Greifswald</i> für die Lufttemperaturwerte (Eigene Darstellung). . . . .	58
8	Regressionsanalyse zwischen den Stationen <i>Arable Bentzin</i> und <i>DWD Greifswald</i> für die Niederschlagswerte (Eigene Darstellung). . . . .	59
9	Regressionsanalyse zwischen den Stationen <i>Arable Bentzin</i> und <i>DWD Greifswald</i> für die relativen Luftfeuchtigkeitswerte (Eigene Darstellung). . . . .	61

## **Abkürzungsverzeichnis**

**BMEL** Bundesministerium für Ernährung und Landwirtschaft

**DWD** Deutscher Wetterdienst

**EOS** Earth Observing System

**ESA** European Space Agency

**IPCC** Intergovernmental Panel on Climate Change

**LiDAR** Light Detection and Ranging

**SDM** Species Distribution Model

**TIR** Thermales Infrarot

**TLS** Terrestrisches Laserscanning

**UN** Vereinten Nationen

**UNFCCC** United Nations Framework Convention on Climate Change

**WMO** Weltorganisation für Meteorologie

# 1 Einleitung

## 1.1 Vorstellung des Themas

Der globale, anthropogene Klimawandel erfordert Forschungen zur Erweiterung des derzeitigen Wissensstandes und die Entwicklung von Anpassungsstrategien und Maßnahmen in den verschiedensten Bereichen, die die Landwirtschaft (siehe: Muldowney et al., 2013), den Städtebau (siehe: Reckien et al., 2018), die Klimaforschung (siehe: IPCC 2014), die Wasserwirtschaft (siehe: Wang et al., 2016), den Artenschutz und weitere abdecken (siehe: Bowgen et al., 2022). Zur Analyse dieser einzelnen Bereiche und ihrer spezifischen Aspekte werden qualitativ hochwertige Daten benötigt, die lokale Bedingungen exakt erfassen können (Suggitt et al., 2011, S. 6–7; Zellweger et al., 2019, S. 327–328). Diese Auseinandersetzung mit klimatischen Bedingungen an einem spezifischen, lokalen Raum ist an den Begriff *Mikroklima* gekoppelt. Dieser Begriff ist eng mit der Entstehung der wissenschaftlichen Disziplin der Klimaforschung verknüpft, welche wiederum auf das 19. Jahrhundert zurückgeht (Geiger 1950, S. XVII). Zu dieser Zeit war die Wissenschaft bestrebt, Methoden zu entwickeln, um über Kilometer hinweg die typischen Ausprägungen der Witterung und des Wetters über die Zeit hinweg für einen Landstrich zu erfassen (Geiger 1950, S. XVII). Diese Messungen wurden und werden in Höhen von mindestens zwei Metern Höhe durchgeführt (Geiger 1950, S. XVII). Unterhalb dieser Marke nimmt der Einfluss der umliegenden Umgebung mit ihren klimatologischen Einflussgrößen wie dem Wind oder der Lufttemperatur so stark zu, dass sie sich auf die Messergebnisse auswirken können (Maclean, Duffy et al., 2021, S. 1398). Dieser Fokus auf die Messung von großräumigen Klimaausprägungen, dem sogenannten *Makroklima*, dominierte die Forschung (Geiger 1950, S. XVIII). Veränderungen an den Anforderungen des Informationsgehaltes von Wetterdaten, lenkten ab dem 20. Jahrhundert den Fokus auf die Erfassung, Beschreibung und Analyse mikroklimatischer Bedingungen, die von den bis dato genutzten Messverfahren nicht berücksichtigt wurden (Geiger 1950, S. XVIII).

Der Begriff Mikroklima stellt somit die Auseinandersetzung mit lokalen Wetter- und Klimaphänomenen dar und beschränkt sich je nach Definition auf einen geografischen Raum, der null bis zwei Meter Höhe abdeckt (siehe: Broll et al., 2017). Es ist das Klima der bodennahen Luftschichten, beziehungsweise das Klima des Pflanzenbestandes, das sich in seiner Lufttemperatur und der relativen Luftfeuchtigkeit aufgrund lokaler Gegebenheiten wie Mikrorelief, Pflanzenbestand, abschwächenden Luftzirkulation von der umliegenden Umgebung (dem Meso- und Makroklima) klimatisch unterscheidet (siehe: Broll et al., 2017; Gebhardt et al., 2007, S. 191). Diese Lokalität von Mikroklimaten hat zur Folge, dass sie von gängigen, groß eingesetzten Messverfahren nicht abgedeckt werden. Nationale Wetterstationen zielen etwa darauf ab, die Wetterbedingungen großräumiger Flächen zu beschreiben, wodurch heterogene Ausprägungen des Wetters und damit des Klimas unbedacht blei-

ben können (Geiger 1950, XVII f.). Anhand lokal installierten Messsensoren lassen sich solche lokal geprägten mikroklimatischen Gegebenheiten erfassen, jedoch geht mit diesen Verfahren hohe Anschaffungskosten und Wartungsarbeiten einher, die einen großräumigen Einsatz solcher Methoden erschwert (Maclean, Duffy et al., 2021, S. 1398).

Um dieser Problematik zu begegnen, werden verschiedene Verfahren angewendet, um mikroklimatische Datensätze, die für die exakte Umsetzung von etwa Artenverteilungsmodellen benötigt werden, im Feld zu erheben oder aus anderen Daten zu modellieren (Maclean, Duffy et al., 2021, S. 1398). In ihrer Ausführlichkeit werden diese Ansätze in den nachfolgenden Kapiteln beschrieben, doch um einen kurzen Ausblick auf diese zu geben, können mikroklimatische Bedingungen aus großräumigen Daten wie nationalen Wetterstationsdaten anhand teils rechenaufwendiger Modelle für lange Zeiträume berechnet werden (siehe: Kearney, Shamakhy et al., 2014; Kearney und Porter 2017; Maclean, Mosedale et al., 2019); anhand von Satelliten-, Luftbild- und Drohnenaufnahmen flächenhaft für bestimmte Zeitpunkte aufgezeichnet und modelliert werden (siehe: Zellweger et al., 2019); oder wenn möglich direkt im Feld gemessen werden (Rebaudo et al., 2016, S. 3). Ein jeder dieser Ansätze weist Stärken und Schwächen auf, die miteinander verglichen und diskutiert werden; der Fokus dieser Arbeit wird jedoch auf die Feldmessung mit mobilen Wetterdaten liegen.

## 1.2 Ziel der Masterarbeit und verwendete Methoden

Das Ziel dieser Arbeit ist es, einen gegebenen Datensatz von Wettermessungen, welcher mit mobilen Wetterstationen der Firma *Arable Labs*, die mit den Sensoren *Arable-Mark-2* ausgestattet sind, sowohl auf seine Datenqualität hin zu überprüfen als auch mit einer anderen Wetterdatenquelle zu vergleichen. Dieses Ziel wird getragen von der folgenden Forschungsfrage: *Eignen sich mobile Wetterstationen wie die Arable-Mark-2, um einen Mehrwert für die Betrachtung, Analyse und Einschätzung von Mikroklimaten zu erzeugen? Wie gut ist die Datenqualität solcher Messsensoren und wie stark weichen die erhobenen In situ Daten von den makroklimatischen Wetterdaten nationaler Wetterstationen ab?* – die sich wie ein roter Faden durch diese Arbeit ziehen wird. Umgesetzt wird dieses Ziel anhand von zwei methodisch getrennten Abläufen: zum einen der Datenqualitätsüberprüfung und zum anderen dem stationsspezifischen Datenvergleich.

Die eingesetzten Methoden der Datenqualitätsüberprüfung bauen auf Expertengesprächen auf und werden gestützt von Fachliteratur der Weltorganisation für Meteorologie (WMO) und des Deutschen Wetterdienst (DWD)), wobei letzteres auf den Ansätzen der WMO aufbaut (siehe: Kaspar et al., 2013; Spengler 2002; Zahumenský 2002; WMO 2021). Diese Testverfahren der Datenqualität decken die aufeinander aufbauenden Bereiche der technischen Überprüfung (Fehlwertüberprüfung, sensorbasierte Überprüfung), der regional logischen Überprüfung (klimatische Überprüfung) und

der inneren Logik der Messwerte (innere als auch zeitliche Konsistenzprüfung) ab (siehe: Kapitel 3.1). Umgesetzt werden diese Testverfahren anhand der Programmierung eines R-Paketes, welches die einzelnen Methoden als Funktionen beschreibt, die innerhalb zweier R-Skripte auf den Datensatz angewendet werden. *R* steht hierbei für eine Programmierumgebung, die für den Einsatz der Datenanalyse und der Modellierung anhand statistischer Verfahren gedacht ist (Tucker et al., 2022, S. 2).

Die Methodik des Messwertevergleiches hingegen basiert auf statistische Verfahren, wie sie innerhalb der Klimatologie angewendet werden (Regressionsanalysen, Erstellung von Boxplots, Histogrammen, Kurvendiagramme und weiteres), um zwei Stationsstandortspaare miteinander zu vergleichen (Zimmermann-Janschitz 2014, S. 314–330). Diese Standorte sind Erdeborn-Querfurt und Bentzin-Greifswald, wovon ersterer von einer Station des Arable-Datensatzes und einer ortsnahen DWD-Station vertreten wird und zweiterer von vier Arable-Stationen und einer ortsnahen DWD-Station. Auf diese Weise soll ein detailliertes Bild über die Datengüte der eingesetzten mobilen Wetterstationen gezeichnet werden, um die Relevanz solcher lokaler Wetterdaten für die Erfassung und Erforschung von Mikroklimaten zu diskutieren und hervorzuheben.

### 1.3 Aufbau der Masterarbeit

Der Aufbau der Masterarbeit folgt einem stringenten Aufbau, der in die Thematik einführt und über die Durchführung und Ergebnisdarstellung zur Diskussion führen wird. Das sich dieser Einleitung anschließende Kapitel lautet **2 Theoretische Grundlagen** und baut auf der bereits erfolgten **Vorstellung des Themas (1.1)** auf. So erfolgt zuallererst die **Thematische Eingrenzung (2.1)** in welcher die Masterarbeit in die Thematik des globalen, anthropogenen Klimawandel eingeordnet wird (wobei die gesellschaftliche wie politische Relevanz dessen beleuchtet werden). Dem schließt sich die ausführliche **Begriffsdefinition (2.2)** und die Darstellung des **Forschungsstandes (2.3)** an. Letzteres dient der Darstellung der generellen Entwicklung des wissenschaftlichen Interesses an der Thematik *Mikroklima*; der geschichtlichen Entwicklung dieses Forschungsfeldes; und aktuellen Forschungsansätzen, die neben den direkten Feldmessungen eingesetzt werden, um mikroklimatische Zusammenhänge und Daten zu analysieren und zu modellieren. Den Abschluss dieses Grundlagenkapitels bildet die **Formulierung der Leitfrage und der Hypothesen (2.4)**, die das Rückgrat der eigentlichen Arbeit darstellen.

Die Kapitel **3 Methodik** und **4 Ergebnisse** sind jeweils zweigeteilt und beleuchten erst die Datenqualitätsüberprüfung und anschließend den Datenvergleich. Ersteres beleuchtet hierbei die Beschreibung der Datengrundlage, an der sich die Methodenbeschreibungen der Datenqualitätsüberprüfungen anschließen. Diese Methodenbeschreibung beinhalten jeweils immer die literarische Be-

schreibung dieser, sowie die R-basierte Umsetzung dieser. Abgeschlossen wird dieses Kapitel vom stationsspezifischen Datenvergleich, der die Beschreibung der verwendeten statistischen Methoden und Darstellungsformen sowie die darauf aufbauend programmierten Skripte beinhalten wird.

**4 Ergebnisse** stellt die auf Basis der beschriebenen Methoden erzeugten Ergebnisse dar; wobei der Fokus auf die **Güte der Messwerte (4.1)** und den **stationsspezifischen Datenvergleich (4.2)** liegt. Diese Ergebnisdarstellung dient dem anschließenden Kapitel **5 Diskussion** als Grundlage. Innerhalb dessen erfolgt die Diskussion von vier zentralen Auseinandersetzungen, die die Diskussion der **Methodenauswahl (5.1)**, die **Ergebnisdiskussion (5.2)** mit vergleichbaren Forschungen, die kritische **Hypothesenüberprüfung (5.3)** und die **Literaturdiskussion (5.4)** umfasst. Letzteres wird die Relevanz der Erfassung von mikroklimatischen Daten anhand von Feldmessungen in den Diskurs mit den anderen aktuellen Forschungsansätzen stellen. Den Abschluss dieser Masterarbeit bildet das Kapitel **6 Fazit**. Dieses dient der Zusammenfassung der wichtigsten Ergebnisse, der Beantwortung der formulierten Forschungsfrage, der Einordnung dieser Arbeit innerhalb des skizzierten Forschungsstandes und dem Aufzeigen potenzieller zukünftiger Forschungen.

## 2 Theoretische Grundlagen

Das Kapitel **Theoretische Grundlagen** behandelt neben der Erschließung des thematischen Rahmens (2.1), in dem die Masterarbeit angesiedelt ist, die Definitionsfestschreibung zentraler Begriffe wie Mikroklima (2.2), die Darstellung des aktuell diskutierten Forschungsstandes innerhalb der relevanten Wissenschaftsdisziplinen (2.3) sowie das Herauskrystallisieren der zentralen Leitfrage und der wissenschaftlichen Relevanz dieser Arbeit (2.4). Diese zentralen Themen bilden gemeinsam den theoretischen Grundbau, auf dem die nachfolgenden Kapitel (**3. Methoden und Daten, 4. Ergebnisse und 5. Diskussion**) aufbauen werden.

### 2.1 Thematische Eingrenzung

Der anthropogene (vom Menschen gemachte) Klimawandel und seine direkten, wie indirekten Folgen sind eine der großen Herausforderungen des 21. Jahrhunderts (Gillis 2013; Storlie et al., 2014, S. 1). Nach Thomas F. Stocker (Ko-Vorsitzender des Intergovernmental Panel on Climate Change (IPCC)) stellt er die derzeit größte Bedrohung für unseren Planeten und unsere modernen Lebensweisen dar (Gillis 2013). Um diesem nachhaltig begegnen zu können, werden auf wissenschaftliche Erkenntnisse beruhende Anpassungen der derzeitig ausgelebten gesellschaftlichen wie

wirtschaftlichen Praktiken angestrebt, um für die Herausforderungen der Zukunft gewappnet zu sein (UNFCCC 2022).

Der Beginn der wissenschaftlichen Auseinandersetzung mit Fragen bezüglich des globalen Klimas, seinen lokal divergent ausfallenden Ausprägungen und seinem Wandel über die Zeit wird im Rahmen der wissenschaftlichen Historik auf das 19. Jahrhundert datiert und an den Wissenschaftler Joseph Fourier gekoppelt (Semprebello et al., 2021, S. 20). Er formulierte, aufbauend auf wissenschaftlichen Ansätzen und gedanklichen Konstrukten der vorangegangenen zwei Jahrhunderte, Hypothesen bezüglich der Entstehung der terrestrischen Temperatur und versuchte, anhand dieser, die auf der Erde auftretenden Temperaturunterschiede zu erklären (Fleming 1999, S. 73; Semprebello et al., 2021, S. 22). In einem Artikel, der 1824 erschien, diskutierte Fourier, dass die globale Erwärmung der Erde auf drei Quellen zurückzuführen sei: die Solarstrahlung, die sich ungleichmäßig über das Jahr über den Planeten verteilt; die Energie, die von den Sternen außerhalb des Sonnensystems ans Weltall abgegeben wird; und die Hitze, die vom Erdinneren ausgestrahlt wird (Fleming 1999, S. 73; Semprebello et al., 2021, S. 22). Innerhalb dieses Artikels postulierte er auf Basis von Hypothesen des Naturforschers Horace Bénédict de Saussure, dass sich die Erde in einer gasartigen Hülle befindet, deren Zusammensetzung im Zusammenspiel mit den drei genannten Quellen der Erwärmung zu einem Phänomen führt, das heutzutage als *Treibhauseffekt* bekannt ist (Fleming 1999, S. 73).

Der Treibhauseffekt gilt als Antrieb des globalen Klimawandels und hat sich gemeinsam mit diesen in den letzten Jahrzehnten als stark diskutierter Begriff fest in den gesellschaftlichen, politischen wie wissenschaftlichen Diskurs verankert (Schneider 1989, S. 771; Schmidt et al., 2013, S. 1233). Seit den 90er-Jahren findet eine verstärkte mediale Auseinandersetzung mit dieser Thematik statt (Schmidt et al., 2013, S. 1233), die sich auch in den Wissenschaften niederschlägt (siehe: Kapitel 2.3.1: Abbildung 1). Auf politischer Ebene gewinnt sie seit spätestens der 80er-Jahre an globaler Bedeutung. Im Zuge des wissenschaftlichen und medialen Diskurses gründeten die Vereinten Nationen (UN) und die WMO 1988 den IPCC, der über 600 Klimawissenschaftler aus über 130 Ländern unter einem Dach vereint und in regelmäßigen Abständen Berichte zum Zustand des anthropogenen Klimawandels veröffentlicht (Fanelli 2014, S. 15). Diese Berichte müssen von einem jedem UN-Mitglied inhaltlich abgesegnet werden und gelten als die am stärksten kontrollierten und rezensierten wissenschaftlichen Quellen, die sich mit dieser Thematik auseinandersetzen (Fanelli 2014, S. 15). Mit der Veröffentlichung des ersten IPCC-Berichtes 1990 wurde der Einfluss des Menschen auf den globalen Klimawandel stark diskutiert und zu 50 % als wahrscheinlich angesehen (Fanelli 2014, S. 16). Als Ursachen wurden damals die Nutzung fossiler Brennstoffe und Veränderungen in der Landnutzung benannt (IPCC 1992, S. 9). Diese Einschätzung wird in heutiger Zeit zu 95 bis 99

% als wahrscheinlich angesehen und geht zurück auf den benannten Ausstoß von Treibhausgasen, die durch menschliche Aktivität (wirtschaftliche, industrielle Produktion) in die Atmosphäre abgegeben werden und dort die Treibhauseffekt verstärken (Fanelli 2014, S. 16).

Dieser Ursachenfindung folgend, formulierte das UN-Organ United Nations Framework Convention on Climate Change (UNFCCC) 1997 das *Protokoll von Kyōto zum Rahmenübereinkommen der Vereinten Nationen über Klimaänderungen* sowie das *Übereinkommen von Paris* im Jahr 2015 (Falkner 2016, S. 1107). Letzteres gilt als das erste universelle, rechtsverbindliche globale Klimaschutzabkommen, das von allen Mitgliedsstaaten der UNFCCC verabschiedet wurde (UNFCCC 2022; Falkner 2016, S. 1107). Dieses Abkommen verfolgt das Ziel, anhand der Reduzierung des weltweiten Treibhausgasausstoßes das Ausmaß des globalen Temperaturanstiegs auf unter 2° Celsius, vorzugsweise 1,5° Celsius, zu beschränken (UNFCCC 2015, S. 3; Falkner 2016, S. 1108). Derzeitige wissenschaftliche Erkenntnisse besagen, dass ein Anstieg der terrestrischen Temperatur auf oder über 2° Celsius dazu führen wird, dass der Mensch keinen nennenswerten Einfluss mehr auf die Entwicklung und die direkten und indirekten Folgen des anthropogenen Klimawandels ausüben kann (Fanelli 2014, S. 18). Im *Climate Change 2014 Synthesis Report* werden als potenzielle Folgen unter anderem die Bedrohung von Ökosystemen und Lebewesen wie Pflanzen benannt, die auf drastische Temperaturveränderungen nicht im gleichen Ausmaß reagieren können (IPCC 2014, S. 29). Über 70 % der marinen Lebewesen könnten durch die Reduktion des Sauerstoffgehaltes im Wasser bedroht werden, wohingegen durch weitere Folgen 12 % aller Vogelarten und 25 % aller Säugetiere betroffen sein könnten (Fanelli 2014, S. 18). Ein Anstieg des Meeresspiegels könnte mehr als 30 % der Küstenregionen bedrohen (Fanelli 2014, S. 18). Zusätzlich wird bei einem Anstieg von über 2° Celsius eine Reduzierung der Leistungsfähigkeit ökologischer Dienstleistungen befürchtet, was eine Zunahme der Verknappung an Ressourcen und der Lebensmittelproduktion mit sich führen könnte (IPCC 2014, S. 29).

## 2.2 Begriffsdefinitionen

Im vorangegangenen Unterkapitel fielen bereits einige Begriffe, die zentral für die Auseinandersetzung dieser Arbeit sind; allen voran der Klima-Begriff. Klima als solches bezeichnet die Bedingungen der Atmosphäre an einem bestimmten Ort über einen längeren Zeitraum, der die dort vorherrschende Witterung und das aktuelle Wetter prägt (Hendl und Liedtke 1997, S. 329; A. H. Strahler und A. N. Strahler 2005, S. 179). Diese drei Begriffe *Klima*, *Witterung* und *Wetter* stellen statistische Betrachtungszeiträume dar, in denen die atmosphärischen Bedingungen festgehalten werden (Hendl und Liedtke 1997, S. 329–330). Wetter bezeichnet den physikalischen Zustand der Atmosphäre

zu einem bestimmten Zeitpunkt, an einem bestimmten Ort (siehe: Broll et al., 2017; Gebhardt et al., 2007, S. 190). Witterung hingegen ist die langfristige Betrachtung des Wetters über festgelegte Zeiträume wie Jahreszeiten oder etwa dem Jahresverlauf des Wetters (Gebhardt et al., 2007, S. 190). Der Betrachtungszeitraum des Klimas hingegen beträgt mindestens 10, möglichst 30 Jahre (Hendl und Liedtke 1997, S. 329); er kann und wird jedoch im Rahmen der Paläoklimatologie anhand der Untersuchung von Boden-, Eis- und Seebohrkernen und damit verbundenen Klimaproxyen auch auf Jahrhunderte und Jahrtausende ausgeweitet (Gebhardt et al., 2007, S. 427 & 543–549). Das Klima und die Witterung werden hierbei repräsentiert von der statistischen Beschreibung der Ausprägungen und Schwankungen der atmosphärischen Elemente und stellen somit immer eine Charakterisierung lokaler Atmosphärenbedingungen über die Zeit dar (DWD 2022b; DWD 2022d).

Das Klima einer spezifischen geografischen Region, gekennzeichnet durch einen für diesen charakteristischen jährlichen Ablauf der Witterung, kann sowohl groß- als auch kleinräumig betrachtet werden (Gebhardt et al., 2007, S. 191). Großräumige Betrachtungsperspektiven unterteilen sich in makro- wie mesoklimatische Bedingungen (Gebhardt et al., 2007, S. 191). Das Makroklima beschreibt globale oder auch kontinentale Zusammenhänge (Meeresströmungen, atmosphärische Zirkulation, Klimatypen) und wird besonders durch die großräumigen Prozesse der Atmosphäre gesteuert (Gebhardt et al., 2007, S. 192). Das Mesoklima hingegen beschreibt regionale Gegebenheiten (Stadtklima, Ökosysteme, Landstriche), die von den Eigenschaften und Beschaffenheiten der Relief- und Oberflächenformen geprägt werden (Gebhardt et al., 2007, S. 192). Die kleinste Betrachtungsebene ist das Mikroklima, das von lokal auftretenden Klimavariablen geprägt wird (Gebhardt et al., 2007, S. 192).

Grundsätzlich wird das Mikroklima als das Klima der bodennahen Luftschichten in Höhen von null bis zwei Meter oder als das Klima von Pflanzenbeständen definiert (siehe: Broll et al., 2017). Es ist die Summe der atmosphärischen Bedingungen, die in einem lokalen, begrenzten Raum auftreten und das Resultat der Heterogenität der erdoberflächigen Umweltbedingungen sind (Vanwalleghem und Meentemeyer 2009, S. 1158). Mikroklimatische Orte sind somit geografische Räume, die, bezogen auf ihre Klimatologie, von ihrer direkten Umgebung abweichen (Wanjohi 2017). Die räumliche Ausbreitung von Mikroklimaten ist nicht fest definiert, sondern abhängig vom jeweiligen Untersuchungsobjekt und kann kleine Areale von Größen von wenigen Zentimetern bis einige Meter oder Wälder und Hanglagen abdecken (Wanjohi 2017).

Geprägt wird das Mikroklima von der lokalen Topografie, der Lufttemperatur, der Luftfeuchtigkeit, dem Wind und Luftturbulenzen, dem Hitzeausgleich und der Evaporation, die nicht nur auf horizontaler Ebene wirken, sondern auch in der Vertikalen (Rafferty 2022; Geiger 1950, S. XVIII; Maclean, Duffy et al., 2021, S. 1398). Diese Variablen beeinflussen und bedingen einander und prägen das

Mikroklima auf komplexe Weise: so stehen sie zum Beispiel im direkten Zusammenhang mit der Verteilung von Pflanzenarten, die stark abhängig vom Zustand des Bodens ist (Stoutjesdijk und Barkman 2014, S. 3). Tiere hingegen sind abhängig vom Vorhandensein oder Fehlens bestimmter Pflanzenarten; ihr Vorkommen oder Nichtvorkommen kann somit ebenfalls durch das Vorhandensein einer bestimmten, mikroklimatischen Ausprägung des Klimas begünstigt werden (Stoutjesdijk und Barkman 2014, S. 3). Als eines von vielen Beispielen nennen die Autoren Stoutjesdijk und Barkman (2014, S. 3) die submediterrane Pflanze *Cirsium eriophorum* (Wolfskopf-Kratzdistel), die nur im Südwesten der Niederlande vorkommt, da dort die Winter milde ausfallen und der Niederschlagsverlauf einem mediterranen Muster folgt. Wenn nun eine Tierart exakt jene Pflanze zum Überleben benötigt, dann wird diese Art nur dort vorkommen können, wo diese Pflanzenart vertreten ist (Stoutjesdijk und Barkman 2014, S. 3–5). Auf einen Nationalstaat bezogen, entstehen so Flickenteppiche von kleinräumigen, lokalen Gebieten, die aufgrund ihrer mikroklimatischen Ausprägungen dazu führen, dass Pflanzen- und Tierarten vorkommen können, die für die Region, für das Land, untypisch sind (Stoutjesdijk und Barkman 2014, S. 5). Das Mikroklima ist somit jene Ausprägung des Klimas, die tatsächlich von Lebewesen wahrgenommen wird und diese direkt beeinflusst (Kearney und Porter 2017, S. 664; Storlie et al., 2014, S. 1).

Der Fokus dieser Arbeit liegt auf dem Begriff *Mikroklima*. Wie dargestellt, ist es unter anderem ein zentrales Forschungsobjekt der Pflanzenökologie oder von Untersuchungen des Vorkommens und Populationsdynamiken einer Art (Stoutjesdijk und Barkman 2014, S. 8). Das Meso- und das Makroklima hingegen stellen typische Forschungsobjekte zum einen im Rahmen von bioklimatischen Studien, Planungsvorhaben oder aber auch der Analyse von urban-ökologischen Prozessen dar, wohingegen letzteres in den Disziplinen der physischen Geografie und der Biogeografie Anwendung findet (Stoutjesdijk und Barkman 2014, S. 8). Wenn im Rahmen dieser Arbeit von Mikroklima gesprochen wird, dann sind damit klimatische Orte wie Wälder, Ackerfelder oder Hanglagen gemeint, die sich von ihrer direkten Umgebung abheben (Wanjohi 2017). Dabei wird sich auf die Betrachtung der Verhältnisse wenige Meter über dem Erdboden konzentriert, statt auf die Messung im Zentimeterbereich. Der Grund dafür liegt in der bereits angerissenen vertikalen Ausprägung der einzelnen Klimavariablen, die unterhalb von 1,5 bis 2 Meter in ihrem Zusammenspiel von den höheren Luftschichten abweichen (Maclean, Duffy et al., 2021, S. 1398–1400).

## 2.3 Forschungsstand

Die vorangegangene Skizzierung der Komplexität und das Zusammenwirken der einzelnen in einem Mikroklima wirkenden Variablen nehmen einen großen Stellenwert ein, wenn die Überlebensfähigkeit, die Wanderungsbewegungen und die Populationsdynamiken von Pflanzen und Tieren in der Gegenwart und unter zukünftigen Klimaentwicklungen untersucht werden sollen. Ein weiterer stark diskutierter Aspekt bei der Analyse von lokalen mikroklimatischen Bedingungen ist die Produktionsfähigkeit von Ökosystemen (siehe: Lembrechts, Nijs et al., 2019; Zellweger et al., 2019; Kollas et al., 2014). Forschungsgruppen wie Forest & Nature Lab versuchen aktiv die Verbindung zwischen Biodiversität und Ökosystemdienstleistungen und den Einfluss des Klimawandels auf diese zu erörtern, um auf dieser Basis Handlungsstrategien für die Zukunft zu entwickeln (Ghent-University 2022).

Eine Methode, die angewendet wird, um besagte Auswirkungen auf Arten und Ökosysteme zu ermitteln, sind Modelle zur Artenverteilung (*species distribution model = SDM*) (Lembrechts, Nijs et al., 2019, S. 1267). SDMs werden angewendet, um vorherzusagen, wie Arten in ihrer räumlichen Verteilung auf Umweltveränderungen reagieren (Wanderungen von Arten, Aussterben von Arten ...) (Lembrechts, Nijs et al., 2019, S. 1267). Als Datengrundlage dienen dieser und ähnlichen Methoden makroskalige Datensätze, die großräumige Klimabedingungen erklären (Lembrechts, Nijs et al., 2019, S. 1267; Zellweger et al., 2019, S. 327). Bei der Betrachtung homogener, geografischer Objekte wie der Landschaftstyp des Flachlandes ist die Anwendung solcher Datensätze möglich; werden jedoch heterogene Landschaftstypen betrachtet, kann es zu fehlerhaften Charakterisierungen der Verhältnisse und kritischen Interpretationen der erzielten Ergebnisse kommen (Lembrechts, Nijs et al., 2019, S. 1268). Heterogene Landschaftstypen, die von mikroklimatischen Einflüssen geprägt sind, sind zum Beispiel Wald- und Gebirgsgebiete (Zellweger et al., 2019, S. 327). Die Ausrichtung von Gebirgshängen kann bei einem Vergleich zwischen Nord- und Südhanglage zu Temperaturunterschieden zwischen diesen von bis zu 20° Celsius führen (Zellweger et al. 2019: S. 327). Waldgebiete hingegen weisen aufgrund des Kronendaches teils starke Pufferungseffekte von bis zu 7° Celsius im Unterholz auf (Zellweger et al., 2019, S. 327). Direkte Sonneneinstrahlung, Windgeschwindigkeiten und -turbulenzen werden von diesem abgemildert und resultieren so in die bereits aufgezeigte Verminderung der Temperatur und der Luftfeuchtigkeit (Frenne et al., 2021, S. 2280). Im Vergleich mit ihrem direkten Umland weisen Wälder somit unterhalb des Blätterdaches geringere Maximaltemperaturen und höhere Minimaltemperaturen auf; die jährliche Variabilität des Temperatursignals ist abgemildert (Frenne et al., 2021, S. 2280).

### 2.3.1 Entwicklung des wissenschaftlichen Interesses an das Mikroklima

Diese erste Einführung in die wissenschaftliche Auseinandersetzung mit dem Mikroklima zeigt auf, dass es ein vielschichtiges Forschungsobjekt ist, das für unterschiedliche Anwendungsfelder Rele-

vanz hat. Bezogen auf Themen wie die Klimaerwärmung und dem Treibhauseffekt wurde bereits aufgezeigt, dass ihre Bedeutung innerhalb der Politik, beziehungsweise den Medien ab den 80ern beziehungsweise 90ern zugenommen hat, doch in welchem Ausmaß genau und wie verhält es sich im Detail mit der Thematik Mikroklima?

Schmidt et al., (2013, S. 1245) untersuchten, wie stark die mediale Berichterstattung (anhand von Zeitungsartikeln) zu dem Thema Klimawandel zwischen 1996 und 2010 in 27 Ländern angestiegen ist. Insgesamt stieg der mediale Aufmerksamkeitsgrad um 0,60 % an (Schmidt et al., 2013, S. 1241). Der mediale Aufmerksamkeitsgrad steht hierbei für die Anzahl an Artikel, die sich mit dem Thema *Klimawandel* auseinandergesetzt haben, im Verhältnis gesetzt zu allen anderen untersuchten Artikeln (Schmidt et al., 2013, S. 1240). Bei Betrachtung kleiner Zeitabschnitte wie 1997-2000, 2001-2005 und 2006-2009 ergeben sich auf Basis der getätigten Untersuchung Werte für den medialen Aufmerksamkeitsgrad von 0,20 %, 0,29 % und 1,26 % (Schmidt et al., 2013, S. 1241). Kehren wir nun zum Begriff Mikroklima zurück und unterziehen diesen einem ähnlichen Vergleich, so ergibt sich ein vergleichbares Bild. 1 stellt beispielhaft dar, wie sich zwischen 1992 und 2021 die Anzahl der veröffentlichten wissenschaftlichen Artikel, die die Begriffe *microclimate* (in schwarz), *microclimate ecosystem* (in orange) und *microclimate forest* (in grün) beinhalten, entwickelt hat. Diese Daten basieren auf Zahlen der Datenbank Web of Science (Clarivate 2022) . Web of Science dient der Recherche veröffentlichter, wissenschaftlicher Artikel und ermöglicht es, diese nach ihrem Erscheinungsjahr zu sortieren. Der in dieser Abbildung dargestellte Anstieg der Veröffentlichungen zum Thema *Mikroklima* gleicht dem Bild, wie es Schmidt et al., (2013, S. 1240–1243) bezüglich der medialen Berichterstattung des Klimawandels gezeichnet haben: Zu Beginn der Betrachtung steigt die Anzahl der Artikel nur geringfügig an. Zwischen 2005 und 2010 kommt es zu einer quasi Verdoppelung der Artikel, die den Suchbegriff *microclimate* enthalten (von 222 Artikel im Jahr 2005 auf 378 Artikel im Jahr 2010). In den Folgejahren erfolgt ein starker Anstieg der veröffentlichten Artikel auf bis zu 1419 im Jahr 2021; was dem Vierzehnfachen aus 1992 (100 Artikel) entspricht. Die Anstiege der Suchbegriffskombinationen *microclimate ecosystem* und *microclimate forest* verlaufen vergleichbar, wenn auch weniger drastisch. *Microclimate ecosystem* steigt um das 30-fache an (von 7 Artikel 1992 auf 230 Artikel im Jahr 2021) und *microclimate forest* um das 14-fache (von 25 Artikel 1992 auf 356 Artikel im Jahr 2021). Die Spannweite der Themengebiete reicht dabei von der Analyse der Auswirkung von Vegetation auf das umgebende Mikroklima in urbanen Räumen anhand von *plant evaluation models* (siehe: Yang et al., 2019); über die Vorhersage der zukünftigen Anbaubedingungen landwirtschaftlicher Nutzpflanzen anhand mikroklimatischer Daten (siehe: Gardner et al., 2021); über Untersuchungen, die versuchen, mikroklimatische Bedingungen aus makroskaligen Daten wie nationalen Wetterdaten zu generieren (siehe: Kearney, Shamakhy et al., 2014); bis hin zu Analysen, ob Modelle zur Vorhersage von Insektenhäufigkeiten anhand der Nutzung von Mikroklimadaten ver-

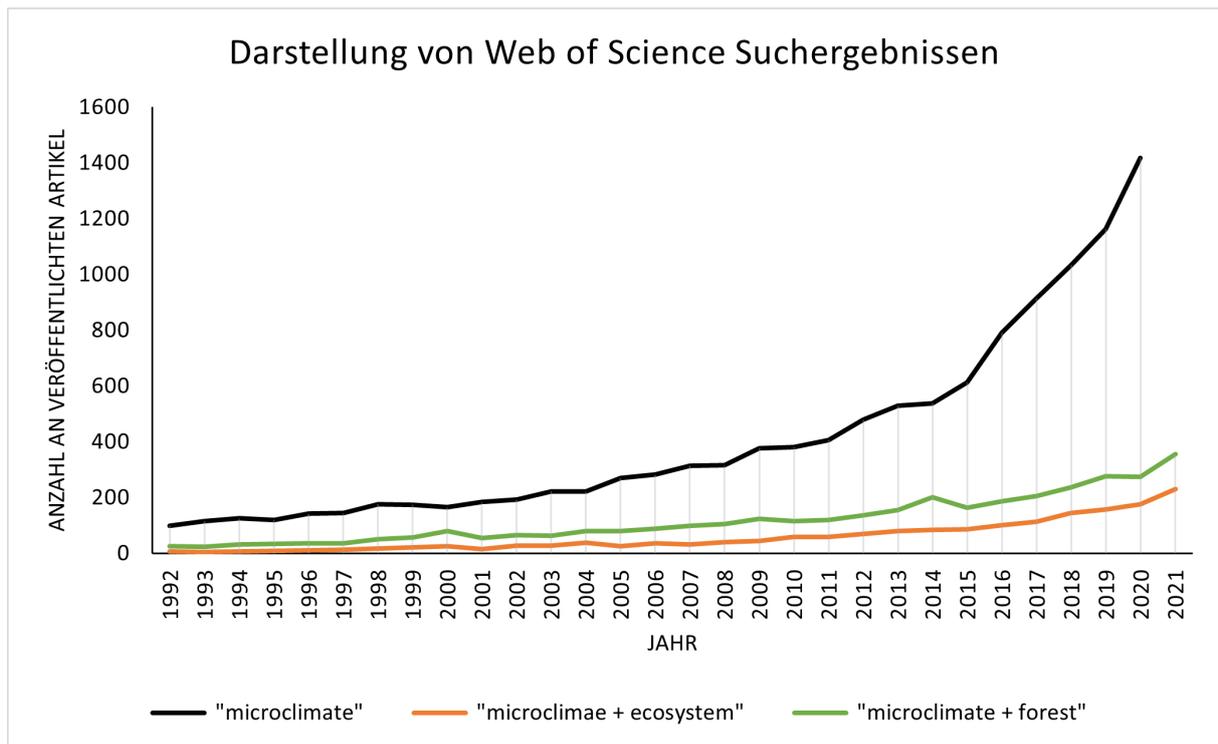


Abbildung 1: Darstellung der Web of Science Suchen als Liniendiagramm (Eigene Darstellung).

bessert werden können und unterstreicht damit die weitgefächerte Reichweite der Thematik (siehe: Rebaudo et al., 2016).

Trotz der unterschiedlichen Untersuchungsobjekte innerhalb der bisher zitierten Literatur zieht sich eine Thematik, ein Konflikt, eine Frage, wie ein roter Faden durch diese hindurch, die für die vorliegende Masterarbeit von zentraler Bedeutung sein wird: *welche Datenskaldimension muss oder kann verwendet werden, um mikroklimatische Bedingungen oder von ihnen abhängige Phänomene zu analysieren?*

### 2.3.2 Rückgriff: Historische Entwicklung der Mikroklimatologie

Um die wissenschaftliche Debatte bezüglich der klimatischen Dimension bezogen auf die Analyse mikroklimatischer Fragen besser greifen zu können, ist ein Blick in die Vergangenheit von Bedeutung. Möchte man den Beginn der wissenschaftlichen Auseinandersetzung mit dem Mikroklima datieren, so muss der Botaniker Gregor Kraus (1841-1915) genannt werden. Der würzburgische Botaniker gilt als einer der Mitbegründer der Mikroklimatologie und veröffentlichte 1911 sein Werk *Boden und Klima auf kleinstem Raum*, das ein wichtiger Grundbaustein nachfolgender Forschungen und Erkenntnisse darstellt (Geiger 1950, S. XX). Prof. Dr. R. Geiger baut mit seinem Werk *The Climate Near the Ground* auf diese und ähnliche Arbeiten auf und definiert in diesem den Begriff

Mikroklima und erläutert, diskutiert und analysiert mikroklimatische Zusammenhänge basierend auf die in diesem wirkenden Klima- und Bodenvariablen (siehe: Geiger 1950). In dem einführenden Kapitel seines Werkes skizziert Geiger die Grundproblematik bei der Analyse von mikroklimatischen Fragestellungen: Mit Beginn der europäischen Wetteraufzeichnung im 19. Jahrhundert wurde auf Basis zahlreicher Experimente die Entscheidung getroffen, die Messsensoren von zu errichtenden Wetterstationen in Höhen von mindestens zwei Meter oder höher über den Boden anzubringen (Geiger 1950, S. XVII). Diese Entscheidung wurde getroffen, weil mit den Wetterstationen keine lokalen, mikroklimatischen Bedingungen gemessen werden sollten, sondern repräsentative Werte, die für ein Gebiet von bis zu mehreren Kilometern valide sein können (Geiger 1950, S. XVII). Aufgrund von steigendem Interesse an Mikroklimadaten seitens der Landwirtschaft vor dem Hintergrund der Gewinnmaximierung zeigte sich die Problematik auf, dass Klimadaten lediglich im makroklimatischen Niveau erhoben wurden, die jedoch nicht die klimatische Realität der Nutzpflanzen widerspiegelt, da diese unterhalb der festgesetzten Höhenmarke existieren (Geiger 1950, S. XVIII). Dieser Erkenntnis folgend, beschreibt Geiger Methoden und Ansätze, die auf die Beschreibung der mikroklimatischen Ausprägungen abzielen, die von späteren, heutigen Bestrebungen weitergeführt werden.

### 2.3.3 Forschungsansätze zur Erforschung von Mikroklimaten

Die Mikroklimatologie ist ein interdisziplinäres Forschungsfeld, das in seinen Anfängen stark von landwirtschaftlichen Interessen geprägt wurde (Kearney, Shamakhy et al., 2014, S. 273). Relevante, mit diesem verknüpfte Forschungsinteressen sind: die Analyse des derzeitigen Artenvorkommens von Flora und Fauna (siehe: Ray et al., 2016), die Populationsentwicklungen dieser und des Menschen (siehe: Rebaudo et al., 2016; Stephenson et al., 2010), die Analyse von vorhandenen Ökosystemdienstleistungen, wie auch die Modellierung zukünftiger Entwicklungen dieser Interessengebiete vor dem Hintergrund des anthropogenen Klimawandels (siehe: Bellard et al., 2012). Die Bedeutung mikroklimatischer Einflüsse auf diese Untersuchungsobjekte und die Relevanz von Datensätzen, die mikroklimatische Gegebenheiten widerspiegeln, wurde seit der Veröffentlichung von *The Climate Near the Ground* von Geiger (siehe: 1950) anhand unterschiedlicher Forschungen nachgewiesen (Rebaudo et al., 2016, S. 1). Exemplarisch kann hierbei die Anwendung von Artenverteilungsmodellen genannt werden. Diese basieren typischerweise auf makroklimatische Datensätze wie nationale Wetterstationssysteme, da mikroklimatische Daten nicht großräumig vorliegen (Zellweger et al., 2019, S. 327). Die Erstellung solcher Modelle auf Basis einer makroklimatischen Datenbasis kann zur Folge haben, dass Schlussfolgerungen über eine Art gezogen werden, die von der Realität abweichen (Gillingham et al., 2012, S. 990–991). Expliziter gesprochen kann das Ergebnis einer solchen Modellierung sein, dass eine Art vom Aussterben bedroht ist, weil sie sich nicht den Klimaänderungen ihrer Umwelt anpassen oder weg wandern kann (Zellweger et al., 2019, S. 328). Wird

ein solches Modell jedoch um mikroklimatische Bedingungen erweitert, kann sich die Schlussfolgerung dahin gehend verändern, dass aufgrund von mikroklimatischen Nischen die Art fortbestehen kann (Scherrer und Körner 2010, S. 414; Wasof et al., 2015, S. 1410). Eine solche Nische kann zum Beispiel durch topologischen Ausprägungen des Reliefs in Gebirgen wie den Alpen hervorgerufen werden, die es bedrohten Pflanzenarten ermöglicht, trotz der Folgen des globalen Klimawandels fortzubestehen (Scherrer und Körner 2010, S. 414; Wasof et al., 2015, S. 1410).

Gegenstand interdisziplinärer Debatten ist an dieser Stelle die Frage, ob solche und ähnliche Fragestellungen für ihre Analyse mikroklimatische Datensätze benötigen oder ob diese häufig nicht vorliegenden Daten auch aus meso- oder makroklimatischen Datensätze modelliert werden können (Rebaudo et al., 2016, S. 1)? Potenzielle Antworten auf diese Frage sollen an dieser Stelle nicht erörtert werden – dafür dient das **Kapitel 5 Diskussion** –, stattdessen werden nachfolgend ein Teil der wichtigsten, derzeitigen Methoden vorgestellt, die im wissenschaftlichen Alltag ausgeübt werden. Diese stellen einen fundamentalen Anteil des aktuellen Forschungsstandes dar und können zu teilen als Gegenentwürfe des Ansatzes dieser Arbeit gewertet werden.

Grundsätzlich gesprochen lassen sich drei Ansätze unterscheiden, die die Anwendung und Analyse mikroklimatischer Bedingungen ermöglichen: die Modellierung aus makroklimatischen Daten oder fernerkundliche Datensätze auf der einen Seite und die Erhebung von in situ Daten (Maclean, Mosedale et al., 2019, S. 281). Makroklimatische Datensätze sind unter anderem nationale Wetterdaten wie jene des DWD oder räumlich hochaufgelöste globale Wetter- und Klimadaten wie jene der *WorldClim*-Datenbank (Fick und Hijmans 2017, S. 4313–4314). Diese werden aufgrund ihrer zeitlichen Verfügbarkeit herangezogen und ermöglichen es, die Variationen des Klimas über die Zeit zu ermitteln (Maclean, Mosedale et al., 2019, S. 281). In ihrer Rohform repräsentieren diese jedoch nicht mikroklimatische, sondern makroklimatische Bedingungen (Slavich et al., 2014, S. 953). Stattdessen dienen diese als Eingangsdatensätze für Modellierungsverfahren, die versuchen, kleinräumige Klimageschehnisse zu berechnen (Kearney und Porter 2017, S. 665; Maclean, Mosedale et al., 2019, S. 281). Die Fülle an methodischen Ansätzen ist dabei mannigfaltig. Regionale Klimamodelle versuchen kleinräumigere Bedingungen anhand der Anwendung von physikalischen Gesetzen zu berechnen (Xuejie et al., 2001, S. 1224–1226). Ihr Einsatz kommt jedoch aufgrund der Komplexität dieser Berechnung und damit einhergehendem konstant ansteigenden Rechenaufwand ab circa 5 Kilometer an ihr Ende (Maclean, Mosedale et al., 2019, S. 281). Stattdessen werden Landoberflächenmodelle und mechanistische Ansätze angewendet, um sich den mikroklimatischen Bedingungen anzunähern. Landoberflächenmodelle wenden ebenfalls physikalische Gleichungen an, werden jedoch punktuell oder entlang eines Gitters angewendet, um die Energie- und Wasserbilanz an jedem Ort zu berechnen, und liefern Informationen zu den bodennahen Temperaturen und

der Luftfeuchtigkeit (Best et al., 2011, S. 693). Eine Weiterentwicklung dieser eher für meteorologische und hydrologische Zwecke entwickelten Modelle sind mechanistische Ansätze, allen voran das R-Paket *NICHE MAPPER* von Kearney und Porter (siehe: 2017). Der *NICHE MAPPER* versucht bodennahe, mikroklimatische Bedingungen anhand von gegebenen Wetterinformationen der Umgebung und den Eigenschaften des Habitats zu ermitteln (Kearney und Porter 2017, S. 665). Besonderes Augenmerk liegt hierbei auf die Modellierung des Energie- und Massehaushaltes von Organismen und ihrem Mikroklima (Maclean, Mosedale et al., 2019, S. 281). So eignen diese sich zum Beispiel, um für landwirtschaftliche Nutzpflanzen anhand von Klimavariablen, die für diese Art relevant sind, ihr Wachstum und ihren Nährstoff- und Kohlenstoffhaushalt vorherzusagen (Gardner et al., 2021, S. 2). Je kleinräumiger die Eingangsdaten dabei sind, desto besser werden die Ergebnisse, jedoch mangelt es bei der Anwendung dieser Modelle an eben jene (Gardner et al., 2021, S. 2).

Einer der aktuellen Ansätze, der entwickelt wurde, um dieses Problem zu umgehen, sieht die Erweiterung mechanistischer Modelle durch die Beigabe von mesoklimatischen Prozesse vor (Lembrechts und Lenoir 2019, S. 337; siehe: Maclean, Mosedale et al., 2019). Die Beigabe von Variablen wie der Kaltluftzufuhr, der Küstenexposition und der Geländehöhe, gekoppelt an der Betrachtung der mikrotopografischen Variationen des Geländes, der Vegetationsbedeckung und der Bodeneigenschaften anhand von Energiehaushaltsgleichungen, sollen verbesserte Ergebnisse ermöglichen (Lembrechts und Lenoir 2019, S. 337; siehe: Maclean, Mosedale et al., 2019). Diese Ergebnisse könnten zum Beispiel Stunden- wie Tageswerte der Lufttemperatur und Bodenfeuchtigkeit sein, die sowohl gegenwärtige als auch zukünftige Prognosen abdecken (Lembrechts und Lenoir 2019, S. 337; siehe: Maclean, Mosedale et al., 2019). Die Anwendbarkeit und Genauigkeit dieser Modelle muss jedoch erst noch validiert werden (Lembrechts und Lenoir 2019, S. 338).

Fernerkundliche Ansätze vereinen Aspekte der Modellierung anhand makroklimatischer Datensätze mit Messungen vor Ort. Unterscheiden lassen sich hierbei zwei grundlegende Arten der Datenerhebung, die jeweils mit der Entfernung des Aufzeichnungsobjektes zu tun haben: Satellitenerhebung und Flugzeug- oder Drohnenbefliegung (Johannsen und Daughtry 2009, S. 251). Diese Verfahren generieren anhand der eingesetzten Sensoren Daten unterschiedlicher räumlicher Auflösung, die von 10 bis 60 m räumlicher Auflösung pro Bildpixel bei den europäischen SENTINEL-2 Satelliten (EOS 2022) und bis zu kleiner als 5 cm bei Drohnenfotos reichen können (Gray et al., 2018, S. 1). Fernerkundliche Erhebungsansätze generieren Datensätze, die für die Erstellung von Artenverteilungsmodellen oder auch Landschaftsklassifikationen von großer Bedeutung sind, da sie es erlauben, großräumig und hochaufgelöstes Kartenmaterial zu generieren (Maclean, Mosedale et al., 2019, S. 281).

Unterscheiden lassen sich fernerkundliche Daten anhand des jeweils genutzten Aufnahmeegerätes, wovon nachfolgend das *Light Detection and Ranging (LiDAR)*-Verfahren, das *Terrestrisches Laserscanning (TLS)*-Verfahren und die Erstellung von Thermalbildern anhand von *Thermales Infrarot (TIR)* vorgestellt werden sollen.

LiDAR-Daten ermöglichen es, räumlich zusammenhängende Daten mit Auflösungen von bis zu unter einem Meter zu sammeln, die wichtige Informationen über die Bodenbeschaffenheit und die Vegetationsstruktur liefern (Lefsky et al., 2002, S. 27–29; Zellweger et al., 2019, S. 329). Eine günstigere Alternative ist die Verwendung von Luftfotografien (Flugzeug oder Drohnen), deren Informationsgehalt jedoch geringer ausfallen, da sie in ihrem räumlichen Ausmaß einen kleineren Ausschnitt abdecken als es Satellitenaufnahmen vermögen (Zellweger et al., 2019, S. 329). Dennoch stellt besonders die Anwendung von Drohnen ein viel genutztes Werkzeug dar, um ortsspezifische Informationen zu sammeln (Nattke 2021, S. 389). Sie stellen eine wichtige Informationsquelle dar, wenn die Phänologie von Vegetationen und damit ihre Auswirkung auf das Mikroklima über die Zeit beobachtet werden soll (Zellweger et al., 2019, S. 329). Terrestrische Laserscanning Verfahren anhand von Satelliten liefern ergänzende Informationen zur Vegetationsstruktur und ermöglichen etwa die Analyse des Unterholzes (Liang et al., 2016, S. 64). Um Oberflächentemperaturen großräumig zu messen und anschließend mit statistischen Interpolationsverfahren für die Analyse mikroklimatischer Bedingungen nutzbar zu machen, eignen sich thermale Infrarotaufzeichnung (Zellweger et al., 2019, S. 331). TIR-Kameras zeichnen die langwellige von Objekten reflektierte Infrarotstrahlung auf, die Aussagen über die Oberflächentemperatur und über den Energiehaushalt von Pflanzen und Tieren ermöglichen (Zellweger et al., 2019, S. 331). Sie liefern jedoch keine direkten Informationen zur Lufttemperatur (Zellweger et al., 2019, S. 331). Durch den Einsatz von Drohnen, die im Zentimeterbereich Daten sammeln können, lassen sich auf dieser Datenbasis Aussagen über den Wasserhaushalt und Wasserstress von Pflanzen oder über die bodennahen Temperaturen treffen, denen auf den Pflanzen Insekten ausgesetzt sind, deren Vorkommen im Rahmen der Landwirtschaft von Relevanz sein können (siehe: Gray et al., 2018; Zellweger et al., 2019, S. 331). Diese und weitere Fernerkundungsdatensätze können in Kombination betrachtet wichtige Informationen zum Mikroklima eines Ortes generieren.

Die dritte Möglichkeit mikroklimatische Daten zu generieren, ist die Messung direkt vor Ort im Feld, statt in der Luft, wie es beim Einsatz von Drohnen erfolgt. Hierbei steht eine Vielzahl an Gerätschaften wie Thermometer zur Verfügung, die manuell oder automatisch Daten generieren. Aktuelle Forschungsartikel betonen diesbezüglich jedoch, dass es, je nachdem wie nah am Erdboden man Messungen vornehmen möchte, kein leichtes Unterfangen ist, akkurate mikroklimatische Daten zu erheben (Maclean, Duffy et al., 2021, 1407 f.). So ergibt sich beispielsweise für die Messung der

Klimavariable *Lufttemperatur* die Problematik, dass Messungen unter der typischen Wetterstationsmarke von 1,2 – 2 Meter über dem Boden Störungsquellen ausgesetzt sein können, die die Messung verzerren (Maclean, Duffy et al., 2021, S. 1398).

In der ersten Edition des *Guide to Meteorological Instruments and Methods of Observation* von 1954 schreibt die WMO, dass für die Messung der Variable Lufttemperatur ein Messgerät von der direkten Einstrahlung der Sonne und der reflektierten Strahlung der umliegenden Objekte abgeschirmt werden muss, da ansonsten Abweichungen der Messungen von der Realität möglich sind (WMO 2018, S. 107–109). Die verwendeten Schilde, um ein Thermometer abzuschirmen, sind bei klassischen Wetterstationen aus reflektierendem Material hergestellt, berühren nicht direkt das Thermometer und ermöglichen eine gute Luftzirkulation innerhalb des Schildes, damit die innere Lufttemperatur der äußeren gleicht (WMO 2018, S. 107). Wird nun die Lufttemperatur unterhalb der typischen Wetterstationshöhe oder gar bodennah gemessen, verringert sich die Zirkulation der Luft und der Einfluss der Umgebungsstrahlung nimmt zu (Maclean, Duffy et al., 2021, S. 1398). Die Problematik, die sich daraus ergibt, ist folgende: nutzt man einen Schild, um das Thermometer vor der direkten Einstrahlung und der reflektierten Strahlung der Umgebung abzuschirmen, kommt es zu Ungenauigkeiten aufgrund der fehlenden Luftzirkulation; gleichzeitig kann der Schild jedoch nicht weggelassen werden, weil sonst der benannte Einfluss der Umgebungsstrahlung zu stark wird (WMO 2018, S. 107–109; Maclean, Duffy et al., 2021, S. 1398).

Diese auf die Lufttemperatur bezogene Problematik für Messungen nahe dem Boden veranschaulicht, dass auch die Erhebung von in situ Daten ein komplexes Unterfangen ist. Es reicht nicht aus, Messsensoren zu montieren, sondern stattdessen muss stark auf die direkte Umgebung geachtet und die richtige Stelle für den richtigen Sensor gefunden werden (Stoutjesdijk und Barkman 2014, 173 ff.). Ein zusätzliches Problem gestaltet sich in der unausweichlichen Geräteanschaffung (Maclean, Duffy et al., 2021, S. 1398). Trotz dessen ermöglichen direkte vor Ort Messungen gezielte Aussagen über den Einfluss des Klimas, des Wetters und der Witterung auf die vorkommenden Pflanzen- und Tierarten ziehen zu können (Field-Studies-Council 2022).

#### **2.3.4 Zusammenfassung und Ausblick**

Die Mikroklimatologie ist ein Forschungsfeld, das in den letzten Jahrzehnten aufgrund des wachsenden Interesses an der Auseinandersetzung mit dem anthropogenen Klimawandel und dessen gegenwärtige und zukünftige Auswirkungen auf unser Leben und das von Pflanzen und Tieren, an Bedeutung gewonnen hat. Besonders ökologische Fragestellungen, im Hinblick auf eine nachhaltige, produktive Landwirtschaft (siehe: BMEL 2022), dem Modellieren des derzeitigen Vorkommens und der zukünftigen Entwicklung von Arten (Bellard et al., 2012, S. 365–367) und ähnlichem, dominieren die Auseinandersetzung mit diesem Forschungsfeld.

Die drei vorgestellten Ansätze erlauben es zu unterschiedlichen Graden mikroklimatische Bedingungen zu erfassen, jedoch weisen sie alle unterschiedlichen Stärken und Schwächen auf. Während nationale Wetterdaten und vergleichbares sich für die Analyse langer Zeitreihen eignen, beschränkt sich ihre Aussagekraft über lokale Mikroklimaten auf die Güte der angewendeten Modellierung (Maclean, Duffy et al., 2021, S. 1398). Fernerkundungsdatensätze decken zum Beispiel bei Satellitendaten mit ihren Messungen große Regionen bildhaft ab, jedoch sind sie in ihrer temporären Auflösung teils stark abhängig von Wetterbedingungen (Wolkendecken), die Umlaufbahn der Satelliten und die Auflösungsqualität der Sensoren, die von hunderten Metern bis zu wenige Meter reichen kann (Al-Wassai und Kalyankar 2013, S. 54). Die Erhebung von Daten direkt im Feld geht mit ihren eigenen Problematiken einher (vertikaler Einfluss von Klimavariablen wie den Wind auf Messsensoren wie zum Beispiel Thermometer, Anschaffungskosten, geringe räumliche Abdeckung, Dauer der Erhebungen), jedoch ermöglicht sie die Messung von lokalen, klimatischen Gegebenheiten für Räume, die sich nicht in unmittelbarer Nähe zu einer nationalen Wetterstation und ähnlichem befinden (Maclean, Duffy et al., 2021, S. 1408–1409). Aufgrund dieser unterschiedlichen Stärken und Schwächen der Ansätze sehen aktuelle Forschungsansätze eine Verknüpfung dieser vor, um durch die Verwendung aller drei Arten die realitätsnächste Abbildung mikroklimatischer Bedingungen erzielen zu können (Lembrechts, Nijs et al., 2019, S. 1274).

## 2.4 Leitfrage und Relevanz der Arbeit

Die literarische Auseinandersetzung mit der Thematik der Erhebung oder auch Modellierung von mikroklimatischen Daten lässt mehrere Anknüpfungspunkte für die eigene wissenschaftliche Arbeit zu. Mehrere Faktoren, darunter die grundsätzliche Verfügbarkeit der Messgeräte und das persönliche Interesse, führten dazu, dass sich diese Arbeit auf den zuletzt angesprochenen Zweig der Erhebungsmethoden stützt: dem Erheben von lokalen Klimadaten direkt im Feld. Diese Daten, auf die im nachfolgenden Kapitel näher eingegangen wird, wurden und werden mit mobilen Wetterstationen der Firma *Arable Labs* erhoben. *Arable Labs* ist Produzent und Dienstleister von mobilen Wetterstationen, der sich dazu verschrieben hat, Prozesse der Digitalisierung in die Landwirtschaft zu überführen (*Arable-Labs* 2022a). Durch das zur Verfügung stellen von lokalen Klimadaten sollen Entscheidungsträger die Leistungsfähigkeit ihrer Nutzpflanzen steigern und ihr Wissen über die klimatischen Bedingungen, denen die Pflanzen ausgesetzt sind, erweitern (*Arable-Labs* 2022b). Da diese mobilen Messstationen für den Einsatz innerhalb der Landwirtschaft konstruiert sind, messen sie in diesem Fall die klimatischen Bedingungen wenige Meter über dem Boden im offenen Feld (*Arable-Labs* 2021, S. 3). Der Fokus dieser Arbeit liegt dabei weniger auf die direkte Anwendung der Messdaten als vielmehr auf die Überprüfung der Datenqualität und darauf, wie sie im Vergleich zu

konventionellen Datensätzen wie nationale Wetterstationen abschneiden. Darauf basierend formulieren sich die Leitfragen dieser Arbeit: *Eignen sich mobile Wetterstationen wie die Arable-Mark-2, um einen Mehrwert für die Betrachtung, Analyse und Einschätzung von Mikroklimaten zu erzeugen? Wie gut ist die Datenqualität solcher Messsensoren und wie stark weichen die erhobenen in situ Daten von den makroklimatischen Wetterdaten nationaler Wetterstationen ab?*

Vor diesem Hintergrund ist ein zentrales Ziel dieser Arbeit, Methoden zur Datenqualitätsüberprüfung zusammenzufassen, auf einen stetig wachsenden Datensatz anzuwenden und diese Daten mit anderen Datenquellen grafisch zu vergleichen, um Ähnlichkeiten oder Unähnlichkeiten herauszuarbeiten. Ferner soll sie Anreize für Forschungsvorhaben liefern, die mit diesen Daten arbeiten werden. Diese beiden Faktoren zusammengenommen bilden die Relevanz dieser Arbeit. Vor dem Hintergrund des bereits vorgestellten Diskurses soll sie eine Möglichkeit aufzeigen, im Feld zeitlich hochaufgelöste Daten zu sammeln, mit deren Hilfe die Güte von Forschungsfragen, die mit lokalen Klimabedingungen arbeiten müssen, zu erhöhen. Gleichzeitig soll sie aber auch dem Beispiel von Maclean, Duffy et al., (siehe: 2021) die Grenzen solcher Messgeräte aufzeigen.

Zum Erreichen dieser Ziele werden sich die nun anschließenden Kapitel, ganz besonders die Interpretation und Diskussion der Ergebnisse an den nachfolgenden Hypothesen orientieren, die es zu überprüfen gilt:

- Der Einsatz mobiler Wetterstationen ermöglicht eine exaktere Abbildung der Realität als es mit makroklimatischen Daten möglich ist (Vergleich: Burnett und Anderson 2019; Kearney, Shamakhy et al., 2014; Kollas et al., 2014; Lembrechts, Nijs et al., 2019; Storlie et al., 2014; Zellweger et al., 2019).
- Wird die Datenqualität der zu überprüfenden Daten als gut befunden, dann werden die Abweichungen zwischen diesen und anderen Wetterstationen wie jenen des DWD nur geringe Messunterschiede aufweisen (Arable-Labs 2020a; Arable-Labs 2020b).
- Wenn die Arable-Mark-2 mobilen Wetterstationen vergleichbare Ergebnisse wie etablierte Wetterstationen erzielen, dann werden sie auch im Feld, wo keine großräumigen Daten erhoben werden, qualitativ gute Daten erheben.
- Wenn die Genauigkeit der Messdaten von mobilen Wetterstationen hoch ausfällt, dann können mit diesen mikroklimatischen Bedingungen beobachtet und analysiert werden.

## 3 Methodik

### 3.1 Datengrundlage

In dieser Arbeit finden zwei Datensätze Anwendung: zum einen die vom Fachbereich *Digital Geography* der Martin-Luther-Universität Halle-Wittenberg zur Verfügung gestellten Klimadaten, die mit den mobilen Wetterstationen Arable-Mark-2 erhoben wurden, und Klimadaten des DWD. Die den zu testenden Arable-Klimadaten zugrundeliegenden mobilen Wetterstationen sind *all-in-one* Lösungen, die neben relevanten Klimaparametern Pflanzeninformationen sammeln, um einen Einblick in das Wachstum und das generelle Wohlbefinden der Nutzpflanzen zu liefern (Arable-Labs 2020a, S. 3). Die Geräte sind witterungsbeständig gebaut und verfügen über Mobilfunkverbindungen, um die gemessenen Daten ohne Feldbetretung abrufen zu können (Arable-Labs 2020a, S. 3). Ferner verfügen sie über GPS, einem 6-Kanal Spektrometer, Radiometer für langwellige wie kurzwellige Strahlung, sowie ein akustisches Disdrometer (Arable-Labs 2020a, S. 3). Zusammengenommen ermöglichen diese Sensoren die Messung von über 40 Pflanzen- und Klimadaten (Arable-Labs 2020a, S. 3).

Der erhobene Arable-Datensatz enthält die gemessenen Wetterdaten von derzeit 48 Stationsstandorten, die über gesamt Mittel- und Norddeutschland verteilt sind. Die ersten Messwerte liegen ab dem **13.06.2021** in stündlicher Auflösung vor, wobei für neuere Stationsstandorte andere Start- und auch Enddaten gelten. Der Datensatz ist somit zeitlich betrachtet stark heterogen und geprägt von aktuell stattfindenden Messungen und Untersuchungen im Feld. Dieser Aspekt spielte bei der Umsetzung der Methodik eine relevante Rolle (siehe: 3.3.1). Für die Untersuchungen der Datenqualität wurde sich auf drei Klimavariablen konzentriert: der Lufttemperatur (im Datensatz = *t\_air*, *t\_min*, *t\_max* und *t\_dew*), dem Niederschlag (im Datensatz = *precip*) und der relativen Luftfeuchtigkeit (im Datensatz = *rel\_hum*). Weitere zwingend erforderliche Variablen sind unter anderem *data\_id*, *station\_id*, *date* und *date\_time*, die der Identifikation der einzelnen Messwerte dienen.

Wichtig bei der Verwendung dieser Daten ist zu wissen, dass sie bereits bis zu einem gewissen Grad korrigiert, beziehungsweise bearbeitet wurden. Arable Labs verwendet automatisierte AI-gestützte Algorithmen, um Klimavariablen zu messen und zu korrigieren (Arable-Labs 2021, S. 1). Dies trifft für jeden einzelnen der zu untersuchenden Klimavariablen zu. Sowohl die Temperatur- als auch die relativen Feuchtigkeitswerte werden nach der Erhebung im Feld durch einen AI-gestützten Algorithmus korrigiert, jedoch geht nicht hervor, bis zu welchem Grad diese Korrektur erfolgt (Arable-Labs 2021, S. 1). Der Niederschlag hingegen wird mit einem Disdrometer ermittelt; das bedeutet, dass anhand des akustischen Signals fallender Regentropfen die Anzahl und die Stärke des Regenereignisses geschätzt werden (Arable-Labs 2021, S. 2). Anhand der gemessenen Frequenz der einzelnen Tropfen berechnet ein Algorithmus die Größe der Tropfen und schließt auf dieser Basis auf die

Gesamtmenge und die Niederschlagsrate (Arable-Labs 2021, S. 2).

Die nachfolgende Tabelle 1 beinhaltet relevante Informationen zu den drei zu untersuchenden Klimavariablen *Lufttemperatur*, *Niederschlag* und der *relativen Luftfeuchtigkeit*. Hervorzuheben ist dabei die Genauigkeit der Sensoren und der Messbereich, für den diese kalibriert sind. Diese Werte fließen in die Methodik ein, die im nachfolgenden Kapitel erläutert wird.

Die zweite Datensatzquelle, die für den Vergleich der Arable-Daten mit anderen verwendet wird,

Tabelle 1: Informationen zu der Sensorik des Arable-Mark-2 (Arable-Labs 2021, S. 1).

Klimavariablen	Einheit	Messbereich	Genauigkeit
Lufttemperatur	° C	Kalibriert: -18° bis 45° C	Kalibriert: ± 0,8° C Nicht-kalibriert: ± 1,5° C
Niederschlag	mm	0 bis 100 mm pro Stunde	± 6 %
Relative Luftfeuchtigkeit	%	1 bis 100 %	± 5 %

sind Wetterdaten des DWD. Diese stammen von den Stationen Querfurt (DWD-ID: 04036) und Greifswald (DWD-ID: 01757) und liegen in stündlicher Auflösung vor. Ausgewählt wurden hierbei die gleichen Klimavariablen wie bei den Arable-Daten, deren Kürzel *TT\_TU* für die Lufttemperatur, *R1* für den Niederschlag und *RF\_TU* für die relative Luftfeuchtigkeit lauten.

Ein wichtiger Hinweis zu den Einheiten der einzelnen Variablen: Während die Temperaturwerte beider Datensätze in Grad Celsius und der Niederschlag in Millimeter angegeben werden, ist die relative Luftfeuchtigkeit bei den Arable-Daten als Dezimalzahl ausgedrückt, wohingegen sie innerhalb der DWD-Daten direkt als Prozentangabe eingetragen sind. Um grafische Vergleiche zwischen diesen beiden Datensätzen zu ermöglichen, werden die relativen Luftfeuchtigkeitswerte der Arable-Stationen zu diesem Zweck mit 100 multipliziert.

### 3.2 Datenqualitätsüberprüfung

Die Standardisierung der Erhebung von Wetter- und Klimadaten ist ein Prozess, der mit der ersten Edition des *Guide to Meteorological Instruments and Methods of Observation* im Jahr 1954 von der WMO angestoßen wurde (siehe: WMO 2018). Auf Fragen wie „Mit welchen Geräten kann welche Klimavariablen gemessen werden?“, „Wie erhält man diese Daten?“ und „Wo und wie errichtet man eine Wetterstation?“ findet man in der Literatur zahlreiche Antworten (siehe: WMO 1993; WMO 2019). Anders verhält es sich mit der Suche nach Methoden zur Überprüfung der Datenqualität solcher Datensätze. Die nachfolgend vorgestellten Methoden, die in dieser Arbeit angewendet werden, stellen eine Zusammenstellung basierend auf Gesprächen mit Experten, die in diesem Feld tätig sind,

dar, die von Literatur des DWD und der WMO getragen werden. Dieses literarische wie Experten basierte Wissen dient während der Überprüfung der Datenqualität der Einhaltung von etablierten Standards.

Das Ziel der Methodendurchführung soll ein Datensatz sein, der die Qualität der Werte der Klimavariablen *Lufttemperatur* (*t\_air*), *Niederschlag* (*precip*) und *relative Luftfeuchtigkeit* (*rel\_hum*) mit sogenannten *quality flags* darstellt. Für jede Variable eines jeden Wertes soll ein solcher Qualitätsflaggingsschlüssel vergeben werden. Dieser ergibt sich auf Basis mehrerer Tests. Die Durchführung dieser Überprüfungsverfahren und Bestimmung des Qualitätsschlüssels soll anhand eines selbst programmierten R-Paketes und Ausführungsskripte erfolgen, die den zur Verfügung gestellten Arable-Datensatz semi-automatisiert überprüfen. Nachfolgend werden die den Tests zugrundeliegenden Methoden und wie diese innerhalb des R-Paketes definiert sind, erläutert. Dabei werden einzelne Codebeispiele vorgestellt, jedoch aus Platzgründen nicht das gesamte Skript schrittweise durchgegangen. Die Funktionsweise dieser lässt sich anhand der Kommentare innerhalb der Einzelskripts, die gemeinsam mit dieser Arbeit abgegeben werden, nachvollziehen.

Nach der Beschreibung der einzelnen Testverfahren und ihren methodischen Hintergründen wird die Erstellung des allgemeinen Qualitätsschlüssels für die drei Klimavariablen aufgeschlüsselt. Der Abschluss dieses Kapitels stellt die Beschreibung der Erstellung des R-Paketes und der R-Skripte dar, die einen fundamentalen Anteil an der methodischen Durchführung ausmachen. Dabei wird auf die Skripte beider methodischen Schwerpunkte (Datenqualitätsüberprüfung der Arable-Mark-2 Daten und die Vergleichsanalyse) eingegangen, da sich diese in ihrer Durchführung voneinander unterscheiden.

Die nachfolgenden Methodendarstellungen 3.2.1 bis einschließlich 3.2.6 sind jeweils immer zweigeteilt. Der erste Teil des jeweiligen Unterkapitels beschreibt die Methode als solche und welche Richtwerte und -linien angewandt werden, wohingegen der zweite Teil die Umsetzung innerhalb der R-Skripte thematisiert, die allesamt im programmierten R-Paket unter dem Verzeichnis **R** zu finden sind. Diese ausgewählten Methoden sind nur eine Auswahl an einer Vielzahl an Testverfahren zur Überprüfung der Datenqualität (siehe: WMO 2021). Die Entscheidung dahin gehend, welche der einzelnen Verfahren angewendet werden, ergab sich auf Basis des Ausgangsdatensatzes, der aufgrund seiner temporären Auflösung nur einen Teil der Methoden zulässt. Dieser zentrale Aspekt wird im Unterkapitel 5.1 näher beleuchtet und diskutiert.

### 3.2.1 Fehlwertüberprüfung

Die Überprüfung des Datensatzes auf Fehlwerte ist ein wichtiger, erster Schritt zur Überprüfung der Datenqualität (Spengler 2002, S. 246). Liegt ein Wert nicht vor, kann dieser nicht überprüft werden. Wie mit diesen Werten umgegangen werden soll, zum Beispiel, ob der fehlende Stundenwert aus der Zeitreihe gestrichen oder durch vorherige und nachfolgende Werte interpoliert wird, ist an dieser Stelle nicht von Bedeutung. Alles, was in diesem ersten Schritt passiert, ist, dass das Fehlen oder Vorhandensein eines Fehlerwertes im Datensatz in einer neuen Spalte des R-Arbeitstabellendokumentes mit einem Schlüssel codiert wird. Dieser Schlüssel kann entweder eine **1** für vorhanden, **-777** für fragwürdig oder **-999** für einen Fehlwert sein.

Innerhalb des R-Paketes wird dieser Schritt mit dem Skript *qcFlag\_NA.R* umgesetzt. In diesem wird eine Funktion *qcFlag\_NA* programmiert, die einen Eingangsdatensatz *x* benötigt und diesen um die drei Spalten *qcFlag\_na\_t\_air*, *qcFlag\_na\_precip* und *qcFlag\_na\_rel\_hum* erweitert. Um diese Spalten mit dem jeweiligen Datenqualitätsschlüssel versehen zu können, wird eine Schleife programmiert, die diese zuallererst mit **1** codiert, um die Menge an Abfragen zu reduzieren. Anschließend wird überprüft, ob für die Messung eines jeweiligen Stundenwertes der drei zu prüfenden Klimavariablen ein Fehlwert vorkommt oder nicht. Wenn ja, dann wird in die jeweilige Flagging-Spalte (wie *qcFlag\_na\_t\_air*) auf der Ebene des jeweiligen Stundenwertes der Wert **-999** eingetragen.

```
x[[i]]$qcFlag_na_precip[is.na(x[[i]]$precip)] <- -999
```

Der obenstehende Codeausschnitt stellt diesen Arbeitsschritt beispielhaft für den Niederschlag dar. Innerhalb des Datensatzes *x* wird eine beliebige Zelle *i* der Spalte *qcFlag\_na\_precip* mit dem Wert **-999** gefüllt, wenn in der korrespondierenden Spalte *precip* der zugehörige *i*-te Wert ein Fehlwert ist. *i* entspricht hierbei der Anzahl der Zeilen des Datensatzes und ermöglicht es dank einer Schleife jeden einzelnen Niederschlagswert nacheinander zu überprüfen. Die Abfrage per *is.na()* ist dadurch möglich, dass Fehlwerte innerhalb des Arable-Datensatzes als *NA* oder auch leere Zeile codiert werden. Eine Änderung dieses Umstandes würde ein Umschreiben dieser einleitenden Funktion erfordern.

Den Abschluss der Funktion bildet eine zweite Schleife, die über den Wert *j* gesteuert wird und es ermöglicht, exakt jeden Einzelwert einer Station zu untersuchen und mit seinen zugehörigen Stundenwerten der anderen Klimavariablen zu vergleichen. Angewendet wird diese Schleife, weil der Datensatz für die Niederschlagswerte eine Besonderheit aufweist. In seltenen Fällen wird für den Niederschlag der Wert **0** ausgegeben, während alle anderen Sensoren des gesamten Datensatzes Fehlwerte enthalten. Da an dieser Stelle unklar ist, ob der Niederschlagssensor tatsächlich etwas gemessen hat oder diese Nullwerte von Arable eingetragen wurden, werden diese besonderen Niederschlagsnullwerte mit **-777** geflaggt, wenn zeitgleich in *qcFlag\_na\_t\_air* und

`qcFlag_na_rel_hum` der Wert **-999** vorzufinden ist. Dabei muss der bereits codierte Schlüssel innerhalb von `qcFlag_na_precip` ungleich **-999** sein; er wurde somit mit **1** codiert.

Diese Abfrage erfolgt über den Befehl `ifelse()`, der den Zelleninhalt eines jeden Niederschlagstundenwertes mit **-777** ersetzt, wenn die so eben beschriebene Bedingung erfüllt ist. In codierter Form sieht diese Funktion wie folgt aus und besagt: ein Zellenwert innerhalb der Spalte `qcFlag_na_precip` bekommt den Wert **-777** zugewiesen, wenn der jeweils entsprechende Wert in `qcFlag_na_t_air` und `qcFlag_na_rel_hum` **-999** beträgt, wobei der jeweils entsprechende Wert der Spalte `qcFlag_na_precip` nicht **-999** sein darf:

```
x[[i]]\$qcFlag_na_precip[j] <-
  ifelse(x[[i]]$qcFlag_na_t_air[j]==-999
        & x[[i]]$qcFlag_na_rel_hum[j]==-999
        & x[[i]]$qcFlag_na_precip[j]!=-999,
        -777,
        x[[i]]$qcFlag_na_precip[j])
```

Aufgrund dieser genauen Abfrage wird verhindert, dass fälschlicherweise korrekt klassifizierte Flaggingwerte mit **-777** überschrieben werden. Abschließend werden die Werte und Spalten mit `return(x)` ausgegeben.

### 3.2.2 Technische Überprüfung: Sensor Range Test

Die technische Überprüfung der gemessenen Werte durch den Arable-Mark-2 erfolgt anhand eines Sensor-Range-Tests; in manchen Werken auch *plausible value check* genannt (Zahumenský 2002, S. 5). Das Ziel dieses Tests ist es, zu überprüfen, ob die gemessenen Werte plausibel sind. Als Basis für diese Einschätzung werden die Messbereiche und Ungenauigkeiten der Messsensorik herangezogen (Zahumenský 2002, S. 5; WMO 2021, S. 23). Wenn ein Wert diesen Test nicht besteht, muss er weiteren Testverfahren oder der kritischen Auseinandersetzung des Nutzers unterzogen werden, bevor er für die Berechnung anderer Parameter oder ähnliches verwendet werden kann (Zahumenský 2002, S. 5). Als Grundlage für die Einschätzung der Plausibilität werden die Werte aus Tabelle 1 auf Seite 20 herangezogen (Arable-Labs 2021, S. 1).

Die methodische Umsetzung erfolgt mit dem Skript `qcFlag_sensorRange.R` des programmierten R-Paketes. Dieses Skript erzeugt sechs Qualitätsschlüssel, wobei einer nur zu Zwecken der Durchführung vergeben wird. Diese Werte sind:

- **1** für valide
- **-999** für Fehlwerte
- **-888** für Werte, die außerhalb des Messbereiches liegen
- **-887** für Werte, die aufgrund von Sensorungenauigkeiten innerhalb des plausiblen Messbereiches liegen könnten
- **-886** für Werte, die im nicht kalibrierten Bereich des Temperaturmesssensors liegen und somit höheren Messungenauigkeiten von bis zu  $\pm 1,5^\circ \text{C}$  ausgesetzt sind (Arable-Labs 2021, S. 1)
- **-555** besagter temporärer Wert, der für die Durchführung der Funktion vergeben wird

Die Funktion für den Sensor Range Test ist in vier Teile untergliedert. Der erste Teil dient der grundlegenden Kalibrierung der drei zu erstellenden Tabellenspalten des Qualitätstests. Um Rechenkosten zu sparen, wird *qcFLag\_sensorRange\_t\_air* mit **-886** codiert, während die Spalten für den Niederschlag und die relative Luftfeuchtigkeit den temporären Wert **-555** zugewiesen bekommen. Nach dieser Vorbereitungsphase werden drei Schleifen aufgerufen, die nacheinander die drei Klimavariablen basierend auf den Werten aus Tabelle 1 codieren.

Alle Werte, die bei der Lufttemperatur zwischen  $-18^\circ$  und  $45^\circ$  Celsius liegen, werden mit **1** codiert. Diese Werte gelten als valide und sollten lediglich um bis zu  $\pm 0,8^\circ$  Celsius von der realen Lufttemperatur abweichen (Arable-Labs 2021, S. 1). Da die Spalte im ersten Schritt mit **-886** vorcodiert wurde, erhalten alle Werte, die außerhalb dieses Bereiches liegen, diesen Qualitätsschlüssel. Abschließend erfolgt die Codierung potenzieller Fehlwerte innerhalb der Spalte *t\_air*. Diese Werte erhalten den Wert **-999**, der identisch mit der Qualitätskontrolle des vorangegangenen Tests sein muss.

Für den Niederschlag muss ein etwas anderes Prinzip angewendet werden, da hier ein tatsächlicher Mess- und kein Kalibrierungsbereich benannt ist (Arable-Labs 2021, S. 1). Dieser Messbereich reicht von 0 bis 100 mm pro Stunde und Werte, die innerhalb dieses liegen, werden als valide betrachtet. Zusätzlich wird die Genauigkeit (*accuracy*) berechnet, die bei 6 % liegt (Arable-Labs 2021, S. 1). Das bedeutet, dass alle Werte bis einschließlich 106 mm als plausibel angesehen werden können. Diese Werte werden mit **-887** codiert, um auf diesen Umstand hinzuweisen. Werte, die selbst mit Berücksichtigung der Sensorungenauigkeit nicht in den Messbereich fallen (Werte über 106 mm), werden mit **-888** codiert. Fehlwerte werden vergleichbar, mit der der Lufttemperaturüberprüfung behandelt.

Die Überprüfung der relativen Feuchtigkeit verhält sich wie jene der Temperatur. Da der Messbereich nur zwischen 0 und 100 % (beziehungsweise 0 bis 1) liegen kann, werden alle Werte als valide erklärt, die in diesem Bereich fallen. Jedoch gilt für alle diese Werte, dass sie um bis zu  $\pm 5\%$  von der Realität abweichen können (Arable-Labs 2021, S. 1). Obwohl der vorgegebene Messbereich

für die relative Luftfeuchtigkeit strikt vorgegeben ist, wird anschließend eine Funktion aufgerufen, die untersucht, ob Werte außerhalb dieses Bereiches fallen. Diese Werte werden automatisch mit **-888** codiert und müssen vom Nutzer des Datensatzes kritisch betrachtet werden. Fehlwerte werden wieder mit **-999** codiert. Nach der Durchführung dieser Methode enthält der Eingangsdatensatz drei neue Spalten, die die Qualitätsschlüssel der einzelnen Klimavariablen beinhalten.

### 3.2.3 Klimatische Überprüfung: Plausible Values

Der *Plausible Values* oder auch *Climate-based range* Test testet, angelehnt an Zahumenský (2002, S. 5), ob ein Einzelwert basierend auf klimatologischen Grenzen plausibel erscheint. Diese Schwellenwerte können grob gefasst werden, um die Klimavariablenschwankungen über ein Jahr zu fassen oder auch spezialisiert für die jeweilige Jahreszeit angepasst werden (WMO 2021, S. 39). Idealerweise werden diese Grenzwerte auf Basis von Klimadaten der jeweiligen Station festgelegt, die mindestens einen Zeitraum von 30 Jahre abdecken, um die klimatischen Bedingungen vor Ort zu erfassen (WMO 2021, S. 39). Wenn dies nicht möglich ist, können diese Grenzwerte allgemein definiert werden (Zahumenský 2002, S. 6) oder anhand einer ortsnahen Station, die einen entsprechenden Zeitraum abdeckt, ermittelt werden (WMO 2021, S. 39).

Da die mobilen Wetterstationen des Arable-Datensatzes über ein räumlich großes Gebiet erstreckt sind und sich die jeweiligen Standorte jederzeit ändern können, wurden, um die zukünftige Wartung und Anpassung des Skriptes zu minimieren, die folgenden Grenzwerte angesetzt:

- **Lufttemperatur:** -30° bis +40° Celsius (Zimmermann et al., 2015, S. 2)
- **Relative Luftfeuchtigkeit:** 0 - 100 % (Zahumenský 2002, S. 6)
- **Niederschlag:** bis zu 40 mm pro Stunde (DWD 2022a)

Diese Grenzwerte orientieren sich stark an den Warnkriterien des DWD und überprüfen den Datensatz auf Extremereignisse hin. Temperaturen von unter 40 Grad werden vom DWD als extreme Wärmebelastung (DWD-Warnstufe 2) und Niederschläge bis 40 mm pro Stunde als heftiger Starkregen (DWD-Warnstufe 3) klassifiziert (DWD 2022a). Temperatur- oder Niederschlagswerte, die diese Grenzwerte überschreiten, liegen zwar im Bereich des möglichen für Stundenwerte, stellen jedoch extreme Wetterereignisse dar und müssen vom Anwender der Daten dahin gehend überprüft werden, ob ein solches Ereignis tatsächlich vorlag (anhand eines Vergleiches mit anderen Stationen und Datensätzen) oder ob es sich hierbei um einen zweifelhaften Messwert handelt (Zimmermann et al., 2015, S. 2). Selbst die Anwendung von Grenzwerten, die anhand der maximalen, beziehungsweise minimalen Ausprägung der Klimasignale über einen Zeitraum von 30 Jahren festgelegt werden, kann aufgrund des globalen Klimawandels dazu führen, dass Extremereignisse der Zukunft als fragwürdig geflaggt werden (WMO 2021, S. 39). Das Ziel dieses Tests ist es nicht, zu überprüfen, ob

ein Messwert im Bereich des Möglichen liegt, sondern um zu überprüfen, ob er basierend auf dem ortstypischen Klima der letzten 30 Jahre oder anderweitig festgelegten klimatischen Grenzwerten plausibel erscheint (WMO 2021, S. 39; Zahumenský 2002, S. 6).

Die Durchführung dieser Methode erfolgt mit dem Skript *qcFlag\_plausibleValues.R*. Dieses Skript setzt sich aus zwei Schleifen zusammen, einer äußeren und einer inneren. Die äußere Schleife dient der Vorbereitung der Tabellenspalten *qcFlag\_plausibleValues\_t\_air*, *[... ]\_precip* und *[... ]\_rel\_hum*, deren Zellen den Platzhalterwert **-555** zugewiesen bekommen. Die innere Schleife berechnet für jeden Wert der zu überprüfenden Klimavariablen einer jeder Station, ob dieser in den definierten Grenzen liegt. Begonnen wird dabei mit der Lufttemperatur. Wenn die Werte zwischen  $-30^{\circ}$  und  $+40^{\circ}$  Celsius liegen, werden sie mit **1** für valide codiert, fragwürdige Werte von Ausreißern, die auf Extremereignissen zurückzuführen sein könnten, erhalten den Wert **-777** und Fehlwerte den Schlüssel **-999**. Diese Abfrage der validen und der potenziellen invaliden Werte erfolgen per *ifelse()* Befehle, die besagen, wenn ein Wert *x* zwischen zwei Werten *y* und *z* liegt, dann bekommt er den Schlüssel **1** oder **-777** zugewiesen. Die Fehlwertabfrage erfolgt über den Befehl *is.na()*. Diese drei Funktionen werden in angepasster Form auf die drei Klimavariablen angewendet und ihre Werte mit *return(x)* ausgegeben.

Der nachfolgende Code stellt diesen Aspekt der Codierung für die Lufttemperatur anhand des Flagging der validen Werte dar:

```
x[[i]]$qcFlag_plausibleValue_t_air[j] <-
  ifelse(x[[i]]$t_air[j]>=(-30)
    & x[[i]]$t_air[j]<=40,
    1,
    x[[i]]$qcFlag_plausibleValue_t_air[j])
```

In die jeweilige Zeile der Spalte *qcFlag\_plausibleValue\_t\_air* wird der Wert **1** geschrieben, wenn der Messwert der Spalte *t\_air* für eine spezifische Station *j* zwischen größer gleich  $-30^{\circ}\text{C}$  und kleiner gleich  $40^{\circ}\text{C}$  liegen. Der Zusatz *x[[i]]\$qcFlag\_plausibleValue\_t\_air[j]* zum Abschluss der Funktion muss angegeben werden, damit die bereits codierten Werte der anderen Zeilen der Spalte *qcFlag\_plausibleValue\_t\_air* nicht überschrieben werden können.

### 3.2.4 Innere Konsistenzprüfung

Die innere Konsistenzprüfung ist eine Methode, die verwendet wird, um einen Klimadatensatz auf seine innere Logik zu überprüfen. Auf Basis der gegebenen Klimavariablen des Arable-Datensatzes und den in der Literatur gefundenen Variablenzusammenhänge wird die Lufttemperatur auf die nachfolgenden Zusammenhänge überprüft:

1. Liegt die gemessene Lufttemperatur zwischen dem zur gleichen Zeit verzeichneten Maximum und Minimum der Lufttemperatur (WMO 1993, S. V1.12; Cerlini et al., 2020, S. 6)?
2. Ist der Taupunkt der Temperatur kleiner oder gleich der gemessenen Lufttemperatur (Zahumenský 2002, S. 7; Cerlini et al., 2020, S. 6)?

Für den Niederschlag und die relative Luftfeuchtigkeit wurden innerhalb der Literatur zwar Zusammenhänge zwischen diesen und anderen Variablen gefunden, jedoch können diese auf den Arable-Datensatz aufgrund des Fehlens dieser Überprüfungsvariablen nicht durchgeführt werden. Ein solcher Zusammenhang wäre zum Beispiel für den Niederschlag der Vergleich mit einer Variable, die angibt, welcher Niederschlagstyp vergeben wurde (Zahumenský 2002, S. 7). Wird etwa ein Niederschlagstyp wie *heftiger Niederschlag* bei einem Niederschlagswert von 0 mm ausgegeben, müsste dieser Wert als fragwürdig markiert werden (Zahumenský 2002, S. 7).

Die Überprüfung der genannten Bedingungen für die Lufttemperatur erfolgt mit dem Skript *qcFlag\_internalConsistency.R*. Damit es durchgeführt werden kann, sind die Variablen *t\_min*, *t\_max* und *t\_dew* vonnöten. Das Skript setzt sich, wie die vorausgegangenen, aus zwei Schleifen zusammen, wovon die äußere in diesem Fall nacheinander die einzelnen Stationen durchgeht und die innere Schleife sich auf die Werte, der durch die äußere Schleife definierten Station bezieht. Um zu überprüfen, ob der gemessene Wert der Lufttemperatur zwischen dem Maximum und dem Minimum dieser liegt, wird der nachfolgende Code verwendet:

```
x[[i]]$[...]t_air_vs_max_min[j] <-
  between(x[[i]]$t_air[j], x[[i]]$t_min[j], x[[i]]$t_max[j])
```

Dieser Code fragt ab, ob *t\_air* wertetechnisch zwischen *t\_min* und *t\_max* liegt und schreibt das Ergebnis dieser Überprüfung in die jeweilige gerade zu überprüfende Zelle der Spalte *qcFlag\_internalConsistency\_t\_air\_vs\_max\_min*. Dieses Ergebnis lässt nur zwei Ausprägungen zu: wahr (*TRUE*) oder falsch (*FALSE*). Nachfolgend werden zwei *ifelse()* Befehle durchgeführt, die jenen der Funktion *qcFlag\_plausibleValues(x)* entspricht, und verwendet werden, um die Ausprägungen *TRUE* und *FALSE* in Qualitätsschlüssel umzuwandeln. Wenn das Ergebnis der *between*-Abfrage mit *TRUE* codiert wird, so wird dieser Wert zu den Qualitätsschlüssel **1** umgeschrieben; ist das Ergebnis *FALSE*,

dann erhält er den Qualitätsschlüssel **-777** für fragwürdig. Die Fehlwertabfrage wird abschließend per *is.na()* ermittelt.

Der Aufbau der Abfrage bezüglich des Taupunktes und der Lufttemperatur folgt dem gleichen Schema. Jedoch werden die *TRUE* und *FALSE* Werte anhand des nachfolgenden Codes einleitend ermittelt:

```
x[[i]]$qcFlag_internalConsistency_t_air_vs_t_dew[j] <-  
  x[[i]]$t_dew[j] <= x[[i]]$t_air[j]
```

Wenn *t\_dew* kleiner oder gleich *t\_air* ist, wird in die jeweilige Zelle der Spalte *qcFlag\_internal Consistency\_t\_air\_vs\_t\_dew* der Wert *TRUE* geschrieben. Ist dem nicht so, ist das Ergebnis *FALSE*. Die anschließende Codierung dieser Ergebnisse ist identisch mit dem Aufbau der ersten Bedingung.

### 3.2.5 Zeitliche Konsistenzprüfung

Die zeitliche Konsistenzprüfung ist eine Methode, die einen Wert *x* mit seinem vorangegangenen Wert vergleicht und überprüft, ob die Differenz zwischen ihnen einen festgelegten Grenzwert überschreitet (Zahumenský 2002, S. 6). Ist dies der Fall, wird dieser Wert als fragwürdig markiert (Zahumenský 2002, S. 6). Je nach Literatur werden unterschiedliche Grenzwerte genannt. Für die Lufttemperatur wird diesbezüglich ein Grenzwert von  $\pm 4^\circ$  Celsius (Cerlini et al., 2020, S. 6) und für die relative Luftfeuchtigkeit ein Wert von 10 % also 0.1 angesetzt (Zahumenský 2002, S. 6). Wichtig zu wissen ist, dass ein solcher Test die meteorologische Varianz von Klimasignalen über die Zeit nicht berücksichtigt und somit bei der Betrachtung von sehr großen Zeiträumen an Bedeutung verlieren kann (WMO 2021, S. 40). Ferner müssen diese Grenzwerte gegebenenfalls aufgrund zukünftiger Klimaveränderungen angepasst werden, wenn diese von den gegenwärtigen Klimabedingungen abweichen (WMO 2021, S. 40). Zusätzlich kann der Niederschlag für dieses Testverfahren nicht berücksichtigt werden. Im Gegensatz zu der Temperatur oder der relativen Luftfeuchtigkeit nimmt die Niederschlagsmenge nicht kontinuierlich zu, sondern tritt je nach Ereignis spontan auf (Platzregen) (DWD 2022c). Diese Methode setzt jedoch voraus, dass ein Anstieg oder auch ein Abfall der Werte kontinuierlich vonstattengeht und nicht plötzlich (WMO 2021, S. 40).

Wie auch die vorangegangenen Methoden wird auch diese in dem programmierten R-Paket als Funktion zur Verfügung gestellt. Diese Funktion *qcFlag\_timeConsistency()* ist in dem identisch benannten R-Skript definiert und setzt sich wie folgt zusammen: Zuerst wird die äußere Schleife definiert, die es ermöglicht, jede Station getrennt voneinander und nacheinander zu untersuchen. Gesteuert wird diese Schleife über die Funktion:

```
i in 1:(length(x))
```

wobei  $i$  für eine beliebige Station steht.

Die innere Schleife hingegen wird wie auch bei den vorangegangenen Funktionen mit dem nachfolgenden Befehl gesteuert:

```
j in 1:(dim(x[[i]]))[1])
```

Dieser Befehl besagt, dass nacheinander eine jede Zelle  $j$  einer bestimmten Station  $i$  beginnend bei 1 bis zum letzten, gemessenen Zellenwert durchlaufen werden soll. Das Ermitteln der Anzahl aller Zellen einer Station wird über den `dim(x[[i]])[1]` Befehl geregelt, der überprüft, wie viele Einträge in dem Datensatz einer jeden Station vorkommen (RDocumentation 2022b).

Beginnend mit dem allerersten erhobenen Wert aller Werte einer Station wird direkt eine Problematik aufgeworfen: Es ist kein vorheriger Wert vorhanden. Deswegen berechnen die Befehle in Zeile 24 und 43 im Skript nicht den derzeitigen und den vorherigen Wert, sondern den nachfolgenden und den derzeitigen Wert und speichern diese in temporäre Tabellenspalten ab:

```
x[[i]]$temporary_(t_air[j] )<- x[[i]]$t_air[j+1] - x[[i]]$t_air[j]
```

Im nächsten Schritt wird für die zu untersuchenden Klimavariablen eine weitere temporäre Tabellenspalte erstellt, deren Zellen mit **-777** vorcodiert werden, um Rechenschritte zu ersparen. Die eigentliche Vergabe der Qualitätsschlüssel findet ab Zeile 31 beziehungsweise 50 statt und erfolgt per `ifelse()` Abfrage:

```
x[[i]]$temporary_(qcflag_(t_air[j] ) )<-
  ifelse(x[[i]]$temporary_(t_air[j] )>=(-4)
    & x[[i]]$temporary_t_air[j]<4,
    1,
    x[[i]]$temporary_qcflag_t_air[j])
```

Wenn ein Wert in der Tabellenspalte `temporary_t_air` zwischen  $-4^{\circ}$  und  $+4^{\circ}$  Celsius liegt, so wird dieser mit **1** codiert. Alle Werte, die dieser Abfrage nicht entsprechen, behalten den vorcodierten Wert **-777** und müssen nachfolgenden Betrachtungen unterzogen werden. Abgeschlossen wird die Bestimmung der Qualitätsschlüssel mit der bereits bekannten Fehlwertüberprüfung per `is.na()`.

Damit ist die innere Schleife fertig programmiert, jedoch besteht noch immer die Problematik, dass die codierten Qualitätsschlüssel um eine Zeile nach oben versetzt sind und nicht ihrem jeweiligen Tabellenwert zugeordnet sind, da für die Berechnung der Qualitätsschlüssel der nachfolgende Wert mit dem derzeitigen verglichen wurde. Um diese Problematik zu lösen, wird die nachfolgende Funktion für die Klassifikationsspalten beider Klimavariablen angewendet:

```
x[[i]]['qcFlag_timeConsistency_t_air'] <-
  c(-999, head(x[[i]]['temporary_qcflag_t_air'], dim(x[[i]])[1] - 1)[[1]])
```

Diese Funktion verschiebt die Werte der Spalte *temporary\_qcflag\_t\_air* beziehungsweise *[...]*\_rel\_hum um genau eine Zeile nach unten. Dadurch werden die Qualitätsschlüssel den richtigen Werten zugeordnet. Gleichzeitig wird der alle erste Spalteneintrag mit **-999** codiert, da dieser aufgrund des fehlenden vorangegangenen Wertes nicht berechnet werden kann. Diese modifizierte temporäre Spalte wird schließlich in die abschließende *qcFlag\_timeConsistency\_t\_air* beziehungsweise *[...]*\_rel\_hum überführt und ausgegeben. Die verbliebenden temporären Tabellenspalten werden per *Select* Abfrage und dem Befehl *subset()* entfernt und tauchen im getesteten Datensatz nicht auf. Die soeben beschriebene Definierung des Qualitätsschlüssels, die den nachfolgenden und den derzeitigen Wert miteinander vergleicht und das Ergebnis am Ende um eine Zeile nach unten verschiebt, damit die Qualitätsschlüssel den richtigen Werten zugeordnet werden, ist historisch erwachsen und ergab sich zwangsweise aus Verhaltensweisen der Programmiersprache R. So führt etwa die Ermittlung der temporären Werte aus dem vorangegangenen mit dem derzeitigen Stundenwert immer zum Abbruch der Funktion, weil für den allerersten Messwert kein vorheriger gegeben ist. Die einzige Lösung, die gefunden werden konnte, um dieses Problem zu umgehen, ist die soeben vorgestellte.

### 3.2.6 Erstellung des allgemeinen Qualitätsschlüssel

Nach der Durchführung der Einzeltests muss ein allgemeiner Qualitätsschlüssel für die drei Klimavariablen generiert werden, dessen Zweck es ist, einem Nutzer des geflaggtten Datensatzes die Möglichkeit zu geben, auf einen Blick erkennen zu lassen, welche Werte als valide deklariert wurden, welche als fragwürdig gelten und welche auf Fehlwerten beruhen. Die Qualitätsschlüssel sind dabei wie folgt zu deklarieren:

- **1** = valide, alle Tests wurden bestanden
- **- 777** = fragwürdig, nicht jeder Test wurde bestanden
- **- 999** = Fehlwert, kein Test wurde bestanden

Umgesetzt wird dieser letzte Schritt mit zwei R-Skripten, wovon das eine zu Beginn der Untersuchung ausgeführt wird und das zweite am Ende. *qcFlag\_initializing.R* erzeugt drei leere Tabellenspalten mit den Namen *qcFlag\_totalScore\_t\_air*, *qcFlag\_totalScore\_precip* und *qcFlag\_totalScore\_rel\_hum*. Diese Tabellenspalten sollen innerhalb des zu erstellenden überprüften Flagging-Datensatzes vor den Testergebnisspalten stehen. Das R-Skript *qcFlag\_totalScore.R* befüllt diese nach Abschluss der Testverfahren. Der generelle Aufbau der Funktion *qcFlag\_totalScore()* gleicht im Kern

den vorangegangenen: Es wird eine äußere wie eine innere Schleife deklariert, wobei die äußere zu Beginn verwendet wird, um die drei genannten *qcFlag\_totalScore* Spalten mit dem Wert **-777** vorzucodieren. Die eigentliche Berechnung erfolgt innerhalb der inneren Schleife und ist in drei Abschnitte unterteilt, die sich nach den drei Klimavariablen *Lufttemperatur*, *Niederschlag* und *relative Luftfeuchtigkeit* richten.

Im ersten Schritt wird per *ifelse()* Abfrage überprüft, ob alle Testergebnisse einer Klimavariablen als valide (1) codiert wurden (siehe Zeile 25-32, 60-64 und 89-94 im Skript *qrFlag\_totalScore.R*). Wenn ja, dann wird der Wert **1** in die jeweilige Zelle der jeweiligen *qcFlag\_totalScore* Spalte geschrieben. Nachfolgend unterscheiden sich die Einzelfunktionen der Klimavariablen. Für die Lufttemperatur wird als nächstes überprüft, ob ein Wert in der Spalte *qcFlag\_timeConsistenz\_t\_air* den Schlüssel **-999** vorweisen kann. Ist dem so, wird der zugehörige *totalScore* auf **-777** gesetzt, da dieser Test für diesen Wert nicht bestanden wurde. Die zweite für die Lufttemperatur besondere Abfrage ergibt sich aus dem *Sensor-Range* Test. Wurde ein Wert in diesem mit **-886** codiert, dann wird der *totalScore* ebenfalls auf **-777** gesetzt, um darauf hinzuweisen, dass dieser Wert außerhalb des kalibrierten Messbereiches des Sensors liegt und somit größere Ungenauigkeiten aufweist. Die abschließende Funktion (Zeile 48) teilen sich alle drei Klimavariablen und überprüft, ob Fehlwerte vorkommen. Wenn in einem der Testergebnisse einer vorkommt, dann wird der jeweilige Wert der jeweiligen *totalScore* Spalte auf **-999** gesetzt.

Abweichungen für den Niederschlag und die relative Luftfeuchtigkeit sind begründet in den unterschiedlichen Qualitätsschlüsseln, die innerhalb des *Sensor-Range* Tests vergeben werden. Für den *totalScore* des Niederschlages werden alle Werte, die mit **-887** (Messwert ist innerhalb des Messbereiches dank Sensorungenauigkeiten) mit **1** und alle Werte, die den Qualitätsschlüssel **-888** aufweisen (Wert liegt außerhalb des Messbereiches) mit **-777** klassifiziert. Letzteres gilt auch für den abschließenden Qualitätsschlüssel der relativen Luftfeuchtigkeit. Zusätzlich wird für den Niederschlag überprüft, ob in der Spalte *qcFlag\_na\_precip* fragwürdige Werte mit **-777** hervorgehoben wurden. Wenn ja, dann wird der *totalScore* für den Niederschlag auf **-777** gesetzt.

Wichtig ist an dieser Stelle zu betonen, dass ein Stundenwert, der einen *totalScore* von **-777** aufweist, nicht zwangsläufig falsch sein muss. Dieser Wert kann aufgrund eines mit dem Stundenwert festgehaltenen Extremereignis oder vergleichbares zustande gekommen sein und bedarf die Überprüfung des Anwenders. Wie mit den **-999** Werten umgegangen werden soll, ist ebenfalls vom Anwender abhängig.

Im nachfolgenden Abschnitt wird erläutert, wie die vorgestellten Funktionen als R-Paket verpackt werden, wie sie aktualisiert werden können und wie sie ausgeführt werden.

### 3.2.7 Erstellung des R-Pakets

Um zu gewährleisten, dass die Funktionen in Zukunft leicht angepasst werden können und um die Übersicht bei deren Sichtung zu gewährleisten, wurden sie als ein R-Paket zusammengefasst. Dieses R-Paket ist im Abgabeordner unter *Abgabedateien* -> *Datenqualitätsüberprüfung\_Arable* -> ***arableQualityControlFlagging*** zu finden und beinhaltet neben den Dateien *NAMESPACE.dat* und *DESCRIPTION.dat*, ein RStudio-Projekt namens *arableQualityControlFlagging.Rproj* und die zwei Verzeichnisse ***R*** und ***man***. In dem Verzeichnis ***R*** befinden sich alle in diesem Kapitel benannten R-Skripte, die jederzeit bearbeitet werden können. Das Verzeichnis ***man*** hingegen stellt für jedes einzelne R-Skript die notwendige Dokumentation zur Verfügung, die mit dem Befehl *help()* innerhalb von R oder R-Studio aufgerufen werden können. Diese Dokumentationen enthalten Informationen über den Aufbau und die Funktionsweise der jeweiligen Funktion (R-Project 2022).

Aufgrund der Komplexität bei der Erstellung eines R-Paketes, bedingt durch die notwendige Dokumentation dessen und dem generellen Aufbau eines R-Paketes, wurde dieser Arbeitsschritt mit dem freiverfügbaren Programm R-Studio umgesetzt (RStudio 2022a). Dieses Programm wurde auch für die Programmierung der Einzelskripts verwendet und erlaubt es, Projekte wie *arableQualityControlFlagging.Rproj* für R-Pakete zu erstellen, die es einem ermöglichen, ein Paket von dem Programm erstellen zu lassen. Für die Erstellung eines R-Paketes muss die zusätzliche Software *RTools* (<https://cran.rstudio.com/bin/windows/Rtools/>) heruntergeladen werden. Nach der erfolgreichen Übertragung der programmierten Funktionen in das Verzeichnis ***R*** und der Dokumentationen in das Verzeichnis ***man***, kann das Paket über den Reiter *Build* -> *Clean and Install* erstellt werden. Eine jede Änderung an den Skripten erfordert die erneute Durchführung dieses Arbeitsschrittes.

### 3.2.8 Erstellung und Aufbau der R-Skripte

Nach der Erstellung des R-Paketes müssen Skripte geschrieben werden, die den Arable-Datensatz automatisch herunterladen, die einzelnen Funktionen der Qualitätsüberprüfung durchführen und anschließend die geflaggtten Werte in die Datenbank übertragen. Die zwei programmierten Skripte lauten *qc\_flagging\_complete\_arable\_dataset.R* und *qc\_flagging\_only\_new\_values.R* und werden für unterschiedliche Zwecke verwendet. Ersteres dient dem Testen des gesamten Arable-Datensatzes. Dieses Skript ist historisch entwachsen und muss nur ausgeführt werden, wenn bewusst der gesamte Arable-Datensatz überprüft werden muss (beispielsweise bei Verlust des geflaggtten Datensatzes *arable\_data\_hourly\_flag*). Zu Beginn des Skriptes werden die notwendigen Bibliotheken/R-Pakete eingebunden und aktiviert. Dazu gehört auch das vorangegangene selbsterstellte Paket *arableQualityControlFlagging*. Im nächsten Schritt erfolgt die Initialisierung der Datenbankverbindung über die Variable *con* und dem Befehl *dbConnect()* ab Zeile 27 (RStudio 2022b). Der nächste Abschnitt dient

dem Herunterladen und Vorbereiten des Arable-Datensatzes. Dieser wird zu Teilen (lediglich die Variablen *'id\_data', 'station\_id', 'date', 'date\_time', 't\_air', 't\_max', 't\_min', 't\_dew', 'precip' and 'rel\_hum'* werden heruntergeladen) in der Variable *arable\_data\_hourly* temporär zwischengespeichert und in die Variable *arable\_data\_hourly\_temporary* überführt. Dieser Datensatz wird anhand der Variable *date\_time* sortiert und anschließend mit dem Befehl *split()* anhand der *station\_id* aufgeteilt. Der so generierte Datensatz setzt sich aus Unterdatensätzen zusammen, die jeweils die Werte einer Messstation beinhalten. Nur auf diese Weise ist es möglich, die Stationen getrennt voneinander zu untersuchen, ohne eine Vielzahl an Datensätzen generieren zu müssen.

Der nächste und vorletzte Abschnitt dient der Durchführung der einzelnen Funktionen auf den erstellten und zerteilten temporären Datensatz. Nacheinander ausgeführt werden die Funktionen: *qcFlag\_initializing()*, *qcFlag\_NA()*, *qcFlag\_sensorRange()*, *qcFlag\_plausibleValues()*, *qcFlag\_internalConsistency()*, *qcFlag\_timeConsistency()* und *qcFlag\_totalScore()*. Der auf diese Weise getestete komplette Datensatz wird im letzten Abschnitt des Skriptes mit dem *unsplit()* Befehl wieder in einen gemeinsamen Datensatz, der tabellarisch die Werte aller Stationen in einer Tabelle darstellt, in *arable\_data\_hourly\_flag* überführt. Diese Datei wird abschließend mit *dbWriteTable()* in die Datenbank hochgeladen, wobei eine bestehende Datei mit dem gleichen Namen überschrieben wird.

Um jedoch den Rechenaufwand und die Notwendigkeit, dass der gesamte bereits geflaggte Datensatz überschrieben wird, zu verringern und zu vermeiden, wird das zweite und benannte Skript *qc\_flagging\_only\_new\_values.R* ausgeführt. Dieses weicht in seiner Komplexität zwar von dem vorangegangenen Skript stark ab, ist jedoch in der Lage, die Berechnungen schneller und effizienter auszuführen und nur die neuen, noch nicht überprüften Werte der Datenbank hinzuzufügen. Der Beginn gleicht dabei dem anderen Skript. Zuerst werden die notwendigen Bibliotheken/Pakete geladen, die Datenbankverbindung hergestellt und der Arable-Datensatz anhand der gleichen *Select*-Abfrage heruntergeladen, um nur den benötigten Teil temporär zur Verfügung zu stellen. Zusätzlich wird jedoch der Datensatz *arable\_data\_hourly\_flag* heruntergeladen, der die bereits überprüften und geflaggtten Werte enthält. Dieser Datensatz wird benötigt, um ermitteln zu können, welche Werte bereits überprüft wurden, da *R* es nicht ermöglicht, solch komplexeren SQL-Abfragen direkt an die Datenbank zu senden.

Um diese Grenzwerte oder auch *Boundary Values* zu ermitteln, die widerspiegeln, welche Werte bereits überprüft wurden, wird der *arable\_data\_hourly* Datensatz mit *order()* anhand der Variable *date\_time* sortiert. Mit dem Befehl *aggregate()* wird anhand der *station\_id* ermittelt, wann zum letzten Mal für eine Station ein Messwert eingetragen wurde (*boundaryValues\_lastFlagged*).

Der nächste Abschnitt dient dem Generieren eines Datensatzes, der für jede Station das Datum und die Zeitangabe des ersten, noch nicht geflaggtten Wert enthält. Dafür werden die beiden Daten-

sätze *dataset\_arableDataHourly* und *dataset\_arableDataHourlyFlag* mit *rbind()* zusammengeführt und dieser Datensatz *combinedDataset* per *subset()* Befehl so beschnitten, dass nur die einmalig vorkommenden Zeilen in *combinedDataset\_uniqueValues* überführt werden. Dieser Datensatz wird anhand der *date\_time* sortiert und per *aggregate()* aggregiert. Der so erzeugte Datensatz *boundaryValues\_lastFlagged\_correction* enthält für jede Arable-Station das Datum und die Zeitangabe (*date\_time*) des ersten, nicht überprüften Wertes. Benötigt werden diese Werte für alle neuen, ungeprüften Stationen und für jene, die bereits überprüft wurden, aber neue, ungeprüfte Werte beinhalten, die zeitlich vor den geprüften liegen. Diese beiden *boundaryValues* Datensätze werden anschließend zu einem gemeinsamen Datensatz überführt:

```
combinedBoundaryValues <-  
  boundaryValues_lastFlagged %>%  
  anti_join(boundaryValues_lastFlagged_correction, by = "station_id") %>%  
  bind_rows(boundaryValues_lastFlagged_correction) %>%  
  arrange(station_id)
```

Der oben stehende Code erzeugt einen neuen Datensatz namens *combinedBoundaryValues*, der zuallererst mit den Werten aus *boundaryValues\_lastFlagged* befüllt wird und dann anschließend mit den Werten aus *boundaryValues\_lastFlagged\_correction* für die betroffenen Stationen überschrieben wird. Die Generierung dieses Datensatzes ist notwendig, da nicht jede Arable-Wetterstation über den gesamten Zeitraum in Betrieb ist und es vorkommen kann, dass frühere Messwerte dem Datensatz hinzugefügt werden (zum Beispiel vom 04.05.2022 für Station *xyz*) obwohl bereits spätere Messwerte für die gleiche Station vorliegen und getestet wurden. Diese beschriebenen Umstände verkomplizieren die Überprüfung der noch nicht getesteten Werte und machen den soeben beschriebenen Ablauf notwendig, der aufgrund der Einschränkungen der möglichen SQL-Abfragen nicht direkt von der Datenbank durchgeführt werden kann.

Neben der Generierung des Datensatzes *combinedBoundaryValues* wird auf die bereits beschriebene Art und Weise mithilfe von *aggregate()* der Datensatz *boundaryValues\_lastFlagged\_correction* generiert, der für jede Station das Datum und die Zeitangabe des zuallererst gemessenen Wert enthält. Dieser Datensatz und *boundaryValues\_lastFlagged* werden beide für die Funktion *function\_subsetByDateTime()* verwendet. *function\_subsetByDateTime()* dient dazu, den kompletten Arable-Datensatz auf jene Werte zu beschränken, die noch nicht überprüft wurden. Dafür wird eine Schleife programmiert, die nacheinander auf die einzelnen Stationen eine *if/else* Abfrage anwendet. Der *if* Teil der Abfrage überprüft, ob die stationsspezifischen *date\_time* Werte aus *combinedBoundaryValues* ungleich der Werte aus *boundaryValues\_firstFlagged\_dateTime* sind. Ist dem so, so wird ein *y*-Wert generiert, der sich aus den jeweiligen stationsbedingten *date\_time* Werten von *combined-*

*BoundaryValues* – 3600 berechnet. Alle Werte, die größer oder gleich dieses  $y$ -Wertes sind, werden aus dem Datensatz *arable\_data\_hourly* übernommen. Diese Werte sind all jene Werte, die noch nicht überprüft wurden, samt dem jeweiligen vorangegangenen Wert, der bereits überprüft wurde (generiert durch die Subtraktion der *date\_time* des ersten ungeflaggen Wertes mit 3600 innerhalb von  $y$ ). So kann sichergestellt werden, dass die Funktion *qcFlag\_timeConsistency()* mit den richtigen Werten arbeitet und ein korrektes Ergebnis erzielt, obwohl nur eine Teilmenge des gesamten Datensatzes überprüft wird.

Die *else* Abfrage der Schleife muss hinzugefügt werden, weil nicht immer zeitgleich alle gemessenen Werte in die Datenbank überführt werden. In diesem Fall wird für die betroffene Station eine provisorische neue erste Zeile hinzugefügt, die eine Stunde vor dem tatsächlich ersten gemessenen Stundenwert angesetzt ist und für die Lufttemperatur und der relativen Luftfeuchtigkeit ein Fehlwert ausgibt. Nur mit dieser Funktion kann garantiert werden, dass die zeitliche Konsistenz trotz der temporär variablen Natur des Arable-Datensatzes korrekt überprüft wird.

Der anschließende Abschnitt dient der Durchführung der Qualitätsüberprüfung anhand der bereits benannten Funktionen der Qualitätsüberprüfung und wird abgeschlossen mit dem nachfolgenden Befehl, der für jede Station die erste Zeile entfernt:

```
for (i in 1:(length(temporary_qualityControl))) {
  temporary_qualityControl[[i]] <- temporary_qualityControl[[i]][-1,]}
```

Diese erste Zeile ist jene, die innerhalb der *if/else* Abfragen der Funktion *function\_subsetBy DateTime()* dem Datensatz der nicht überprüften Werte hinzugefügt wurde, damit die zeitliche Konsistenz korrekt berechnet werden kann. Da diese ersten Zeilen einer jeder Station entweder bereits im Datensatz *arable\_data\_hourly\_flag* vorhanden sind (*if* Abfrage) oder manuell generiert wurden (*else* Abfrage), müssen sie entfernt werden. Abschließend wird der aufgeteilte Datensatz per *unsplit()* wieder zu einem gemeinsamen überführt und der Datei *arable\_data\_hourly\_flag* innerhalb der Datenbank per *dbAppendTable()* hinzugefügt.

Zusammengefasst lässt sich festhalten, dass die einzelnen Methoden der Datenqualität in einzelne Skripte umgesetzt wurden, die gemeinsam in dem R-Paket *arableQualityControlFlagging* verpackt sind. Um diese Einzelskripte als Funktionen ausführen zu können, stehen zwei Skripte zur Verfügung, wobei *qc\_flagging\_complete\_arable\_dataset.R* die Methoden der Datenüberprüfung basierend auf den Richtlinien der WMO auf den gesamten Datensatz anwendet (siehe: Zahumenský 2002; WMO 2021) und einen neuen namens *arable\_data\_hourly\_flag* erzeugt. *qc\_flagging\_only\_new\_values.R* hingegen wendet diese Methoden lediglich auf die neu erhobenen und nicht überprüften Messwerte an. Durch die Variabilität des Einspeisens der einzelnen Messwerte, führt die

Abfrage der noch nicht überprüften Werte zu einer Zunahme der Komplexität des Skriptes, jedoch ist sie dem anderen vorzuziehen, weil nicht jedes Mal der gesamte Datensatz überschrieben wird, sondern lediglich die noch ungeflaggt Werte überschrieben werden.

### 3.3 Stationsspezifischer Datenvergleich (Arable/DWD)

Den Abschluss des Kapitels Methodik bildet der Datenvergleich ausgewählter Arable-Stationen untereinander und mit ortsnahen Wetterstationen des DWD. Auf diese Weise soll ermittelt werden, wie ähnlich, beziehungsweise unterschiedlich die jeweiligen Messergebnisse ausfallen. Dabei ist zu beachten, dass dieser Vergleich lediglich ein Indiz für die Korrektheit der Daten sein kann. Ein jeder Sensor einer Wetterstation ist mit Ungenauigkeiten behaftet (PRODATA 2022). Vergleicht man mehrere Stationen mit teils unterschiedlichen Sensoren miteinander oder auch identische Stationen mit den gleichen Sensoren, kann es zu Ungenauigkeiten kommen, die eine direkte Validierung erschweren (PRODATA 2022). Zusätzlich werden solche direkten Vergleiche von lokalen mikro- wie mesoklimatischen Bedingungen beeinflusst, die schon auf kleinstem Raum die Vergleichbarkeit problematisch gestalten können (PRODATA 2022). Um ein bestmögliches Ergebnis zu generieren, müssten die Messsensoren ähnlichen Bedingungen ausgesetzt sein, wie jene der Vergleichsstation. Diese Bedingungen können ähnliche Messverfahren und Bedingungen (zum Beispiel, dass die Temperatur mit Bezug auf mögliche Reliefunterschiede in der exakt gleichen Höhe gemessen wird) sein (PRODATA 2022). Da dies jedoch nicht immer gegeben sein kann, wenn Datensätze unterschiedlicher Quellen verglichen werden, muss an dieser Stelle auf diese potenzielle Quelle für festgestellte Abweichungen hingewiesen werden.

Die Methode unterteilt sich in ihrer Durchführung in zwei unterschiedlichen Fragestellungen: Im ersten Schritt soll untersucht werden, wie vergleichbar die Messergebnisse einer ausgewählten Arable-Station mit einer ortsnahen DWD-Station ausfallen. Hierfür werden die Daten der Station mit der *station\_id* 20 in der Nähe von Erdeborn mit den DWD-Daten der Station in Querfurt verglichen. Der zweite Schritt erweitert diesen Ansatz dahin gehend, dass eine ortsnahe DWD-Station mit mehreren nah beieinander aufgestellten Arable-Stationen verglichen wird. Als Referenzstation dient die DWD-Station nahe der Stadt Greifswald in Mecklenburg-Vorpommern. Die zu untersuchenden Arable-Stationen sind jene mit der *station\_id* 60, 61, 62 und 63, welche im circa 17 Kilometer entfernten Bentzin in geringer Distanz zueinander auf einem Acker aufgestellt sind. Dieser Vergleich dient der Erweiterung der Erkenntnisse aus dem vorangegangenen und soll diese vertiefen.

Der Vergleich zwischen der Arable-Station in Erdeborn mit der DWD-Station in Querfurt erfolgt über einen Zeitraum von einem Jahr (14.06.2021 bis einschließlich dem 14.06.2022), wohingegen der Vergleich zwischen den Bentzin Stationen mit der DWD-Station in Greifswald den Zeitraum vom

04.05.2022 bis zum 04.07.2022, somit zwei Monate abdecken wird.

Durchgeführt wird diese Methodik anhand eines separat dafür programmierten Skriptes, welches für die beiden Vergleiche angepasst werden muss, sich im Kern jedoch gleicht. Nachfolgend soll der Aufbau des Grundkonzeptes dieser Skripte beschrieben werden. Die Ergebnisdarstellung sowie die Diskussion dieser erfolgen dann in den nachfolgenden Kapiteln **4. Ergebnisse** und **5. Diskussion**.

### 3.3.1 Aufbau des Skriptes für den datenspezifischen Stationsvergleich

Die erwähnten und anzuwendenden Skripte für die Durchführung des Datenvergleiches mehrerer Stationen sind im Ordner *Vergleichsanalyse\_Arable\_DWD* unter *Abgabedateien* des Abgabeverzeichnisses zu finden und heißen:

- *Arable\_Erdeborn\_vs\_DWD\_Querfurt.r*
- *Arable\_Bentzin\_vs\_DWD\_Greifswald.r*

Nachfolgend wird der generelle Aufbau der Skripte anhand des erstgenannten beschrieben. Jenes für den Vergleich der Bentzin Stationen mit der DWD-Station Greifswald folgt einem identischen, wenn auch komplexeren Schema, da für die Abbildung nicht zwei, sondern fünf Stationen miteinander verglichen werden. Den Anfang des Skriptes macht die Festlegung des Arbeitsverzeichnisses über den Befehl `setwd()`. Dieser legt fest, welches Verzeichnis auf einer Festplatte als Arbeitsverzeichnis dienen soll (Lathiya 2021). Anschließend erfolgt die Installation und das Einbinden der Bibliotheken und R-Pakete, die zur Durchführung dieser Methode benötigt werden (zum Beispiel `ggplot2`, um Abbildungen generieren zu können). Wie auch für die Skripte, die unter 3.2.8 beschrieben werden, wird eine Verbindung mit der Datenbank aufgenommen, um den benötigten Arable-Datensatz herunterzuladen. Dieser wird mit `subset()` auf die wesentlichen Variablen für die Stations-ID, Datumsangaben, Lufttemperatur- und Niederschlagswerte, wie auch Messwerte der relativen Luftfeuchtigkeit zusammengefasst und mit dem bereits bekannten `split()` Befehl temporär in mehrere kleinere Datensätze aufgeteilt.

Im nächsten Schritt werden die benötigten Messwerte vom Server des DWD bezogen. Dies erfolgt über die Befehle `selectDWD()`, `dataDWD()` und `readDWD()` des R-Paketes *rdwd* (BRRY 2022). Heruntergeladen wird sowohl für die Station Querfurt (Stations-ID = 04036) als auch Greifswald (Stations-ID = 01757) die Datensätze *air\_temperature* und *precipitation* in stündlicher Auflösung aus dem Verzeichnis der aktuellen Datensätze (*DWD recent*). Der Datensatz *air\_temperature* beinhaltet neben den Werten Lufttemperatur auch jene der relativen Luftfeuchtigkeit. Diese Datensätze werden per `merge()` anhand der Spalte *MESS\_DATUM* zusammengeführt und in einen neuen Datensatz überführt, der die Spalten *station\_id*, *date\_time*, *t\_air*, *precip* und *rel\_hum* beinhaltet, die mit den korrespondierten Spalten *STATIONS\_ID.x*, *MESS\_DATUM*, *TT\_TU.Lufttemperatur*,

*R1.Niederschlagshoehe* und *RF\_TU.Relative\_Feuchte* des so eben zusammengeführten Datensatzes gefüllt werden. Den Abschluss dieses vorbereitenden Arbeitsschrittes bildet die Auswahl der zu untersuchenden Arable-Stationen anhand des Befehls:

```
arable_data_hourly_erdeborn<- data.frame(arable_data_hourly_temporary[[20]])
```

Die 20 steht hierbei für die Stations-ID der jeweiligen Arable-Station und wird für die Bentzin Stationen durch 60, 61, 62 und 63 ersetzt.

Nach diesem vorbereitenden Arbeitsschritt erfolgt die eigentliche methodische Durchführung. Dazu wird eine Vorauswahl der Daten mit *subset()* generiert, um jeweils einen der zwei angesprochen Zeiträume vom 14.06.2021 bis 14.06.2022 und 04.05.2022 bis 04.07.2022 abzudecken. Dies erfolgt ab den Zeilen 125 beziehungsweise 135 in den jeweiligen Skripten. Aufbauend auf diesen Datensätzen werden nacheinander Boxplots, Liniendiagramme und Histogramme erstellt, die aufgeteilt nach den drei Klimavariablen Lufttemperatur, Niederschlag und relative Luftfeuchtigkeit die jeweiligen Stationen gegenüberstellen.

Die Erzeugung der Boxplots erfolgt über den Befehl *ggplot()+geom\_boxplot()*. Boxplots sind standardisierte Abbildungen, die es ermöglichen, die Verteilung, die Schiefe und die Streuung eines Datensatzes grafisch darzustellen (Zimmermann-Janschitz 2014, S. 327). Dabei werden die fünf statistischen Kennwerte Minimum, das erste Quantil, der Median, das dritte Quantil und das Maximum als Grenzen des Boxplots verwendet (Galarnyk 2018; Zimmermann-Janschitz 2014, S. 326–327). Sowohl das Minimum als auch das Maximum des Boxplots, auch Whisker genannt, sind nicht die kleinsten oder größten Werte eines Datensatzes, sondern jene, die innerhalb eines vordefinierten Abstandes zum ersten und dritten Quantil liegen (Zimmermann-Janschitz 2014, S. 326–327). Dieser Abstand kann zum Beispiel durch  $Q1 - 1.5 * IQR$  beschrieben werden, wobei Q1 für das erste Quantil und IQR für den InterQuantilsabstand stehen (Galarnyk 2018; Zimmermann-Janschitz 2014, S. 326–327). Werte, die außerhalb dieses Bereiches liegen, gelten als Ausreißer (Galarnyk 2018). Ferner erlauben es Boxplots Aussagen über die Verteilung der Werte schließen zu können, zum Beispiel, ob sich diese normalverteilt um den Median herum verteilen oder ob die Verteilung links- oder rechtssteil ist (Williamson et al., 1989, 916 ff.).

Die Erzeugung der Liniendiagramme beginnt für den Vergleich Erdeborn Querfurt ab Zeile 187 und wird über *ggplot()+geom\_line()* generiert. Während die Boxplots und auch die nachfolgenden Histogramme die Stundenwerte als Basis verwenden, wird für die Liniendiagramme Tagesmittelwerte berechnet. Diese Mittelwertberechnung erfolgt vor der Erstellung der Diagramme mithilfe des R-Paketes *dplyr*, das es einem erlaubt, mehrere Arbeitsschritte in einem Befehl zusammenzufassen. So wird der Datensatz *arable\_erdeborn\_subset* ab Zeile 190 des Skriptes zuallererst anhand

der Datumsangabe gruppiert (alle 24 Werte eines Tages werden als Gruppe zusammengefasst) und anschließend der Mittelwert aus diesen berechnet.

```
arable_erdeborn_subset_t_air <- arable_erdeborn_subset %>%  
  group_by(date) %>%  
  dplyr::summarize(t_air_mean = mean(t_air,na.rm=TRUE))
```

Dieser Arbeitsschritt wird auf die Arable, wie auch DWD-Datensätze angewendet und erzeugt für beide jeweils immer drei Datensätze, die Tagesmittelwerte der Lufttemperatur und der relativen Luftfeuchtigkeit, beziehungsweise die Tagessummenwerte der Niederschläge beinhalten. Dieser Arbeitsschritt entfällt für die Histogramme. Diese benötigen lediglich die vorbereiteten Datensätze, die auf die spezifischen Zeiträume zugeschnitten wurden und werden ebenfalls anhand von *ggplot()* erzeugt. Für die Darstellung der Diagramme benötigt *ggplot()* Angaben bezüglich der Eingangsdaten, welche Art von Diagramm erstellt werden soll (für Histogramme wäre dies *geom\_histogram()*), wie der Titel lauten soll *ggtitle()*, Bezeichnungen für die Achsenbeschriftungen, die über *xlab()* und *ylab()* gesteuert werden und weitere Angaben bezüglich der Textgröße und der Form der Legende, die unter *theme()* zusammengetragen werden.

Erzeugt werden diese weiteren Abbildungsarten, weil sie zusätzliche Informationen in grafischer Form liefern, die die inhaltliche Interpretation, die anhand der Boxplots bereits vollzogen werden kann, erweitern. Die Liniendiagramme vergleichen jeweils die Tagesmittelwerte für die Erdeborn- und die Querfurt-Station in Einzelbetrachtung anhand der gegebenen Klimavariablen.

Liniendiagramme werden verwendet, um Trends über die Zeit erkenntlich zu machen (Bahrenberg et al., 2017, S. 57–61). Hierfür werden Datenpunkte (in diesem Fall Tagesmittelwerte) in einem *xy*-Diagramm abgetragen und mit einer Trendlinie verbunden. Die Histogramme hingegen dienen der Darstellung der Verteilung von Datenwerten einer spezifischen Variable (Zimmermann-Janschitz 2014, S. 314–315 & 322). Dabei werden in einem *xy*-Diagramm Balken eingezeichnet, die immer eine vordefinierte Spannbreite von Werten umfassen (beispielsweise alle Werte von 0 bis 5 werden von einem Balken repräsentiert, wohingegen die Werte von 6 bis 10 durch einen weiteren Balken dargestellt werden) (STHDA 2022). Auf die Lufttemperatur bezogen, repräsentiert die *x*-Achse der zu erstellenden Histogramme die Reichweite der gemessenen Lufttemperaturwerte, wohingegen die *y*-Achse für die Häufigkeit des Vorkommens dieser steht (Vergleich: STHDA 2022). Wenn zum Beispiel der Wert 30° Celsius zehnmal im Datensatz vorkommt, wird auf der *y*-Achse eine 10 eingetragen. Histogramme können als Erweiterung der Boxplots gedacht werden, die auf eine andere Art und Weise die Verteilung der Daten darstellen.

Nach dem Abschnitt der Visualisierung der jeweiligen Datensätze und Stationswerte werden in beiden Skripten, die diesen Abbildungen zugrundeliegenden statistischen Maße berechnet. Hierfür werden die beiden Datensätze (zum Beispiel: *arable\_erdeborn\_subset* und *dwd\_querfurt\_subset*), die auf die zu untersuchenden Zeiträume beschränkt wurden, mit *merge()* anhand der Variable *date\_time* zusammengeführt. Anschließend wird ein neuer Datensatz generiert, der von diesem zusammengefassten nur die relevanten Klimavariablen (*t\_air*, *precip*, *rel\_hum*) beinhaltet. Die eigentliche Berechnung des Medians, des Mittelwertes, des ersten und dritten Quantils samt Minimum und Maximum (in Anlehnung an die Boxplots) erfolgt mit dem Befehl *summary()*. Diese Werte werden anschließend als csv-Datei exportiert und in Tabellen übertragen, die im Kapitel 4 Ergebnis vorzufinden sind.

Zusätzlich zur Berechnung der statistischen Maße wie Mittelwert und Median werden der Korrelationskoeffizient und das Bestimmtheitsmaß berechnet. Diese werden jedoch nicht verwendet, um Zusammenhänge zwischen zwei unterschiedlichen Variablen wie Temperatur und relative Luftfeuchtigkeit zu messen, sondern um den Zusammenhang zwischen der Lufttemperatur zum Beispiel der Erdeborn Station mit der Querfurt Station zu generieren. Die hierfür verwendeten Befehle sind *summary(lm(x, y))* und *cor(x, y, use=„complete.obs“)*. *summary(lm())* besteht aus zwei Funktionsteilen, wobei *lm()* verwendet wird, um ein lineares Regressionsmodell zu berechnen, das anhand von *summary()* in tabellarischer Form ausgegeben wird (RDocumentation 2022d). Um den Korrelationskoeffizienten zu ermitteln, muss *cor()* verwendet werden (RDocumentation 2022a). Auch diese Ergebnisse werden tabellarisch als csv-Datei ausgegeben.

Ein solches lineares Regressionsmodell beschreibt, wie gut eine unabhängige Variable *x* eine abhängige Variable *y* beschreiben kann (Freund und Sa 2006, S. 36; Sykes 1993, 3 f.). Im oben beschriebenen Fall wird somit überprüft, wie gut etwa die Temperaturwerte der Erdeborn Arable-Station die korrespondierenden Messergebnisse der DWD-Station in Querfurt erklären können. Dargestellt werden diese linearen Regressionen im nachfolgenden Kapitel als Graphen, die Punktwolken darstellen, die um eine Regressionsgerade herum streuen. Je stärker diese Punktwolke einer Linie gleicht, desto größer ist der lineare Zusammenhang zwischen den beiden getesteten Variablen. Ausgedrückt wird dieser Zusammenhang anhand des Korrelationskoeffizienten *r* (Taylor 1990, S. 36). Dieser kann Werte zwischen **-1** und **+1** annehmen, wobei **0** bedeuten würde, dass kein linearer Zusammenhang zwischen den Variablen besteht und **±1**, dass ein nahezu perfekter linearer Zusammenhang besteht (Taylor 1990, S. 36). Bezogen auf den zu untersuchenden Sachverhalt würde eine **+1** bedeuten, dass hohe Temperaturwerte, die mit den Arable-Stationen gemessen wurden, mit vergleichbaren Werten bei der DWD-Station einhergehen (Taylor 1990, S. 36). Je stärker der Korrelationskoeffizient gegen **0** strebt, desto schwächer wird die erklärende Aussagekraft; desto unterschiedlicher sind die Messergebnisse zwischen den Stationen (Zimmermann-Janschitz 2014,

S. 287).

Als ergänzende Angabe wird im nachfolgenden Kapitel neben dem Korrelationskoeffizient  $r$  das Bestimmtheitsmaß  $r^2$  angegeben. Das Bestimmtheitsmaß beschreibt, wie viel Prozent der Varianz (Streuung) zwischen zwei oder mehreren Variablen durch eine lineare Regression erklärt werden kann (Nakagawa et al., 2017, 1 f.). Werte können zwischen 0 und 100 % (0 bis 1) liegen und je höher der Wert ist, desto besser ist die Güte des Modells (Nakagawa et al., 2017, 1 f.). Erstellt wurden die Regressionsgraphen mit dem generischen Befehl von Base-R `plot()`. `plot()` ermöglicht es auf einfache Art und Weise beispielsweise zwei Variablen in Form von `plot(Variable 1, Variable 2, ...)` gegenüberzustellen (RDocumentation 2022c). Um eine Grafik zu erzeugen, die mehrere Regressionsdiagramme darstellt, kann ein Datensatz generiert werden, der die einzelnen Variablen in sich vereint und anstelle von den zwei Variablen im `plot()` Befehl angegeben wird.

Wenn nicht anders angegeben, dann wurden die vorgestellten Funktionen auf Basis von `help()` und bereits vorhandenem Wissen aus vorangegangenen Veranstaltungen zur Programmierung mit R erstellt.

## 4 Ergebnisse

Das Kapitel Ergebnisse ist wie die Darstellung der Methoden zweigeteilt: zuallererst werden die Ergebnisse der Datenqualitätsüberprüfung besprochen, die auf Basis der Methoden des Unterkapitel 3.2 durchgeführt wurden, anschließend erfolgt die Ergebnisdarstellung des stationsspezifischen Datenvergleiches.

### 4.1 Datenqualitätsüberprüfung und -flagging

Die Datenqualitätsüberprüfung generierte drei Typen von Ergebnissen: die zwei R-Skripte und das R-Paket, die zur Durchführung der Überprüfung und dem Flagging benötigt werden, und der Datensatz `arable_data_hourly_flag` der auf Basis des R-Paketes und der zwei Skripte entweder erzeugt oder erweitert werden kann. Die Beschreibung des funktionalen Aufbaus der Funktionen des R-Paketes, basierend auf den Methoden zur Datenqualitätsüberprüfung, wie auch der zwei R-Skripte, erfolgte bereits im Methodenteil unter 3.2. Auf diesen Ausführungen sei diesbezüglich an dieser Stelle verwiesen. Statt diesen Inhalt zu wiederholen, wird an dieser Stelle die Aufmerksamkeit auf den Datensatz `arable_data_hourly_flag` und seinen Inhalt gelenkt, der als Ergebnis der R-Skripte angesehen werden kann.

Der Datensatz *arable\_data\_hourly\_flag* setzt sich zusammen aus 27 Spalten und zum Zeitpunkt des Verfassens dieser Arbeit aus 123.740 Zeilen, also 123.739 Stundenwerten, die sich über derzeit 48 teils aktive, teils inaktive Stationen erstrecken. Von diesen 123.739 Stundenwerten wurden 118.173 innerhalb der Tabellenspalte *qcFlag\_totalScore\_t\_air* mit einer **1** geflaggt. Das bedeutet, dass 118.173 Lufttemperaturwerte als valide klassifiziert wurden. Das entspricht 95,50 % aller Lufttemperaturstundenwerte. Als fragwürdig (**-777**) wurden 2.083 Lufttemperaturwerte geflaggt. Dies entspricht einem prozentualen Anteil von 1,68 %. Die Anzahl aller Fehlwerte beläuft sich auf 3.483 und entspricht einem prozentualen Anteil aller Lufttemperaturstundenwerte von 2,81 %. Bezüglich des Niederschlages (*qcFlag\_totalScore\_precip*) gilt, dass 96,79 % aller Niederschlagswerte als valide geflaggt wurden, wobei 2,72 % den fragwürdigen (**-777**) und 0,49 % den fehlenden Werten (**-999**) zugeschrieben werden. Die validen Werte der relativen Luftfeuchtigkeit belaufen sich hingegen auf 112.708, was 91,09 % aller Luftfeuchtigkeitswerte entspricht. Als fragwürdig gelten 6,10 % und Fehlwerte belaufen sich auf 3483 (2,81 %). Diese Werte werden für den Überblick in Tabelle 2 zusammengefasst und liefern einen ersten Überblick über die Datenqualität. Die Darstellung der

Tabelle 2: Ergebnisse der Datenqualität (Eigene Darstellung).

	<b>totalScore_t_air</b>	<b>totalScore_precip</b>	<b>totalScore_rel_hum</b>
<b>Anzahl: valide (1)</b>	118.173	119.761	112.708
<b>Anzahl: fragwürdig (-777)</b>	2.083	3.368	7.548
<b>Anzahl: fehlend (-999)</b>	3.483	610	3.483
<b>Anzahl: summiert</b>	123.739	123.739	123.739
<b>%-Anteil: valide (1)</b>	95,50	96,79	91,09
<b>%-Anteil: fragwürdig (-777)</b>	1,68	2,72	6,10
<b>%-Anteil: fehlend (-999)</b>	2,81	0,49	2,81

Ergebnisse der *Totalscores* liefert einen ersten Eindruck über die Qualität der Daten, jedoch geht aus diesen nicht hervor, wie sich die einzelnen Testverfahren auf diese Werte auswirkten. Die nachfolgenden Tabellen 3 (S. 43), 4 (S. 44) und 5 (S. 44) stellen der Reihe nach die Anzahl der validen, fragwürdigen und nicht vorhandenen Werte dar und ermöglichen einen Einblick darüber, welche Tests prozentual weniger Stundenwerte als valide klassifizierten.

Für Tabelle 3 (S. 43) der stündlichen Temperaturwerte gilt: dass für die Tests *Fehlwertüberprüfung* (3.2.1), *Technische Überprüfung: Sensor Range Test* (3.2.2), *Klimatische Überprüfung: Plausible Values* (3.2.3), und der *Inneren Konsistenzprüfung* (3.2.4) bezogen auf den Vergleich zwischen der Lufttemperatur (*t\_air*) und dem Minimum (*t\_min*) und Maximum (*t\_max*) der Lufttemperatur die glei-

che Anzahl an Werten als valide, fragwürdig oder fehlend eingestuft werden. Für diese Tests beläuft sich die Anzahl der validen Stundenwerte auf 120.263 (97,19 %), jene der fragwürdigen Werte auf 0 und der Fehlwerte auf 3.476 (2,81 %). Anders verhält es sich mit den Werten der *Inneren Konsistenzprüfung* (3.2.4) bezogen auf den Vergleich zwischen der Lufttemperatur ( $t_{air}$ ) und dem Taupunkt ( $t_{dew}$ ) sowie für die *Zeitliche Konsistenzprüfung* (3.2.5). Für ersteres gilt, dass 97,06 % (120.107) aller stündlichen Temperaturwerte als valide klassifiziert wurden, wohingegen sich die Anzahl der fragwürdigen auf 0,12 % (149) und jene der Fehlwerte auf 2,81 % (3.483) beläuft. Für die zeitliche Konsistenzprüfung fallen diese Werte wie folgt aus: 95,62 % (118.323) gelten als valide, 1,20 % (1.487) erscheinen fragwürdig zu sein und 3,18 % (3.929) der Werte wurden als Fehlwerte klassifiziert. Für den Niederschlag ergibt sich eine homogene Verteilung der Werte zwischen den

Tabelle 3: Ergebnisse der einzelnen Datenqualitätstests: Lufttemperatur (Eigene Darstellung).

	Fehlwerte	Sensor Range	Plausible Values	*1. Innere Konsistenz	**2. Innere Konsistenz	Zeitliche Konsistenz
<b>Anzahl: valide</b>	120.263	120.263	120.263	120.263	120.107	118.323
<b>Anzahl: fragwürdig)</b>	0	0	0	0	149	1.487
<b>Anzahl: fehlend</b>	3.476	3.476	3.476	3.476	3.483	3.929
<b>Anzahl: summiert</b>	123.739	123.739	123.739	123.739	123.739	123.739
<b>%: valide</b>	97,19	97,19	97,19	97,19	97,06	95,62
<b>%: fragwürdig</b>	0,00	0,00	0,00	0,00	0,12	1,20
<b>%: fehlend</b>	2,81	2,81	2,81	2,81	2,81	3,18

(\*1. *Innere Konsistenz*:  $t_{min} < t_{air} < t_{max}$  | \*\*2. *Innere Konsistenz*:  $t_{dew} < t_{air}$ )

Methoden Technische Überprüfung: *Sensor Range Test* (3.2.2) und *Klimatische Überprüfung: Plausible Values* (3.2.3). Wie der Tabelle 4 (S. 44) entnommen werden kann, beläuft sich die Anzahl der validen Ergebnisse dieser zwei Tests auf 123.129 Stundenwerte, was einem Anteil von 99,51 % entspricht. Fragwürdige Werte sind mit diesen Tests nicht zu verzeichnen, wohingegen sich die Anzahl der Fehlwerte auf 610 (0,49 %) beläuft. Abweichend flaggt die Überprüfung der Fehlwerte 119.761 (96,79 %) aller niederschlagsbedingten Stundenwerte als valide. 3.476 (2,81 %) gelten als fragwürdig und 502 gehen auf Fehlwerte zurück. Zu guter Letzt folgen die Stundenwerte der relativen Luftfeuchtigkeit, die mit den Testverfahren *Fehlwertüberprüfung* (3.2.1), *Technische Überprüfung: Sensor Range Test* (3.2.2), *Klimatische Überprüfung: Plausible Values* (3.2.3), und der *Zeitlichen Konsistenzprüfung* (3.2.5) geflaggt wurden. Für diese ergibt sich, wie der Tabelle 5 (S. 44) ein ähnliches Bild wie schon für die Temperaturwerte. So wurden für die Tests der Fehlwertüberprüfung, der technischen wie klimatischen Überprüfung die gleiche Anzahl an Werten valide

Tabelle 4: Ergebnisse der einzelnen Datenqualitätstests: Niederschlag (Eigene Darstellung).

	Fehlwerte	Sensor Range	Plausible Values
<b>Anzahl: valide</b>	119.761	123.129	123.129
<b>Anzahl: fragwürdig)</b>	3.476	0	0
<b>Anzahl: fehlend</b>	502	610	610
<b>Anzahl: summiert</b>	123.739	123.739	123.739
<b>#: valide</b>	96,79	99,51	99,51
<b>#: fragwürdig</b>	2,81	0,00	0,00
<b>#: fehlend</b>	0,41	0,49	0,49

Tabelle 5: Ergebnisse der einzelnen Datenqualitätstests: Rel. Luftfeuchtigkeit (Eigene Darstellung).

	Fehlwerte	Sensor Range	Plausible Values	Zeitliche Konsistenz
<b>Anzahl: valide</b>	120.256	120.256	120.256	112.708
<b>Anzahl: fragwürdig)</b>	0	0	0	7.099
<b>Anzahl: fehlend</b>	3.483	3.483	3.483	3.932
<b>Anzahl: summiert</b>	123.739	123.739	123.739	123.739
<b>#: valide</b>	97,19	97,19	97,19	91,09
<b>#: fragwürdig</b>	0,00	0,00	0,00	5,74
<b>#: fehlend</b>	2,81	2,81	2,81	3,18

(120.256), fragwürdig (0) und fehlend (3.483) eingestuft, jedoch weicht wie bei den Temperaturwerten die zeitliche Konsistenzüberprüfung von diesen Ergebnissen ab. Für diese gilt, dass 112.708 aller Werte als valide, 7.099 als fehlerhaft und 3.932 als fehlend eingestuft wurden. Die Gründe für diese aufgezeigten Abweichungen zwischen der inneren und zeitlichen Konsistenzüberprüfungen von den anderen Testverfahren sind eines der Themen, die im Zuge des nachfolgenden Kapitels Diskussion thematisiert werden.

## 4.2 Stationsspezifischer Datenvergleich (Arable/DWD)

Aufbauend auf der allgemeinen Datenqualitätsüberprüfung wurden ausgewählte Stationen des Arable-Datensatzes mit Wetterstationen des DWD verglichen. Dabei wurden, wie unter 3.3 beschrieben, zwei voneinander getrennt betrachtete Arable und DWD-Stationspaare miteinander verglichen. Dies sind zum einen die Arable-Station **20** nahe Erdeborns mit der DWD-Station **04036** in Querfurt, und zum anderen die vier Arable-Stationen **60**, **61**, **62** und **63**, welche mit der DWD-Station **01757**

in Greifswald verglichen werden. Die aus diesen Vergleichen generierten Ergebnisse werden nun nacheinander vorgestellt. Dabei wird innerhalb der nachfolgenden Unterkapitel nur ein Teil der erstellten Abbildungen direkt innerhalb des Textes präsentiert werden. Die restlichen finden sich aufgrund der Anzahl der insgesamt erstellten Abbildungen im Anhang dieser Arbeit. Wenn nötig, werden sie mit dem jeweiligen Verweis auf ihre Position im Anhang im Text thematisiert.

#### 4.2.1 Vergleich der Arable-Station Erdeborn mit der DWD-Station Querfurt

Der erste Vergleich erfolgt zwischen der Arable-Station bei Erdeborn und der nächstgelegenen DWD-Station, welche im circa 10 Kilometer entfernten Querfurt gelegen ist. Der betrachtete Zeitraum beginnt am 14.06.2021 und endet mit dem 14.06.2022, wobei stündliche Werte verglichen werden. Paarweise werden in der nachfolgenden Tabelle 6 (S. 45) die ermittelten statistischen Maße der zu untersuchenden Klimavariablen gegenübergestellt. Diese sind das Minimum, das erste Quantil, der Median, der Mittelwert, das dritte Quantil und das Maximum; jeweils aufgeschlüsselt nach den beiden Stationen. Zusätzlich beinhaltet die Tabelle den Korrelationskoeffizienten  $r$  und das Bestimmtheitsmaß  $r^2$ , die zwischen den Klimavariablen der jeweiligen Stationen gebildet werden (etwa die Temperaturwerte von Erdeborn mit den Temperaturwerten von Querfurt).

Tabelle 6: Statistischer Überblick über die Klimavariablen *Lufttemperatur*, *Niederschlag* und *relative Luftfeuchtigkeit* für die Stationen *Arable Erdeborn* und *DWD Querfurt* (Eigene Darstellung).

	Lufttemp. Erdeborn (in °C)	Lufttemp. Querfurt (in °C)	Niederschlag Erdeborn (in mm)	Niederschlag Querfurt (in mm)	Rel. Luft- feuchtigkeit Erdeborn (in %)	Rel. Luft- feuchtigkeit Querfurt (in %)
<b>Minimum</b>	-12,60	-12,10	0,00	0,00	23,00	19,00
<b>1. Quantil</b>	4,40	4,50	0,00	0,00	69,00	65,00
<b>Median</b>	9,80	9,90	0,00	0,00	84,00	81,00
<b>Mittelwert</b>	10,38	10,43	0,06	0,07	78,59	75,96
<b>3. Quantil</b>	16,10	16,00	0,00	0,00	91,00	89,00
<b>Maximum</b>	33,10	33,10	28,50	39,10	100,00	100,00
<b>r</b>	0,9850654		0,5464303		0,9223089	
<b>r<sup>2</sup></b>	0,9704		0,2986		0,8507	

Die statistischen Maße bezogen auf die Lufttemperatur fallen vergleichbar aus. So beträgt der Median für Erdeborn 9,80 °C, während er für Querfurt bei 9,90 °C liegt. Die Mittelwerte liegen ähnlich nah beieinander: 10,38 °C für Erdeborn und 10,43 °C für Querfurt. Der kleinste gemessene Wert liegt für Erdeborn bei -12,6 °C und für Querfurt bei -12,1°. Die Werte des 1. Quantils belaufen sich auf 4,40°

(Erdeborn) und 4,50 °C (Querfurt), wohingegen das 3. Quantil jeweils bei 16,1 °C beziehungsweise 16 °C angesetzt werden. Die Höchstwerte belaufen sich für beide Stationen auf 33,1 °C. Der Korrelationskoeffizient zwischen den Temperaturwerten der Erdeborn- und der Querfurt-Station beträgt 0,9851 wohingegen das Bestimmtheitsmaß 0,9704 beträgt, was auf einen starken erklärenden Zusammenhang der Variable Lufttemperatur zwischen den einzelnen Stationen hindeutet.

Für den Vergleich der Niederschlagswerte ergeben sich gleiche Werte für die Minima, das 1. Quantil, den Median und dem 3. Quantil von 0,00 mm. Der Mittelwert wird für Erdeborn mit 0,06 mm und für Querfurt mit 0,07 mm beziffert. Die Maxima hingegen liegen für Erdeborn bei 28,50 mm und für Querfurt bei 39,10 mm. Der Korrelationskoeffizient beträgt hingegen 0,5464 und das Bestimmtheitsmaß 0,2986; die Werte der Erdeborn Station erklären nur zu geringen Anteilen jene der DWD-Station. Die Ursache für diese Werte wird anhand der später erfolgenden Betrachtung der Abbildungen 5 (S. 49), 7 (S. 50) und 6 (S. 50) klarer und wird Gegenstand der Diskussion sein.

Die Werte der relativen Luftfeuchtigkeit sind in Prozent angegeben und reichen von 0 bis 100 %. Die Minimum-Werte liegen bei 23 % (Erdeborn) und 19 % (Querfurt) und weichen um 4 Prozentpunkte voneinander ab. Dieselbe Verschiebung weisen auch die Werte des 1. Quantils auf. So erreicht Erdeborn 69 % und Querfurt 65 %. Median und Mittelwert liegen für Erdeborn hingegen bei 84 % und 78,59 %, wohingegen Querfurt 81 % und 75,96 % zu verzeichnen hat. Das 3. Quantil wird jeweils mit 91 % (Erdeborn) oder 89 % (Querfurt) beziffert, wohingegen das Maxima für beide Stationen bei 100 % liegt. Der Korrelationskoeffizient erzielt zwischen den Werten Erdeborns und Querfurts 0,92, wohingegen das Bestimmtheitsmaß auf circa 0,85 kommt. Es besteht somit zwischen den Stationen für diese KlimaVariable ein ausgeprägter, linearer Zusammenhang, der zu 85 % die Varianz dieser Werte voneinander beschreiben kann.

Zur besseren Veranschaulichung dieser statistischen Maße und um einen vereinfachteren Zugang zu der Vergleichsanalyse zu ermöglichen, wurde auf die Darstellungsmethoden der Boxplots, Liniendiagramme, Histogramme und linearer Regressionsmodelle zurückgegriffen. Die nachfolgende Abbildung 2 (S. 47) stellt die Temperaturwerte der Stationen Erdeborn und Querfurt als Boxplots in Orange und Grün dar. Die untere Linie der Boxen repräsentiert die in Tabelle 6 (S. 45) angegebenen Werte für das 1. Quantil, wohingegen die obere Linie der Box dem 3. Quantil entspricht. Die schwarz eingezeichnete Linie innerhalb der Box repräsentiert jeweils den Median. Vergleicht man diese beiden Boxen miteinander, wirken diese identisch zueinander. Große Unterschiede zwischen den Temperaturmessungen der Arable- und der DWD-Station scheinen nicht vorhanden zu sein. Abbildungen 3 (S. 47), 18 (S. 93) und 19 (S. 93) liefern vergleichbare Ergebnisse. So weist der Kurvenverlauf der durchschnittlichen, täglichen Temperaturwerte der Arable-Station (dargestellt in Grün) einen ähnlichen Verlauf auf wie jene der DWD-Station (dargestellt in Orange). Abweichungen sind zwar vorhanden, abgesehen von diesen sind die Kurvenverläufe nahezu deckungsgleich.

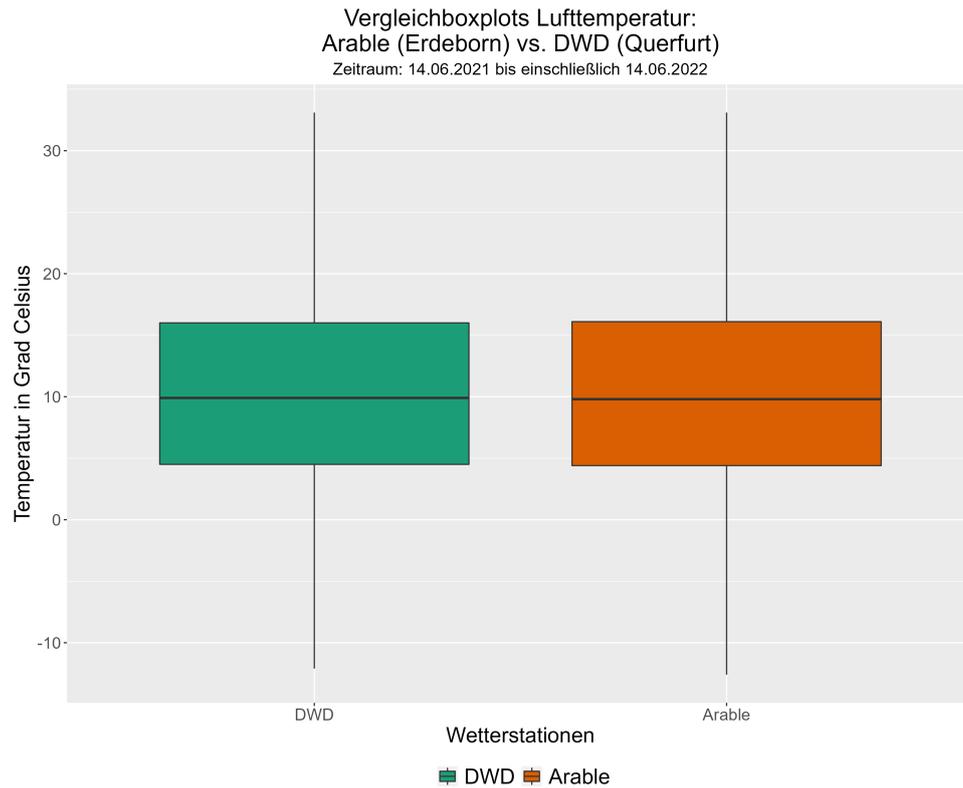


Abbildung 2: Vergleich der Boxplots der Stationen *Arable Erdeborn* und *DWD Querfurt* für die Lufttemperaturwerte (Eigene Darstellung).

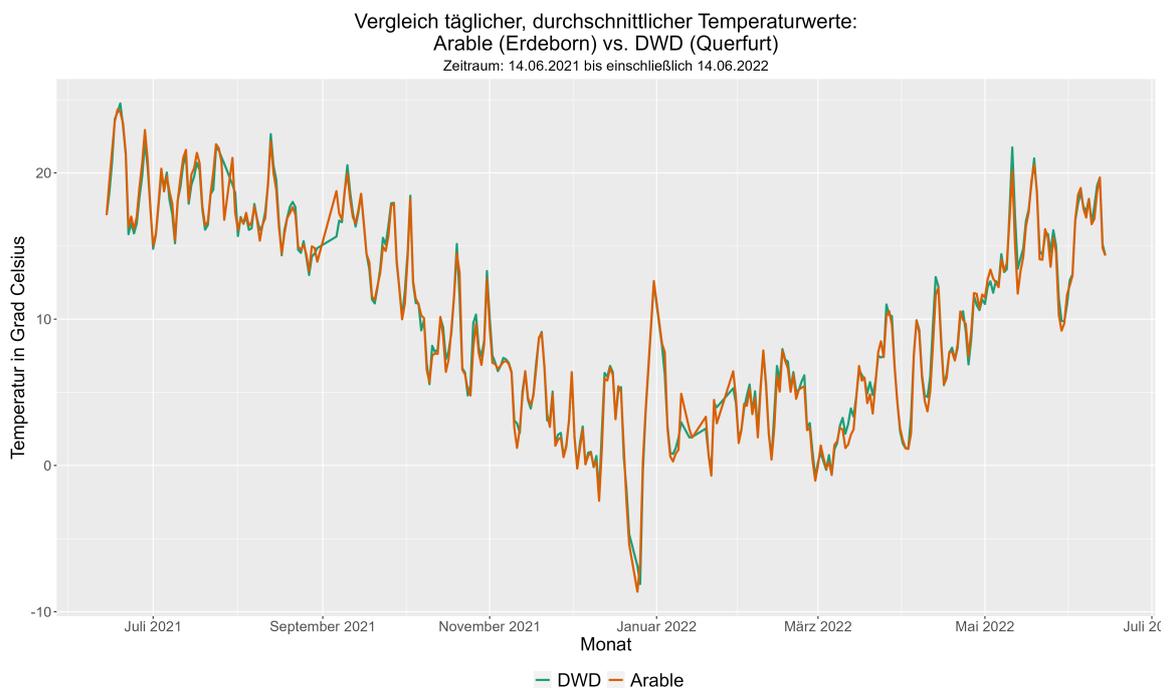


Abbildung 3: Vergleich der täglichen Durchschnittstemperaturwerte der Stationen *Arable Erdeborn* und *DWD Querfurt* anhand eines Liniendiagramms (Eigene Darstellung).

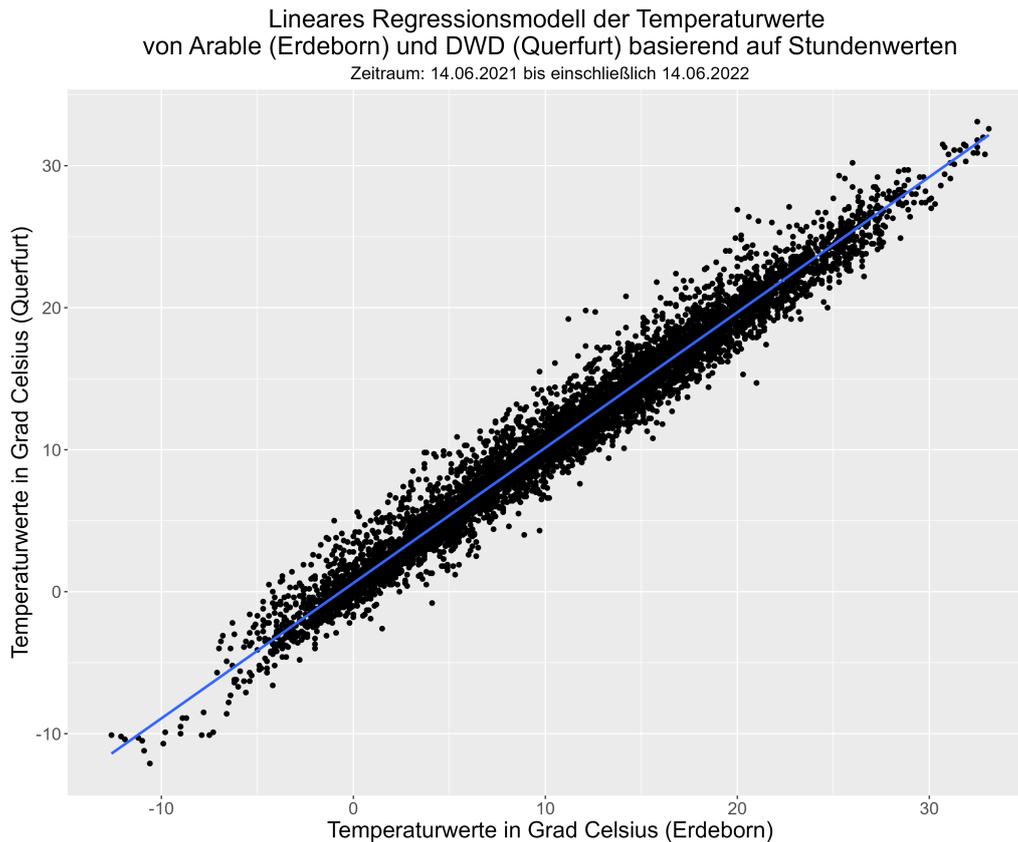


Abbildung 4: Einfaches lineares Regressionsmodell der Temperaturwerte der Stationen *Arable Erdeborn* und *DWD Querfurt* (Eigene Darstellung).

Abbildung 4 (S. 48) vertieft diesen Eindruck, in dem diese das berechnete, lineare Regressionsmodell der Temperaturwerte beider Stationen grafisch darstellt und die Brücke zu den in Tabelle 6 (S. 45) bezifferten Regressionskoeffizienten und dem Bestimmtheitsmaß schlägt. Auf der x-Achse werden die Temperaturwerte der Arable-Station und auf der y-Achse jene der DWD-Station abgetragen. Diese schwarzen Punkte streuen dicht um die blaue Regressionsgerade herum und veranschaulichen, dass ein gemessener Temperaturwert  $x$  innerhalb des Datensatzes der Erdeborn-Station zu einem bestimmten Zeitpunkt  $y$  mit einem vergleichsweise ähnlich stark oder schwachen Temperaturwert in dem Querfurt-Datensatz einhergehen. Zwischen den Werten dieser beiden Stationen scheint ein starker linearer Zusammenhang zu bestehen. Für den Niederschlag ergibt sich ein abweichendes Bild von der Verteilung der Werte. Abbildung 5 (S. 49) stellt die Boxplots der Niederschlagswerte beider Stationen dar. Da bis auf die Werte der Maxima sämtliche statistischen Maße beider Stationen gleich oder sehr nah an 0 mm liegen, können keine Boxen eingezeichnet. Dargestellt werden anhand schwarzer Punkte die Ausreißer, die nicht von der Spannweite der Whisker abgedeckt werden. Optisch sammelt sich die Mehrzahl dieser zwischen 0 bis 10 mm und weisen Ähnlichkeiten zueinander auf. Die Ausreißer oberhalb der 10 mm Marke weichen jedoch stärker voneinander ab.

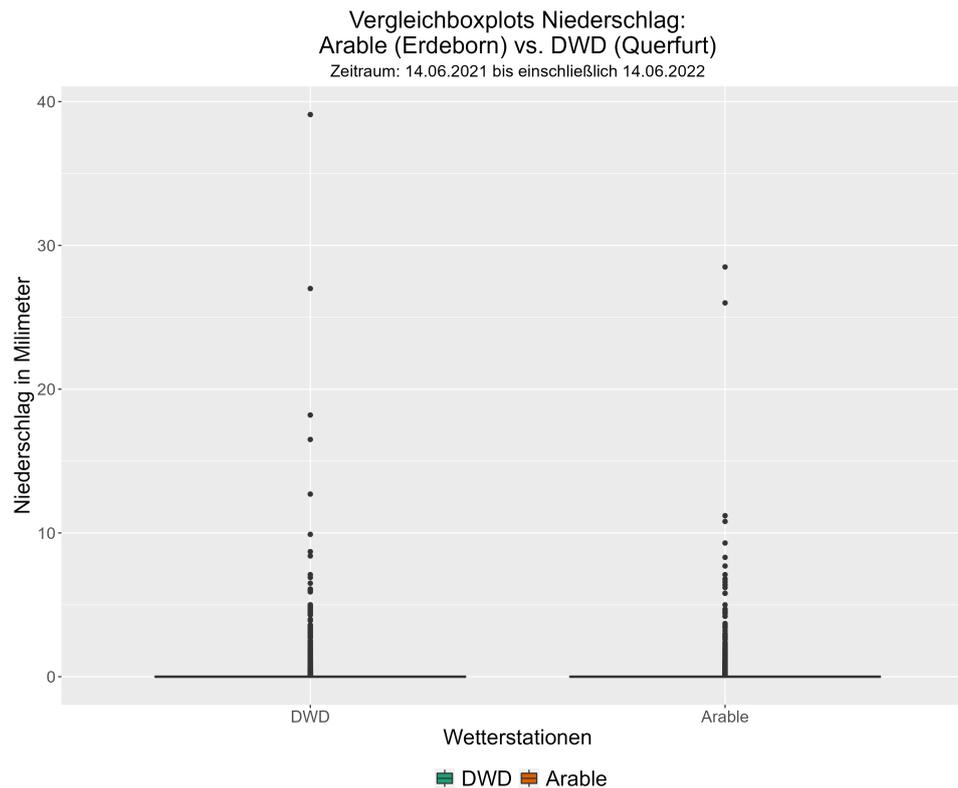


Abbildung 5: Vergleich der Boxplots der Stationen *Arable Erdeborn* und *DWD Querfurt* für die Niederschlagswerte (Eigene Darstellung).

Von den vier Ausreißern der Erdeborner Station liegen zwei nahe der 10 mm Marke, wohingegen die anderen beiden nahe bei 26 und 28,5 mm liegen. Querfurt hat fünf Ausreißer über der 10 mm Marke zu verzeichnen, wovon der erste bei circa 12 mm, die nächsten beiden bei circa 16 und 18 mm liegen. Die verbleibenden zwei Ausreißereignisse liegen bei circa 27 und 39,1 mm. Dieser Eindruck der starken Ansammlung (Clustering) der gemessenen Niederschlagswerte im unteren Segment, gekoppelt mit teils stark ausgeprägten Ausreißern, spiegelt sich auch in der Betrachtung des nachfolgenden linearen Regressionsdiagrammes (Abbildung 6 (S. 50)) wider:

Der Großteil der eingetragenen Messpunkte ist zwischen 0 und 5 mm auf der x-Achse und 0 und 5 mm auf der y-Achse vorzufinden. Sie clustern sich um den Punkt 0|0 mm herum. Verglichen mit der Abbildung 4 (S. 48) für die Regression der Lufttemperaturwerte (wo hohe/niedrige Temperaturwerte von Erdeborn mit hohen/niedrigen Werten für Querfurt einhergehen), sind für die Niederschlagsmessungen Werte zu verzeichnen, die von der Arabl Station gemessen werden, jedoch nicht im selben Ausmaß von der DWD-Station in Querfurt und umgekehrt (siehe Punkte, die entlang der Achsen aufgereiht sind). Dasselbe gilt nicht nur für die niedrigen Niederschlagsmessungen, sondern auch für die verzeichneten Ausreißer.

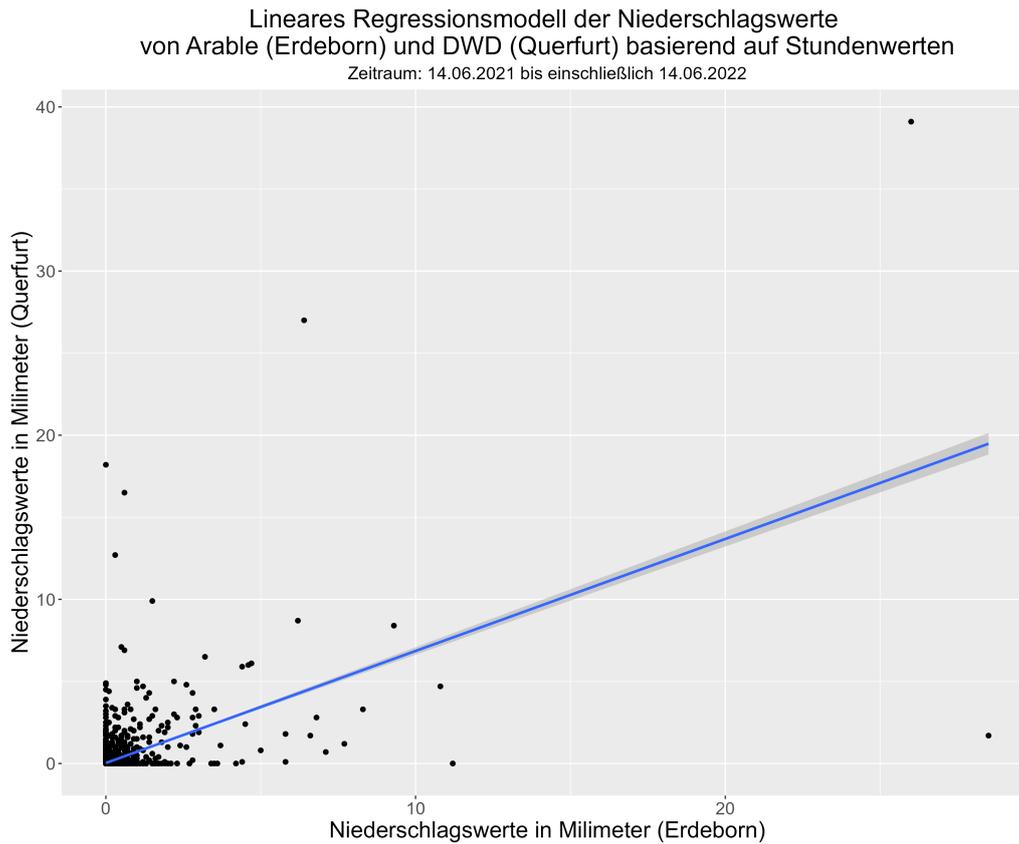


Abbildung 6: Einfaches lineares Regressionsmodell der Niederschlagswerte der Stationen *Arable Erdeborn* und *DWD Querfurt* (Eigene Darstellung).

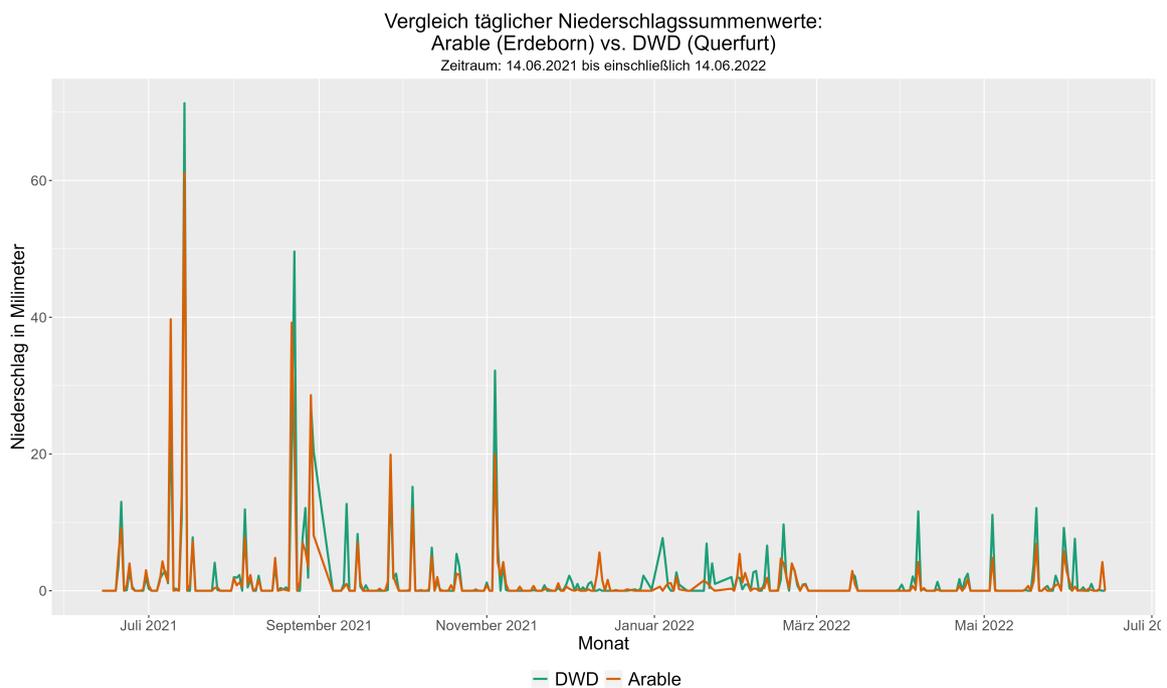


Abbildung 7: Vergleich der täglichen Summenwerte des Niederschlages der Stationen *Arable Erdeborn* und *DWD Querfurt* anhand eines Liniendiagramms (Eigene Darstellung).

Betrachtet man an dieser Stelle die täglichen Niederschlagssummenwerte, wie sie in Abbildung 7 (S. 50) dargestellt sind, fällt auf, dass stärker ausgeprägte Niederschlagsereignisse von beiden Stationen am gleichen Tag gemessen wurden, jedoch in ihrer Stärke variieren. So weist das DWD Signal höhere Werte auf, bei gleichen Niederschlagsereignissen.

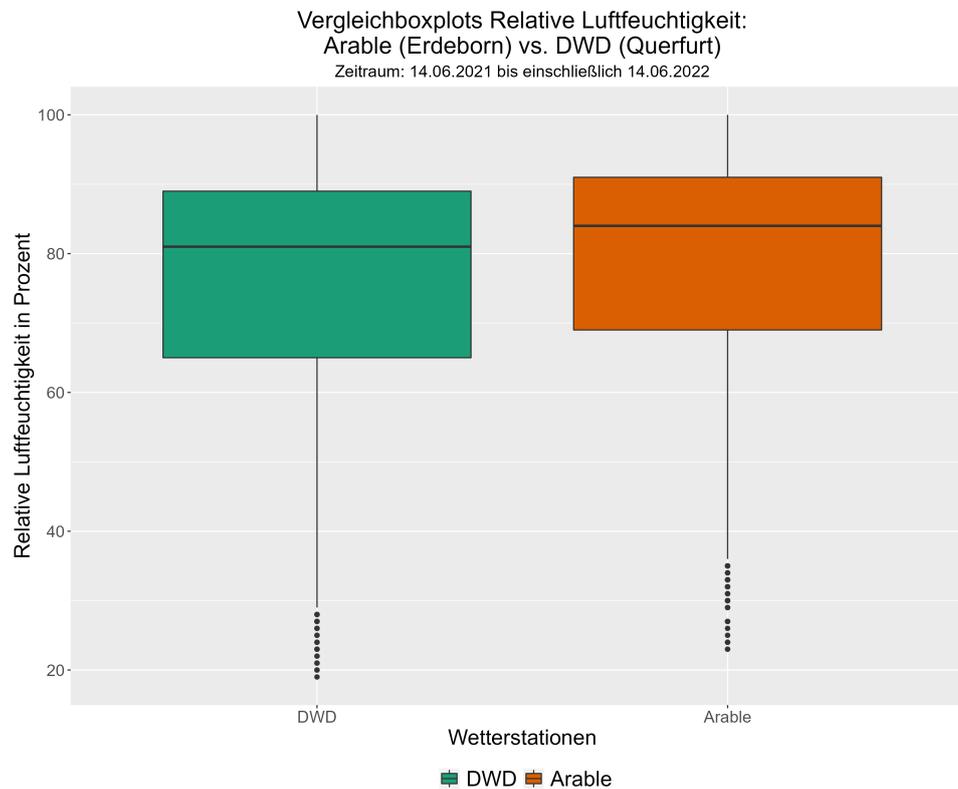


Abbildung 8: Vergleich der Boxplots der Stationen *Arable Erdeborn* und *DWD Querfurt* für die relativen Luftfeuchtigkeitswerte (Eigene Darstellung).

Ausgehend von den in Tabelle 6 (S. 45) eingetragenen Werten, die sich zwischen Erdeborn und Querfurt ähneln, kann der Eindruck entstehen, dass die Stationen vergleichbare Werte für die relative Luftfeuchtigkeit liefern. Abbildung 8 (S. 51) stellt, wie auch schon zuvor für die Temperatur- und Niederschlagswerte, die Boxplots dar. Wie schon anhand der Tabelle aufgezeigt wurde, verzeichnet die Arable-Station insgesamt höhere Werte bezüglich der statistischen Maße bezogen auf die Quantile und dem Median. Insgesamt ist die Spannweite des Arable Boxplots geringer als die des DWD (3. Quantil = 91 %, 1. Quantil = 69, Spannweite zwischen diesen 22 verglichen mit einer Spannweite von 24 für Querfurt).

Die Betrachtung der täglichen Durchschnittswerte anhand von Abbildung 9 (S. 52) zeigt, dass diese für die relative Luftfeuchtigkeit die Kurven der Arable und DWD-Station im Zeitraum von Juni 2021 bis circa Januar 2022 Ähnlichkeiten aufweisen. Bei genauerer Betrachtung fällt jedoch die bereits beobachtete Tatsache auf, dass die DWD Luftfeuchtigkeitswerte niedriger ausfallen als die Erdeborn

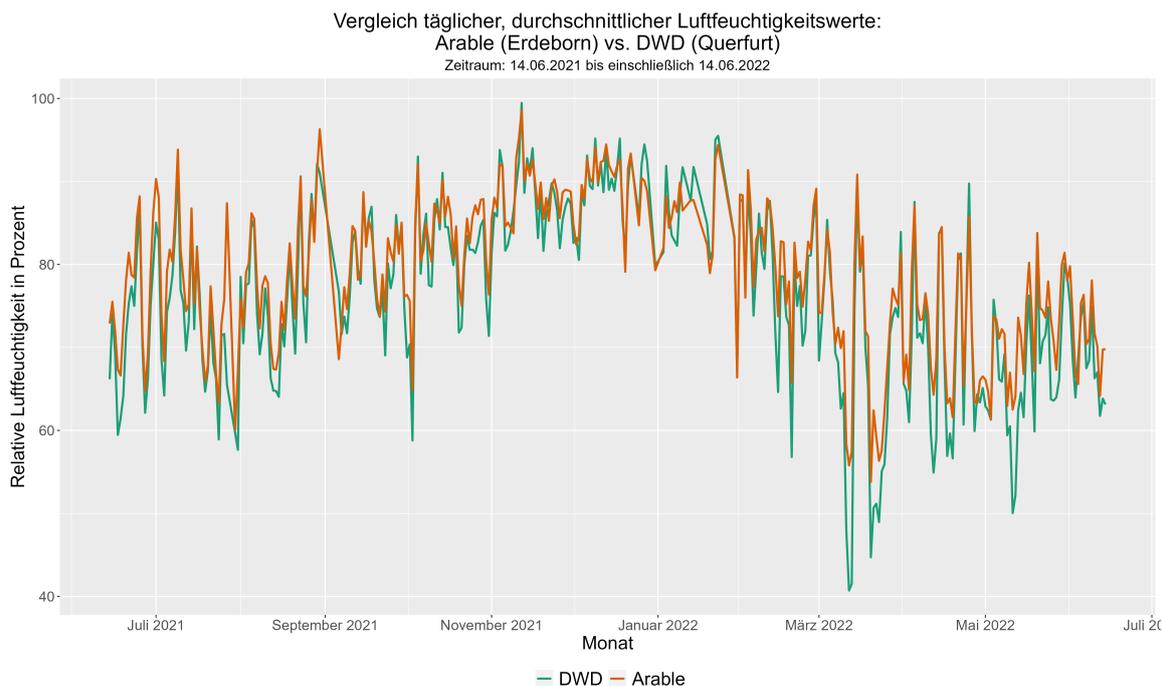


Abbildung 9: Vergleich der täglichen Durchschnittswerte der relativen Luftfeuchtigkeit der Stationen *Arable Erdeborn* und *DWD Querfurt* anhand eines Liniendiagramms (Eigene Darstellung).

Werte. Mitte Juni und Oktober 2021 sind dafür gute Beispiele. Ab Mitte Februar nehmen die Abweichungen zu. Im Zeitraum März bis Mai 2022 fallen die niedrigen Luftfeuchtheitsdurchschnittswerte geringer aus als für die Arable-Station. Ergänzend dazu sei an dieser Stelle auf die Abbildung 17 im Anhang auf Seite 91 verwiesen. Diese stellt für den Zeitraum vom 14. Juli 2021 bis 21. Juli 2021 die Tageswerte der gemessenen relativen Luftfeuchtheitswerte ausschnitthaft dar. Diese Nahaufnahme zeigt, dass die gemessenen Werte einander zwar ähneln, sie jedoch in ihren minimalen und maximalen Ausprägungen nicht identisch sind. So werden zum Beispiel in den Nächten vom 14. und 15. Juli 2021 die relative Luftfeuchtigkeit für die Arable-Station als geringfügig höher eingeschätzt, als für die Querfurter Station. Zudem scheint ein zeitlicher Versatz der Messungen zwischen den Stationen vorzuliegen.

Mit einem Regressionskoeffizienten von 0,9223089 und einem Bestimmtheitsmaß von 0,8507 deutet vieles darauf hin, dass es zwischen der Verteilung der Erdeborn Luftfeuchtheitswerte und den Querfurtern einen positiven Zusammenhang gibt. Abbildung 10 (S. 53) zeigt, dass die Luftfeuchtheitswerte sich über den gesamten Messwertebereich annähernd homogen verteilen. Zwischen 20 und 40 % sind die wenigsten Werte zu verzeichnen. Darüber hinaus erscheint die Werteverteilung über den Messbereich von 40 bis 100 % gleichmäßig entlang der Regressionsgerade zu erfolgen. Generell lässt sich anhand dieser Grafik festhalten, dass hohe Luftfeuchtheitswerte in der einen Wetterstation mit hohen in dem Datensatz der anderen einhergehen. Wie bereits aufgezeigt wurde, gibt es jedoch auch Wertepaare, wo etwa die Erdeborn-Station einen höheren Wert gemessen hat

als die Querfurter und andersherum.

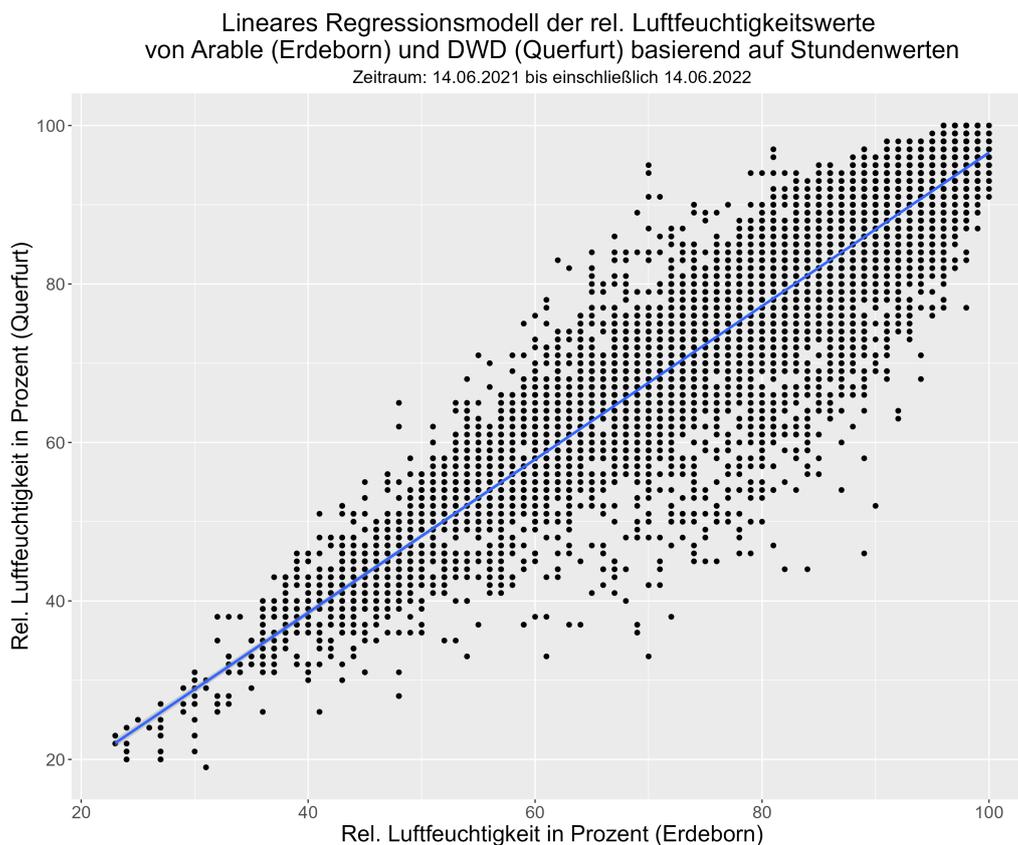


Abbildung 10: Einfaches lineares Regressionsmodell der relativen Luftfeuchtwerte der Stationen *Arable Erdeborn* und *DWD Querfurt* (Eigene Darstellung).

Neben diesen Abbildungen wurde für jede Station und für jede KlimaVariable jeweils ein Histogramm und Nahaufnahmen eines kürzeren Betrachtungszeitraumes anhand der Stundenwerte für die Variablen Temperatur und Niederschlag erstellt, wie sie so eben bei der Ergebnisdarstellung der relativen Luftfeuchtwerte erwähnt wurde. Diese Abbildungen 18 (S. 93), 19 (S. 93), 20 (S. 94), 21 (S. 94), 22 (S. 95), 23 (S. 95), 15 (S. 90), 16 (S. 91) und 17 (S. 91) finden sich im Anhang und unterstreichen die dargestellten Ergebnisse weiter. So weisen die Temperaturkurvenverläufe der beiden Stationen in Abbildung 15 (S. 90) starke Ähnlichkeiten zueinander auf. Auch die Histogramme (Abbildungen 18 (S. 93) und 19 (S. 93)) gleichen einander, wohingegen jene für die relative Luftfeuchtigkeit wie bereits dargestellt marginal stärker voneinander abweichen, als es für die Temperaturwerte der Fall ist (Abbildungen 22 (S. 95) und 23 (S. 95)). Die Histogramme für die Niederschlagswerte (Abbildungen 20 (S. 94) und 21 (S. 94)) sind ebenfalls stark identisch und heben die extreme Anzahl von Stundenwerte hervor, an denen kein Niederschlag gemessen wurde (in beiden Fällen bis an die 8.000 niederschlagslose Ereignisse). Die starken Niederschlagsereignisse über 10 mm treten in beiden Datensätzen derart selten auf, dass sie innerhalb der Histogramme nicht erkenntlich sind.

#### 4.2.2 Vergleich der Arable-Stationen Bentzin mit der DWD-Station Greifswald

Der Vergleich zwischen der Arable-Station **20** in Erdeborn und der DWD-Station in Querfurt hat sich auf eine einseitige Betrachtung mit jeweils nur einer Station beschränkt. Um die Erkenntnisse dieses Vergleiches zu erweitern, wurden kurzfristig die Messwerte der Stationen **60**, **61**, **62** und **63** des Arable-Datensatzes herangezogen und untereinander und mit der DWD-Station in Greifswald verglichen. Da diese Arable-Stationen seit Anfang Mai 2022 Wetterdaten sammeln, beschränkt sich der nachfolgend dargestellte Zeitraum der Daten auf den 04.05.2022 bis zum 04.07.2022. Diese kurzfristige Betrachtung über zwei Monate hebt jedoch wichtige Besonderheiten der Arable Messdaten hervor, die im Umgang mit diesen zu beachten sind und die Grenzen der automatischen Datenqualitätsüberprüfung aufzeigen.

##### 4.2.2.1 Vergleich der Boxplots anhand Stundenwerte

Die nachfolgende Abbildung 11 (S. 55) stellt eine Zusammenstellung der anhand der Klimavariablen *Lufttemperatur*, *Niederschlag* und *Relative Luftfeuchtigkeit* erstellten Boxplots für die genannten fünf Stationen dar. Die Farbcodierung ist dabei einheitlich gehalten, sodass die DWD-Station Greifswald in dunkelgrün, Arable-Station **60** in Orange, **61** in Lila, **62** in Pink und **63** in Hellgrün dargestellt werden. Beginnend mit den Boxplots basierend auf den Temperaturwerten fällt auf, dass die Arable Boxplots von circa 15 bis 21 °C reichen, wohingegen die DWD-Station bei 15 °C beginnt und bei circa 18 °C endet. Diese Unterschiedlichkeit ist auch bei der relativen Feuchtigkeit zu beobachten. Hier liegen die Werte des 1. und 3. Quantils der DWD-Station zwischen circa 60 und 85 %, wohingegen die Arable-Stationen Werte von 50 % bis 85 % oder im Fall der Station **62** gar 90 % erreichen. Die Niederschlagswerte hingegen weisen in ihrer Darstellung Ähnlichkeiten zu den Boxplots des Erdeborn-Querfurt Vergleiches auf. Die eigentlichen Boxen sind nicht erkennbar und die als Punkte dargestellten Ausreißer liegen zwischen 0 und circa 8 mm. Der Großteil der Niederschlagsereignisse ist im unteren Segment bis 2,5 mm vorzufinden. Die Ereignisse über dieser Marke fallen alle in ihren Messwerten unterschiedlich aus. Vergleichbare Ergebnisse liefern die Stationen Arable **60** und **62**. Verschiebungen im Signal der Messstärke treten verglichen zu diesen bei Arable **61** und **63** auf. Die Höchstwerte hat die Station **61** zu verzeichnen.

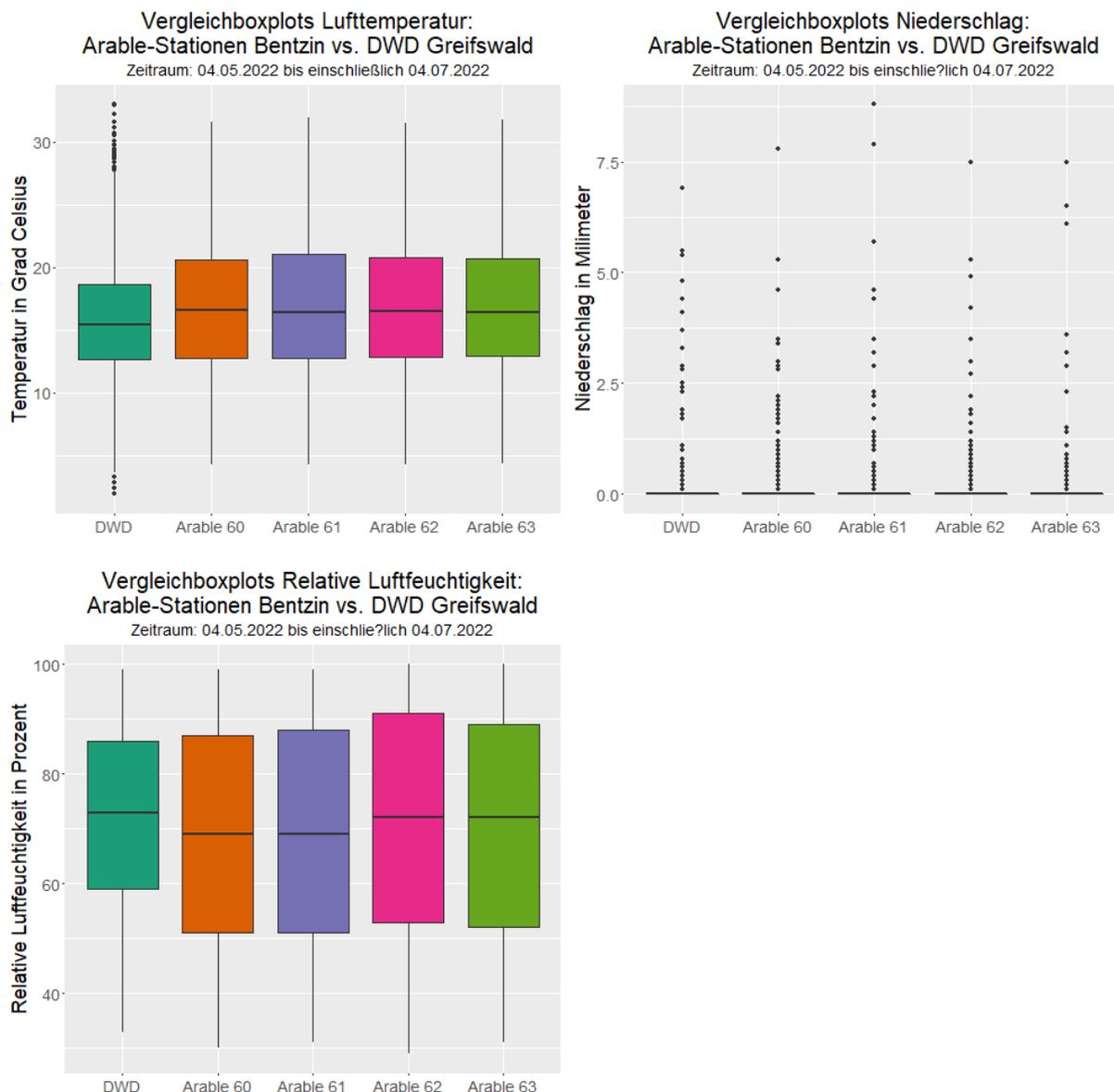


Abbildung 11: Gemeinsame Darstellung der Boxplots der Stationen *Arable Bentzin* und *DWD Greifswald* der Klimavariablen *Lufttemperatur*, *Niederschlag* und *Relative Luftfeuchtigkeit* (Eigene Darstellung).

#### 4.2.2.2 Vergleich der Liniendiagramme anhand täglicher Werte

Die Betrachtung der nachfolgenden drei Liniendiagramme der Abbildungen 12 (S. 56), 13 (S. 57) und 14 (S. 57) zeigen auf, wo die innerhalb der Boxplots beobachtete Abweichungen auftreten. Abbildung X stellt die durchschnittlichen Tageswerte der Lufttemperatur aller fünf Stationen als Linienverläufe dar. Die Farbcodierung ist hierbei identisch mit jener der Boxplots. Um die DWD-Station besser abzuheben, wird diese als schraffierte Linie dargestellt. Auf der x-Achse werden die Tage, an denen die Messungen stattfanden, abgetragen und auf der y-Achse die gemessenen, täglichen Durchschnittstemperaturwerte. Die Betrachtung der Linienverläufe zeigt auf, dass es zwischen dem 04.05.2022 bis zum 10.05.2022 zu auffälligen Diskrepanzen zwischen den Werten der

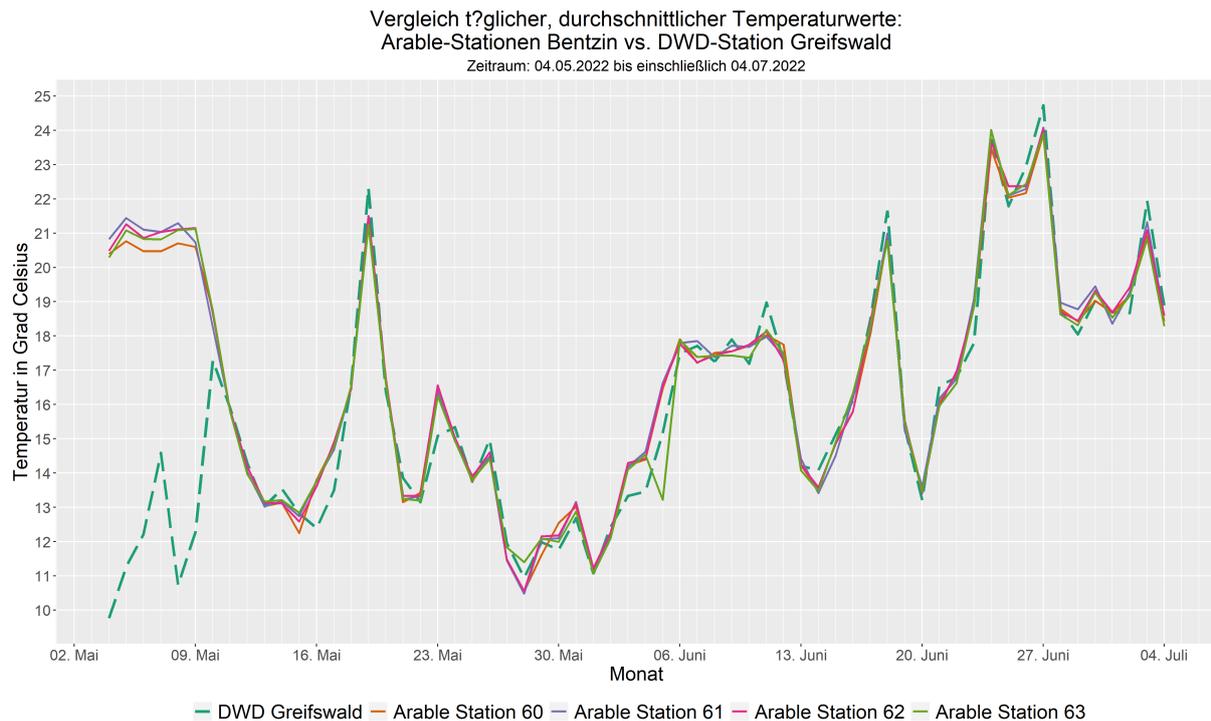


Abbildung 12: Vergleich der t?glichen Durchschnittstemperaturwerte der Stationen *Arable Bentzin* und *DWD Greifswald* anhand eines Liniendiagramms (Eigene Darstellung).

Arable-Stationen und der DWD-Station kommt. Ab dem 10.05.2022 gleichen sich die f?nf Kurven einander an und verlaufen, abgesehen von leichten Abweichungen zwischen Greifswald und den Bentzin Stationen in den Zeitr?umen 16.05. bis 18.05.2022 und 02.06 bis zum 06.06.2022, nahezu deckungsgleich.

Ein anderes Bild liefern die Niederschlagswerte, die in Abbildung 13 (S. 57) dargestellt sind. Wie schon anhand der Boxplots erkennbar war, verlaufen diese Kurven der aufsummierten Tageswerte keinesfalls identisch zueinander. Erkennbar ist, dass wenn Niederschlagsereignisse auftraten, diese von allen Stationen aufgenommen wurden. Daf?ur ist die Betrachtung der Arable-Stationen untereinander umso relevanter, da diese ortsnah voneinander aufgestellt wurden und teils voneinander abweichen. Zwei solcher auff?lligen Ereignisse sind jene am 28. Mai und 2. Juni 2022. Im ersten Fall schlug besonders die Arable-Station **60** stark aus und erreicht Werte von bis zu circa 15,5 mm. Die restlichen Arable-Stationen erreichen um die 8 mm. Das Ereignis am 2. Juni hat zwei Stationen mit starken Signalen von bis zu 17 mm zu verzeichnen (Station **61** und **63**), wohingegen die Stationen **60** und **62** 5 mm erreichen. Die potenziellen Messfehler zwischen dem 4. und 10. Mai sind in dieser Abbildung nicht erkennbar.

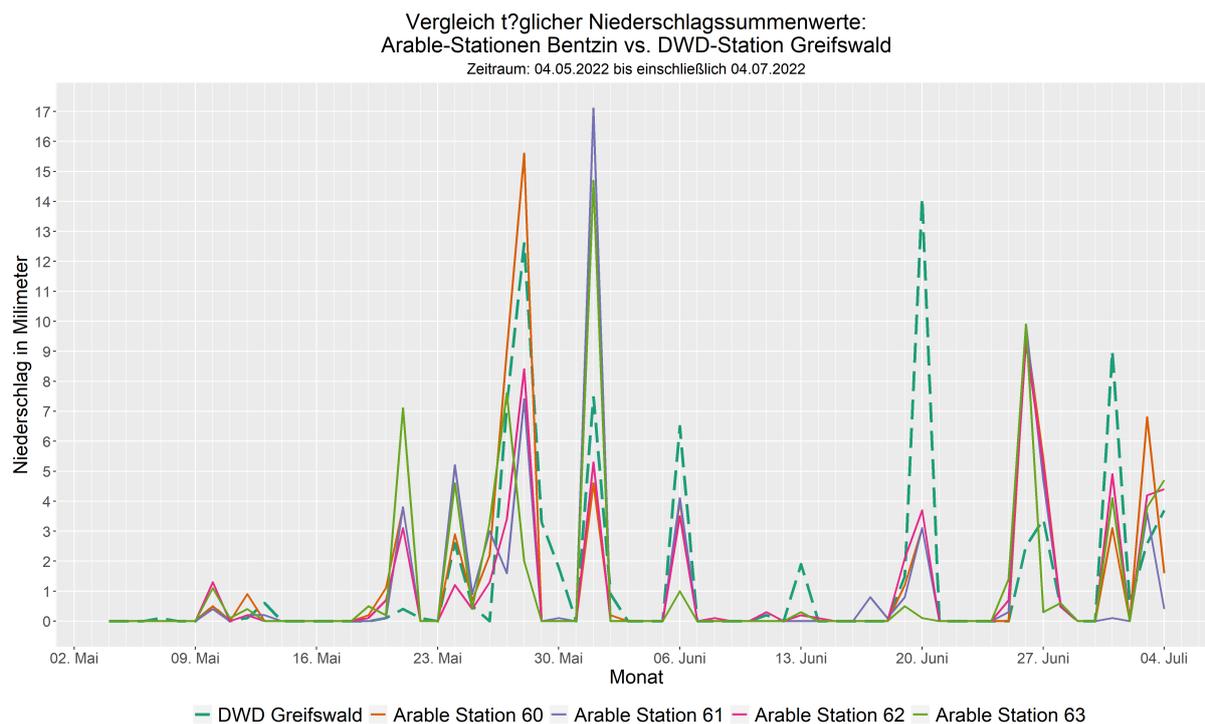


Abbildung 13: Vergleich der t?glichen Niederschlagssummenwerte der Stationen *Arable Bentzin* und *DWD Greifswald* anhand eines Liniendiagramms (Eigene Darstellung).

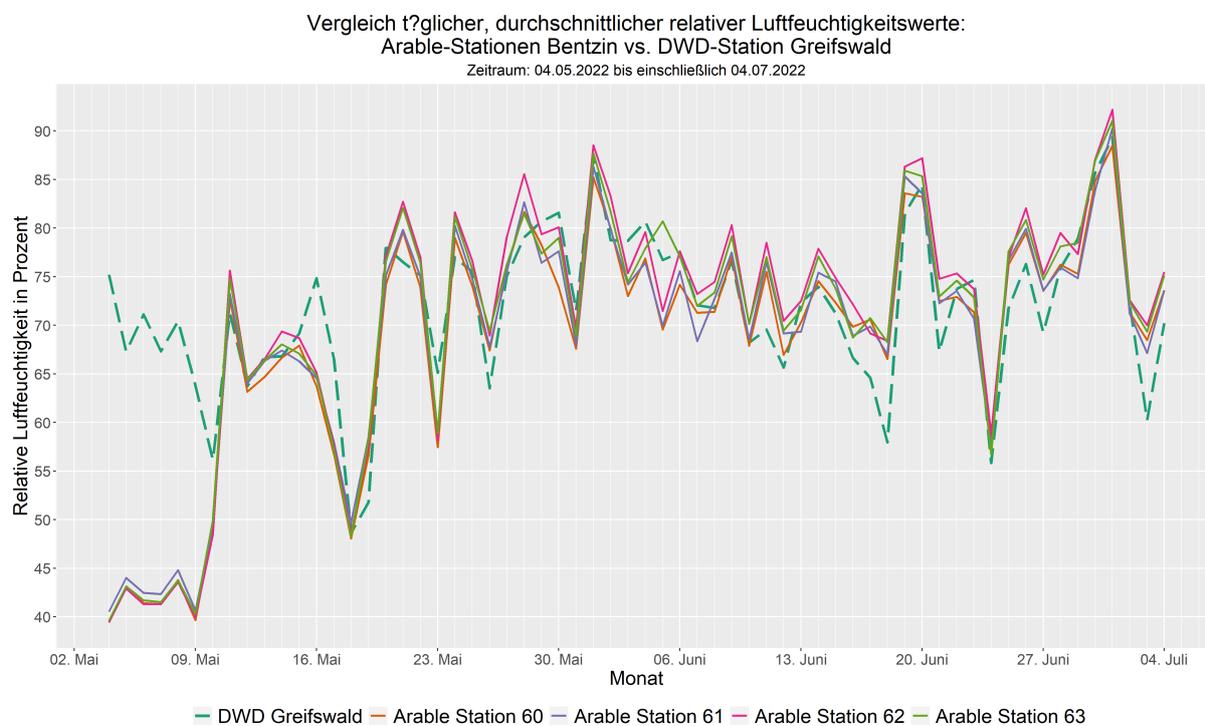


Abbildung 14: Vergleich der t?glichen Durchschnittswerte der relativen Luftfeuchtigkeit f?r die Stationen *Arable Bentzin* und *DWD Greifswald* anhand eines Liniendiagramms (Eigene Darstellung).

#### 4.2.2.3 Vergleich der Histogramme basierend auf Stundenwerte

Neben den Boxplots und Liniendiagrammen wurden auch Histogramme erstellt, die im Anhang ab Seite 96 finden. Der Vergleich der Arable-Stationen untereinander liefert vergleichbare Ergebnisse für alle drei Klimavariablen. Ausgeprägte Abweichungen treten nur im Vergleich mit der DWD-Station auf. Die Abbildungen 25 (S. 97), 27 (S. 99) und 29 (S. 101) stellen den gleichen Datensatz dar, jedoch wurde dieser um dem Zeitraum 4. bis 10. Mai bereinigt. Das Ergebnis sind stark vergleichbare Histogramme zwischen allen Stationen. Besonders für die Temperatur- und Niederschlagswerte sind die verzeichneten Abweichungen zur DWD-Station in den Abbildungen 25 (S. 97) und 29 (S. 101) kaum noch vorhanden. Die bereits aufgestellten Abweichungen der DWD mit den Arable-Stationen in den Histogrammen bleibt jedoch trotz der Bereinigung bestehen (Abbildung 27 (S. 99) ).

#### 4.2.2.4 Vergleich der durchgeführten Regressionsanalysen

Aufbauend auf diesen Erkenntnissen wurde die Berechnung der Regressionskoeffizienten ohne den Zeitraum vom 4. bis 10. Mai durchgeführt. Auf diese Weise kann ausgeschlossen werden, dass Messwerte einfließen, die aufgrund von Sensorkalibrierungen oder anderen Einflüssen von der Realität abweichen. Warum dieser Zeitraum zu Beginn der Messungen solche Unterschiede zwischen den Wetterdatensätzen aufweist, wird im Unterkapitel 5.2 erörtert.

Tabelle 7 (S. 58) trägt die ermittelten Korrelationskoeffizienten und Bestimmtheitsmaße paarweise ab. Zu lesen ist die Tabelle dabei so, dass zum Beispiel der Korrelationskoeffizienten  $r = 0,994$  in der dritten Zeile, vierten Spalte jener der linearen Regression der Stationen Arable 60 und Arable 61 entspricht. Die hervorgehobenen Korrelationskoeffizienten und Bestimmtheitsmaße sind jene, zwischen

Tabelle 7: Regressionsanalyse zwischen den Stationen *Arable Bentzin* und *DWD Greifswald* für die Lufttemperaturwerte (Eigene Darstellung).

	<b>DWD Greifswald</b>	<b>Arable 60</b>	<b>Arable 61</b>	<b>Arable 62</b>	<b>Arable 63</b>
<b>DWD Greifswald</b>	$r = 1,000$ $r^2 = 1,000$	$r = 0,940$ $r^2 = 0,884$	$r = 0,942$ $r^2 = 0,887$	$r = 0,946$ $r^2 = 0,895$	$r = 0,951$ $r^2 = 0,904$
<b>Arable 60</b>	$r = 0,940$ $r^2 = 0,884$	$r = 1,000$ $r^2 = 1,000$	$r = 0,994$ $r^2 = 0,988$	$r = 0,995$ $r^2 = 0,990$	$r = 0,994$ $r^2 = 0,988$
<b>Arable 61</b>	$r = 0,942$ $r^2 = 0,887$	$r = 0,994$ $r^2 = 0,988$	$r = 1,000$ $r^2 = 1,000$	$r = 0,994$ $r^2 = 0,988$	$r = 0,994$ $r^2 = 0,988$
<b>Arable 62</b>	$r = 0,946$ $r^2 = 0,895$	$r = 0,995$ $r^2 = 0,990$	$r = 0,994$ $r^2 = 0,988$	$r = 1,000$ $r^2 = 1,000$	$r = 1,000$ $r^2 = 0,986$
<b>Arable 63</b>	$r = 0,951$ $r^2 = 0,904$	$r = 0,994$ $r^2 = 0,988$	$r = 0,994$ $r^2 = 0,988$	$r = 0,993$ $r^2 = 0,986$	$r = 1,000$ $r^2 = 1,000$

der DWD-Station Greifswald und den einzelnen Arable-Stationen.

Im Vergleich zwischen der DWD-Station mit der Arable-Station **60** besteht ein starker linearer Zusammenhang zwischen den Variablen von 0,94. Hohe Temperaturwerte in der DWD Messung gehen mit vergleichbar hohen Werten bei der Arable-Station einher. Diese Aussage spiegelt sich in dem Bestimmtheitsmaß von 0,88 wider. 88 % der Varianz kann von dem linearen Modell erklärt werden. Der höchste Regressionskoeffizient tritt zwischen der DWD-Station und Arable **63** auf und beträgt gerundet 0,951. Die Werte der Arable-Stationen untereinander liegen, deutlich näher beieinander. Der niedrigste Korrelationskoeffizient tritt zwischen Arable **62** und **63** auf und beträgt 0,993, während die anderen sich auf 0,994 oder 0,995 belaufen. Dieser Eindruck spiegelt sich in der Kurvenverlaufdarstellung der Temperaturwerte in Abbildung 12 (S. 56) wider.

Abbildung 30 im Anhang auf Seite 102 stellt diese Tabelle als Regressionsdiagramme dar. Der Aufbau und die Lesart ist dabei identisch zu dieser Tabelle und stellt die linearen Zusammenhänge bildlich dar. Die Lufttemperaturwerte der Arable-Stationen untereinander liegen eng beieinander und bilden eine dünne, diagonale Linie. Starke Ausreißerwerte treten keine auf. Im Vergleich mit der DWD-Station streuen diese etwas stärker, doch auch für diese reihen sich die Einzelpunkte der Regressionsgerade entlang auf und spiegeln die hohen Korrelationskoeffizienten und Bestimmtheitsmaße wider.

Für die Niederschlagswerte bestätigt sich das bereits gezogene Ergebnis, dass die DWD-Station verglichen mit den Arable-Stationen weniger Gemeinsamkeiten aufweisen, als die Arable-Stationen untereinander. Für die Arable-Stationen untereinander ergeben sich Korrelationskoeffizienten, die

Tabelle 8: Regressionsanalyse zwischen den Stationen *Arable Bentzin* und *DWD Greifswald* für die Niederschlagswerte (Eigene Darstellung).

	<b>DWD Greifswald</b>	<b>Arable 60</b>	<b>Arable 61</b>	<b>Arable 62</b>	<b>Arable 63</b>
<b>DWD Greifswald</b>	$r = 1,000$ $r^2 = 1,000$	$r = 0,2777$ $r^2 = 0,077$	$r = 0,182$ $r^2 = 0,033$	$r = 0,303$ $r^2 = 0,092$	$r = 0,006$ $r^2 = 0,000$
<b>Arable 60</b>	$r = 0,277$ $r^2 = 0,077$	$r = 1,000$ $r^2 = 1,000$	$r = 0,785$ $r^2 = 0,616$	$r = 0,926$ $r^2 = 0,857$	$r = 0,576$ $r^2 = 0,332$
<b>Arable 61</b>	$r = 0,182$ $r^2 = 0,033$	$r = 0,785$ $r^2 = 0,616$	$r = 1,000$ $r^2 = 1,000$	$r = 0,814$ $r^2 = 0,663$	$r = 0,840$ $r^2 = 0,706$
<b>Arable 62</b>	$r = 0,303$ $r^2 = 0,092$	$r = 0,926$ $r^2 = 0,857$	$r = 0,814$ $r^2 = 0,663$	$r = 1,000$ $r^2 = 1,000$	$r = 0,571$ $r^2 = 0,326$
<b>Arable 63</b>	$r = 0,006$ $r^2 = 0,000$	$r = 0,576$ $r^2 = 0,332$	$r = 0,840$ $r^2 = 0,706$	$r = 0,571$ $r^2 = 0,326$	$r = 1,000$ $r^2 = 1,000$

zwischen 0,571 und 0,926 liegen. Der geringste lineare Zusammenhang scheint dabei zwischen den Stationen **62** und **63** (0,571) wie **60** und **63** (0,576) zu bestehen. Das nächste Wertepaar (**60** und **61**) erreicht bereits einen Wert von 0,785, gefolgt von 0,814 (**61** und **62**) und 0,84 (**61** und **63**). Die größte Korrelationskoeffizient tritt zwischen Arable **60** und **62** auf und beträgt 0,926. Die Arable-Stationen untereinander liefern Ergebnisse, die vergleichbar sind, jedoch in den Ausprägungen der gemessenen Niederschlagsereignisse voneinander abweichen. Dieser Eindruck wurde bereits von Abbildung 13 (S. 57) vermittelt, die Ähnlichkeiten und Unterschiede in der Signalstärke einzelner Stationspaare aufzeigte, die sich in diesen Werten nun wiederfinden.

Abbildung 31 im Anhang auf Seite 103 stellt diese beschriebenen linearen Zusammenhänge grafisch dar. Anhand dieser lässt sich für den Vergleich der Arable-Stationen untereinander festhalten, dass sie eine starke Ansammlung an Einzelpunkten nahe 0|0 aufweisen. Die wenigen gemessenen Niederschlagsereignisse reihen sich einer gedachten diagonalen Regressionsgeraden entlang auf, wenn für diese Stationspaare hohe Korrelationskoeffizienten und Bestimmtheitsmaße ausgegeben werden. Beispiele hierfür sind die Stationen **60** und **62** oder **61** und **63**. Stationen mit niedrigen Regressionsmaßen weisen optisch eine stärkere Streuung auf. Dies ist zum Beispiel für **60** und **63** der Fall, wo hohe Niederschlagswerte bei der einen Station mit vergleichbar niedrigen Werten bei der anderen einhergehen.

Für den Vergleich der DWD-Station mit den Arable-Stationen ergibt sich ein abweichendes Bild. Der Großteil der Werte sammelt sich zwar um 0|0 herum, doch die ausschlaggebenden Niederschlagsereignisse stimmen selten überein. Hohe Werte bei den einzelnen Arable-Stationen gehen mit niedrigen DWD Werten einher und umgekehrt. Dementsprechend fallen die Korrelationskoeffizienten innerhalb der Tabelle 8 niedrig aus und betragen: 0,277 (DWD / Arable 60), 0,182 (DWD / Arable 61), 0,303 (DWD / Arable 62) und 0,006 (DWD / Arable 63). Zwischen den Niederschlagswerten der zwei verwendeten Datensätze besteht kein linearer Zusammenhang.

Dieses Kapitel abschließend folgt nun die Betrachtung der Regressionsanalyse der relativen Luftfeuchtigkeitswerte. Zur Erinnerung: Anhand der vorangegangenen Ergebnisdarstellung anhand der Boxplots (Abbildung 11 (S. 55)), dem Liniendiagramm (Abbildung 14 (S. 57)) und den Histogrammen (Abbildung 29 (S. 101)), wurde herauskristallisiert, dass die relativen Luftfeuchtigkeitswerte der Arable-Stationen untereinander vergleichbare Ergebnisse erzielen, wohingegen zur DWD-Station in Greifswald ausgeprägtere Abweichungen auftreten. Die nachfolgende Tabelle 9 (S. 61) und Abbildung 32 im Anhang auf Seite 104 liefern an dieser Stelle Auskunft darüber, wie stark diese Abweichungen und Zusammenhänge tatsächlich sind.

Für den Vergleich zwischen Greifswald und den Arable-Stationen ergeben sich Regressionskoeff-

Tabelle 9: Regressionsanalyse zwischen den Stationen *Arable Bentzin* und *DWD Greifswald* für die relativen Luftfeuchtwerte (Eigene Darstellung).

	<b>DWD Greifswald</b>	<b>Arable 60</b>	<b>Arable 61</b>	<b>Arable 62</b>	<b>Arable 63</b>
<b>DWD Greifswald</b>	$r = 1,000$ $r^2 = 1,000$	$r = 0,880$ $r^2 = 0,774$	$r = 0,879$ $r^2 = 0,773$	$r = 0,883$ $r^2 = 0,780$	$r = 0,886$ $r^2 = 0,785$
<b>Arable 60</b>	$r = 0,880$ $r^2 = 0,774$	$r = 1,000$ $r^2 = 1,000$	$r = 0,993$ $r^2 = 0,986$	$r = 0,993$ $r^2 = 0,986$	$r = 0,992$ $r^2 = 0,984$
<b>Arable 61</b>	$r = 0,879$ $r^2 = 0,773$	$r = 0,993$ $r^2 = 0,986$	$r = 1,000$ $r^2 = 1,000$	$r = 0,991$ $r^2 = 0,982$	$r = 0,993$ $r^2 = 0,986$
<b>Arable 62</b>	$r = 0,883$ $r^2 = 0,780$	$r = 0,993$ $r^2 = 0,986$	$r = 0,991$ $r^2 = 0,982$	$r = 1,000$ $r^2 = 1,000$	$r = 0,992$ $r^2 = 0,984$
<b>Arable 63</b>	$r = 0,886$ $r^2 = 0,785$	$r = 0,992$ $r^2 = 0,984$	$r = 0,993$ $r^2 = 0,986$	$r = 0,992$ $r^2 = 0,984$	$r = 1,000$ $r^2 = 1,000$

fizienten von 0,88 (DWD / Arable 60), 0,879 (DWD / Arable 61), 0,883 (DWD / Arable 62) und 0,886 (DWD / Arable 63). Zwischen den gemessenen relativen Luftfeuchtwerten der DWD-Station verglichen mit den einzelnen Arable-Stationen besteht jeweils ein ausgeprägter positiver Zusammenhang. Hohe Messwerte der DWD-Station gehen eher mit vergleichbar hohen Messwerten bei den einzelnen Arable-Stationen einher. Wie den Bestimmtheitsmaßen von beispielsweise 0,774 (DWD / Arable 60) entnommen werden kann, werden die einzelnen Punktpaare innerhalb der grafischen Darstellung dieser Zusammenhänge zu einem gewissen Teil streuen und somit die Messergebnisse der DWD-Station von den einzelnen Arable-Stationen abweichen. Ein Blick auf Abbildung 32 veranschaulicht diese Aussage. Der Großteil der DWD-Stationswerte geht mit vergleichbar ausgeprägten Werten für die relative Luftfeuchtigkeit bei den Arable-Stationen einher. Jedoch streuen die Punktpaare um die gedachte, diagonale Regressionsgerade herum, wodurch die einzelnen Regressionsgraphen im Bereich zwischen 30 und 60 Prozentpunkte stärkere Variationen zu verzeichnen haben und die Punktwolken bauchig wirken. Wie anhand des Vergleiches zwischen der DWD-Station und Arable **60** ersichtlich wird, kommen Ausreißer vor. Niedrige Messergebnisse der Station **60** von 30 % gehen mit hohen Werten im DWD-Datensatz von bis zu 80 % einher. Solche Ausreißer sind für jeden Vergleich mit einer Arable-Station zu verzeichnen.

Abweichend dazu weisen die Arable-Stationen innerhalb der Regressionsgraphen untereinander vergleichbar starke Ähnlichkeiten zu jenen der Temperaturwerte in Abbildung 30 auf. Die Wertepaare streuen bei gemessenen Bestimmtheitsmaßen von zum Beispiel 0,986 also 98,6 % für die Stationen **60** und **62** nur geringfügig. Die linearen Regressionsmodelle können die Varianz nahezu perfekt beschreiben. Aufgrund der Korrelationskoeffizienten, die allesamt zwischen 0,991 und

0,993 liegen, reihen sich die Punktpaare eng entlang der gedachten, diagonal verlaufenden Regressionsgeraden auf. Die Werte der relativen Luftfeuchtigkeit fallen für die Arable-Stationen nahezu deckungsgleich aus.

## 5 Diskussion

Der Dreh- und Angelpunkt dieser Arbeit ist neben dem Programmieren eines R-Skriptes, das die aufgezeichneten Messdaten der eingesetzten Arable-Mark-2 Sensoren automatisch anhand gegebener Richtwerte und Bedingungen qualitativ vor überprüft, der Vergleich dieser mit bereits etablierten und zur Verfügung stehenden Datensätzen; in diesem Fall der Fokus auf die Datenquelle des DWD. Dieser Vergleich setzte ursprünglich nur auf den Stationsvergleich zwischen Erdeborn und Querfurt, wurde jedoch während des Schreibens dieser Zeilen um den Vergleich zwischen den Arable-Stationen in Bentzin mit der DWD-Station Greifswald erweitert; glücklicherweise, da sich so Aspekte des Arable-Datensatzes hervorheben lassen, die bei der Station in Erdeborn nicht beobachtbar sind. Allem voran die Besonderheit der Messwerte zwischen dem 04.05. bis 10.05.2022, die für alle vier Bentzin Stationen zu verzeichnen ist: das Auftreten von Werten, die sich auch bei der Betrachtung der Stundenwerte für knapp eine Woche nicht grundlegend verändern. Die stündlichen Temperaturwerte schwanken in dieser Zeit zwischen 20 und 22 °C, wohingegen die Messwerte der relativen Feuchtigkeit zwischen 40 und 45 Prozentpunkte angesiedelt sind. Eine genaue Ursache für diesen Effekt kann an dieser Stelle nicht genannt werden, da der Autor nicht an der Aufstellung der Sensoren beteiligt war, jedoch soll eine mögliche Ursache benannt werden, die in ihrer Nachvollziehbarkeit schwer zu überprüfen ist: Auf welche Art und Weise werden die Messergebnisse genau erhoben, beziehungsweise wie stark werden diese korrigiert?

Als Rückgriff auf **3.1 Datengrundlage** wird daran erinnert, dass sowohl die Messwerte der Lufttemperatur als auch der relativen Luftfeuchtigkeit AI-gestützt sind (Arable-Labs 2021, S. 1). Inwiefern diese potenziellen Korrekturen die Rohdaten beeinflussen, ist nicht in Erfahrung zu bringen. Für die Niederschlagswerte gilt, dass diese nicht physisch aufgefangen und die Menge des Niederschlages auf diese oder ähnliche Weisen gemessen wird, sondern stattdessen wird die Niederschlagsmenge anhand akustischer Signale, die Regentropfen erzeugen, berechnet (Arable-Labs 2021, S. 2). Die genauen Prozesse dieses Messverfahrens sind nicht bekannt und so kann sich anhand der noch zu diskutierenden Ergebnisdarstellung die Frage bilden: Sind die unterschiedlichen Ergebnisse der ortsnah aufgestellten Bentzin Messstationen bedingt durch dieses Verfahren (also Sensorungenauigkeiten) oder auf tatsächlich gemessene Differenzen der Niederschlagsstärke auf kleinem Raum zurückzuführen?

Eine Antwort auf diese Fragen kann aufgrund fehlender Informationen nicht formuliert werden. Von Bedeutung ist an dieser Stelle zu wissen, dass diese Unbekannte existiert und auf Basis der Vergleiche mit anderen Ortschaften nicht ersichtlich wird, warum sich die Stundenwerte der Temperatur und der relativen Luftfeuchtigkeit für die vier Bentzin Stationen so verhalten, wie sie es tun. Ausführlicher werden diese Ergebnisse jedoch unter **5.2** diskutiert. Worauf stattdessen einleitend der Fokus gelenkt werden soll, ist eine potenzielle Schwäche des Datenqualitätsskriptes und seiner Testverfahren, die, unabhängig davon ob die Messwerte des Zeitraumes 04.05. bis 10.05.2022 korrekt sind oder nicht, nichts an diesen Werten auszusetzen haben. Sie werden nicht als fragwürdig klassifiziert.

Wie vorgestellt wurde, arbeiten die Testverfahren mit Grenz- und Schwellwerte. Je nachdem ob diese über- oder unterschritten werden, werden den Stundenwerten passende Qualitätsschlüssel vergeben, die diese als valide, fragwürdig oder fehlend markieren. Was die Testverfahren nicht überprüfen können, ist, ob die gemessenen Werte tatsächlich der Realität entsprechen. Solange sie im Rahmen der Grenz- und Schwellenwertprüfung keine Auffälligkeiten aufweisen, werden diese als valide markiert. Dieser Aspekt zeigt eine wichtige Komponente der Datenqualitätsüberprüfung auf, die grundlegender Bestandteil dieser ist: automatische Kontrollmethoden liefern erste Richtwerte bezüglich der Datenqualität, jedoch muss ein solcher Datensatz idealerweise immer auch von einem Experten gesichtet und eventuell die jeweils vergebenen Flags angepasst werden, wenn solche oder andere Aspekte beobachtet werden (Cai und Zhu 2015, S. 5 & 8; WMO 2021, S. 5). Dieser Prozess der erweiterten Qualitätskontrolle (*Extended Quality Control Procedures*) kann auf den gleichen, bereits eingesetzten Methoden aufbauen (jedoch sollten diese dann anhand langer Zeiträume durchgeführt werden), weitere Testverfahren heranziehen, die die Daten im Detail überprüfen können, oder aber auch statistische Methoden verwenden, um Besonderheiten wie Ausreißer und Ähnliches herauszuarbeiten (Zahumenský 2002, S. 7).

## **5.1 Auswahl der angewandten Methoden**

Bevor an dieser Stelle weiter über die Güte der erzielten Ergebnisse und des Arable-Datensatzes im Allgemeinen gesprochen wird, muss eine wichtige Thematik vorgezogen werden, die bereits im Methodenkapitel angeschnitten wurde: Warum wurden die vorgestellten Methoden angewendet und nicht andere?

Eine grundlegende Problematik bei der Auswahl der Testverfahren bezüglich der Qualität von Wetterdaten ergab sich bereits ganz zu Beginn der Recherche und manifestierte sich in Form ungenauer Methodenbeschreibungen innerhalb der gesichteten Literatur. Testverfahren wurden genannt, jedoch nicht beschrieben, wie diese angewendet wurden. Einziger Nenner dieser war die in unter-

schiedlichen Formen auftretende Aussage, dass alle Entscheidungen bezüglich der Datenqualitätsüberprüfung auf Richtlinien der WMO aufbauen. 2021 veröffentlichte diese den Leitfaden *Guidelines on Surface Station Data Quality Control and Quality Assurance for Climate Applications*, welcher den Leitfaden *Guidelines on the Quality Control of Surface Climatological Data* (WMO/TD-No. 111) aus dem Jahr 1986 ersetzt (WMO 2021, S. 1). In diesem Werk werden nicht nur grundlegende Richtlinien für das Erheben von Wetterdaten und deren Überprüfung aufgestellt, sondern auch 41 unterschiedliche Testverfahren benannt und beschrieben (siehe: WMO 2021, 24 ff.). Diese Tests gelten entweder als obligatorisch, empfohlen oder optional. Die in dieser Arbeit verwendeten Verfahren gelten alle als obligatorisch und wurden, wenn keine Grenzwerte für die einzelnen Testverfahren genannt wurden, in Anlehnung an den *Guidelines on Quality Control Procedures for Data from Automatic Weather Stations* oder aber auch Angaben des DWD und weiteren Autoren ermittelt (siehe: Zahumenský 2002; DWD 2022a; Zimmermann et al., 2015; WMO 2021). Die Auswahl der Methoden erfolgte anhand der Einzelbeschreibungen dieser, wobei einige der in der Richtlinie von 2021 genannten, obligatorischen Verfahren fallen gelassen werden mussten, da sie Voraussetzungen an den Datensatz stellen, der nicht erfüllt werden konnte. So benötigt es zum Beispiel für den *Domain Test*, der eine Ergänzung der bereits getätigten **klimatestischen Überprüfung (3.2.3)** darstellt, einen Datensatz, der mindestens einen Zeitraum von 30 Jahren abdecken kann (WMO 2021, S. 28). Anhand eines solchen Datensatzes wird für jeden Wert überprüft, ob dieser zwischen der 0,3 und 99,7 Perzentile aller Werte der letzten 30 Jahre fällt (WMO 2021, S. 28). Die zeitliche Begrenzung des Arable-Datensatzes verhindert automatisch die Anwendbarkeit einiger Testverfahren, die auf Datensätze wie nationale Wetterstationen des DWD angewendet werden könnten. Andere Testverfahren wie *zero precipitation spatial test* setzen voraus, dass Nachbarstationen gegeben sind und überprüft, ob ein starkes Niederschlagsereignis, das von einer Station verzeichnet wurde, auch von Nachbarstationen beobachtet wurde (WMO 2021, S. 30). Ist dem nicht so, kann das darauf hindeuten, dass dieser Wert für den falschen Tag eingetragen wurde oder eine Aufsummierung des Signals über mehrere Tage hinweg entspricht (WMO 2021, S. 30). Im Fall des Arable-Datensatzes scheitert die Anwendung dieser Methode anhand der Tatsache, dass nicht immer eine Arable Nachbarstation in direkter Umgebung gegeben ist. Deswegen ist *zero precipitation spatial test* eines von jenen Verfahren, die wie eingangs erwähnt von einem Anwender der Daten eingesetzt werden sollte, um diese im Detail zu überprüfen (WMO 2021, S. 5).

Die angewandten Methoden stellen somit eine Auswahl jener typischen Datenqualitätsmethoden dar, die auf den zeitlich und räumlich begrenzten Arable-Datensatz angewandt werden können und diesen auf obligatorische Bedingungen getestet. Sie zählen zu den grundlegenden Testverfahren, die bei manuellen und automatischen Erhebungsverfahren durchgeführt werden sollen (WMO 2021,

S. 10). Würden einzelne Stationen innerhalb dieses Datensatzes über Jahre hinweg in Betrieb sein und Daten generieren, wäre es ratsam, weitere Testverfahren wie den Domain-Test hinzuzunehmen, um die Konsistenz der Daten auch über einen längeren Zeitraum als wenige Monate oder Jahre prüfen zu können. Ferner gelten die in **3.2 Datenqualitätsüberprüfung** genannten Grenzwerte der einzelnen Testverfahren nur für die gegenwärtigen klimatischen Bedingungen. Sollte es in der Zukunft zu starken Veränderungen der zum Beispiel durchschnittlichen Höchsttemperaturen im Sommer oder im Jahresverlauf kommen, so müssten diese Werte gegebenenfalls angepasst werden.

## 5.2 Ergebnisdiskussion

Die dargestellten Ergebnisse in Form von Tabellen, Abbildungen und schriftlichen Ausführungen für die beiden Themenfelder dieser Arbeit erfolgte in ausführlicher Form, jedoch stehen sie derzeit noch unangefochten im Raum. Um diese kritisch betrachten zu können, wird im Zuge der nachfolgenden Auseinandersetzung Literatur herangezogen werden, die auf wissenschaftlichen Untersuchungen aufbauen, die sich ebenfalls mit den Arable-Mark-2 Sensoren und den von ihnen erhobenen Daten auseinandergesetzt haben. Weiterhin sollen die Ergebnisse vor dem Hintergrund der eingangs formulierten Hypothesen überprüft und diskutiert werden.

96 % aller Temperaturwerte, 97 % aller Niederschlagswerte und 91 % aller relativen Luftfeuchtigkeitswerte gelten aufgerundet als valide (Tabelle 2). Wie herausgearbeitet werden konnte, ging dabei der Großteil der als fragwürdig geflaggt Werte auf die zeitliche Konsistenzprüfung zurück. Um es in Erinnerung zu rufen: diese überprüft, wie groß der Abstand zwischen zwei aufeinander folgenden Stundenwerte ist und wenn dieser einen festgelegten Grenzwert überschreitet, wird der derzeit betrachtete Wert als fragwürdig (-777) klassifiziert. Für die Lufttemperatur wurde hierbei ein Grenzwert von  $\pm 4^\circ$  Celsius (Cerlini et al., 2020, S. 6) und für die relative Luftfeuchtigkeit ein Wert von 10 % also 0.1 pro Stunde angesetzt (Zahumenský 2002, S. 6). Dieses Testverfahren verlangt die Überprüfung der Werte von einem Experten. Wenn ein Wert als fragwürdig gekennzeichnet wurde, dann immer, weil der Abstand zwischen diesen und dem vorangegangenen Wert zu groß ist. Jedoch kann dieser Abstand von unterschiedlichen Faktoren beeinflusst werden. Ungenauigkeiten der Sensoren oder auch die Messerhebungsmethoden als solche, die AI-gestützt korrigiert werden, können Auswirkungen haben. Wurde ein Wert mit -777 geflaggt, bedeutet das lediglich, dass manuell eingeschätzt werden muss, ob dieser Wert als plausibel gilt. Diese **Expertenüberprüfung** ist ein zentraler Aspekt der Datenqualitätsüberprüfung und muss idealtypisch immer durchgeführt werden (WMO 2021, S. 5). Wenn dem nicht möglich ist, weil es ein sehr zeit- und arbeitsaufwendiger Schritt

ist, dann müssen mindestens die fragwürdige -777 Werte manuell gesichtet werden (WMO 2021, S. 5)

Basierend auf dem, was die Testverfahren überprüfen, wird das Ergebnis der Datenqualitätsüberprüfung als gut eingeschätzt. Jedoch muss dabei betont werden, dass eine weiterführende Datenqualitätsüberprüfung bei der Anwendung dieser Daten durchgeführt werden sollte, um durch etwa statistische Verfahren Ungereimtheiten innerhalb der Daten, die von dem Skript nicht überprüft werden können, festzustellen. Wie anhand des Bentzin und Greifswald Vergleich für den Zeitraum vom 4. bis zum 10. Mai 2022 aufgezeigt wurde, ist dieser Schritt für die Anwendung und Analyse der Daten von großer Bedeutung.

Das Einschätzen der Güte der Messungen anhand der Stationen des DWD gestaltet sich komplizierter. Unterschiedliche Faktoren wie die Distanz der Stationen zwischen einander oder die unterschiedlichen Sensoren, die für die Messungen verwendet werden, können Einflüsse auf die Ergebnisse haben. Die getätigten Vergleiche zwischen den Stationsstandortpaare Erdeborn und Querfurt oder Bentzin und Greifswald erlauben im Detail keine Aussagen darüber, welche Stationsart bessere Ergebnisse liefert. Selbst wenn eine Arable und eine DWD-Station in direkter unmittelbarer Nähe zueinanderstehen würden, verwenden sie immer noch unterschiedliche Sensoren, die unterschiedlich kalibriert sein können, wodurch die Ergebnisse voneinander abweichen können. Das Ziel dieses Vergleiches ist zum Glück jedoch nicht, zu entscheiden, welche Stationsart bessere Ergebnisse erzielt, sondern dient der Überprüfung der Arable Messergebnisse und ob diese den DWD Messwerten gleichen. Für den Erdeborn-Querfurt Vergleich wurde die Hypothese angesetzt, dass je stärker die Ergebnisse einander gleichen, desto genauer messen die Arable-Mark-2 Sensoren ihre Umwelt. Diese Annahme wurde dank der kurzfristigen Hinzunahme der Betrachtung der Bentzin Arable-Stationen mit der DWD-Station in Greifswald erweitert, um einen Eindruck darüber gewinnen zu können, wie vergleichbar die Messergebnisse mehrerer Arable-Mark-2 Sensoren untereinander ausfallen, wenn diese ortsnah aufgestellt sind.

Vergleiche der Lufttemperaturmesswerte weisen über alle Stationspaare hinweg starke Gemeinsamkeiten auf. Sowohl die Vergleiche von täglichen Durchschnittswerten als auch Stundenwerten erzielen sehr hohe Korrelationen und würden es zum Beispiel im Fall der Bentzin Stationen ermöglichen, beim kurzfristigen Ausfall einer Station die Nachbarstationen zum Interpolieren der Fehlwerte zu verwenden (Xu et al., 2013, 7452 ff.). Für die relative Luftfeuchtigkeit ergibt sich, dass die Bentzin Stationen untereinander stark vergleichbar sind und nahezu das Gleiche messen, wohingegen geringe Unterschiede zu der DWD-Station Greifswald auftreten. Selbiges gilt auch für das Stationspaar Erdeborn-Querfurt und ist ein zu erwartendes Ergebnis. Sensoren und räumliche Distanzen

von mehreren Kilometern zwischen den Stationen können Verschiebungen der Messwerte zur Folge haben (Abimbola et al., 2022, S. 6–7). Stärkere Abweichungen treten beim Niederschlag auf. Wenn Niederschlagsereignisse auftreten, werden sie zwar fast immer von allen betrachteten Stationen registriert, jedoch schwanken sie teils stark in der Ausprägung dieser. Gerade an dieser Stelle muss die bereits getätigte Aussage wiederholt werden, dass der Vergleich keine Aussage über die Korrektheit der Messwerte treffen kann. Die Stationen sind im Fall von Erdeborn und Querfurt circa 10 km voneinander entfernt, wohingegen die Stationen Bentzins und Greifswald auf circa 15 km kommen. Nach Jensen und Pedersen (2005, S. 269 & 276–277) kann es bereits bei einer Distanz von 0,5 km zu Abweichungen von bis zu 100 % zwischen zwei Stationen kommen. Kleinere Distanzen von nur 250 m können darüber hinaus Unterschiede von bis zu 26 % hervorrufen (Pedersen et al., 2010, S. 449). Niederschlagsmessungen zwischen Stationen zu vergleichen, ist somit immer mit Ungenauigkeiten verbunden. Und selbst die Ungenauigkeiten zwischen den Arable-Stationen in Bentzin untereinander müssen nicht weniger oder mehr korrekt sein, sondern können aufgrund der Distanzen von 100 bis 200 m zwischeneinander oder den mehrfach erwähnten Messungenauigkeiten auftreten.

Weil die Ausprägungen der Klimavariablen Lufttemperatur und relative Luftfeuchtigkeit zwischen Arable-Stationen, aber auch im Vergleich zu DWD-Stationen teils stark vergleichbare Ergebnisse erzielen und das, obwohl sie mehrere Kilometer voneinander entfernt sind und nicht unter den gleichen Bedingungen erhoben werden (die untersuchten Arable-Stationen stehen zum Beispiel auf Ackerflächen, wohingegen die DWD-Stationen stadtnah aufgestellt sind), können die mit dem Arable-Mark-2 Sensoren erhobenen Daten als gut eingeschätzt werden. Die Überprüfung des Niederschlages gestaltet sich im Stationsvergleich als schwieriger und lässt keine allgemeingültige Antwort bezüglich der Güte der Messungen zu, jedoch erscheinen sie plausibel. Ortsnahe Arable-Stationen untereinander verzeichnen die gleichen Niederschlagsereignisse und erzielen im Schnitt Regressionskoeffizienten von 0,785 bis 0,926 untereinander. Vergleichbare Niederschlagsereignisse werden zur gleichen Zeit auch von weiter entfernten DWD-Stationen gemessen. Ob es sich bei den Messwerten der Bentzin Stationen vom 04.05. bis 10.04.2022 um realitätsnahe Bedingungen handelt, lässt sich schwer überprüfen. Vergleich der Stundenwerte zwischen der DWD-Station Greifswald oder externen Wetterarchiven von Wetterstationen von nächstgelegenen Ortschaften wie Demmin zeigen auf, dass diese stärkere Variationen des stündlichen Temperatursignals aufweisen (siehe: Timeanddate 2022). Die Zweifel an diesen Ergebnissen werden dabei nicht dadurch hervorgerufen, dass diese von den anderen Datenquellen abweichen, sondern weil sie im Fall der Lufttemperatur sich über Tage hinweg kaum ändert, nur um dann schlagartig sich den Werten der DWD-Station Greifswald anzugleichen (siehe: Abbildungen 12 und 14). Diese Beobachtung ist über beide Vergleiche hinweg einmalig.

Vergleichbare Ergebnisse anderer Forschungsprojekte zu finden, gestaltet sich aufgrund der relativen Neuheit der Arable-Mark-2 Sensoren als schwierig. Die einzig vergleichbare Studie, die gefunden werden konnte, ist der Artikel *Improving crop modeling to better simulate maize yield variability under different irrigation managements* von Abimbola et al., (2022). In diesem wird untersucht, wie gut sich der Maisertrag anhand von in situ Daten vorhersagen lässt (Abimbola et al., 2022, S. 1). Solche Modellierungsversuche, bezogen auf den Ertrag von Nutzpflanzen innerhalb der Landwirtschaft, werden üblicherweise anhand von Daten generiert, die nicht die lokalen Bedingungen widerspiegeln müssen (Abimbola et al., 2022, S. 1). Dieser Einsatz von Makroklimadaten oder Daten von nahegelegenen Wetterstationen führt zu einer Über- oder Unterschätzung der Erträge (Abimbola et al., 2022, S. 1). Einer der genannten Gründe für diese Abweichungen sind zum einen die mikroklimatischen Unterschiede zwischen (in diesem Fall) den zu untersuchenden Ackerflächen und der nächstgelegenen nationalen Wetterstation oder aber auch Unterschiede in der Sensorkalibrierung und Sensorart (Vergleich: Abimbola et al., 2022, S. 2; mit: Pedersen et al., 2010, S. 449; Jensen und Pedersen 2005, S. 276–277). Für die erarbeiteten Ergebnisse dieser Masterarbeit sind die von den Autoren angestellten Vergleiche zwischen den Messwerten der von ihnen eingesetzten Arable Mark-1 und Mark-2 Sensoren mit ortsnahen Wetterstationen von Relevanz. Die Ergebnisse dieser angestellten Vergleiche besagen, dass die Lufttemperaturmessungen starke Ähnlichkeiten aufweisen. Die Durchführung von linearen Regressionen erzielt ein Bestimmtheitsmaß  $r^2$  von 0,93 (Abimbola et al., 2022, S. 6). Die relativen Luftfeuchtigkeitswerte erzielen ein  $r^2$  von 0,71, was einem Korrelationskoeffizienten von circa 0,84 entspricht (Abimbola et al., 2022, S. 6). Ausgeprägte Abweichungen zwischen den untersuchten Wetterstationen ergeben sich für die Niederschlagswerte. Als Ursache werden dabei die unterschiedlichen Messverfahren und die räumliche Distanz zwischen den Stationen genannt (Abimbola et al., 2022, S. 7).

Vergleicht man diese Ergebnisse mit jenen dieser Masterarbeit und jenen, die Arable Labs selbst durchgeführt hat, um die Güte ihrer Sensoren zu testen (Arable-Labs 2020a, 8 ff.), lassen sich vergleichbare Aussagen ziehen: Die mit Arable-Mark-2 gemessenen Lufttemperaturwerte und auch die relativen Luftfeuchtigkeitswerte weisen starke Ähnlichkeiten zu anderen Wetterdatensätzen auf (wobei der lineare Zusammenhang zwischen den Lufttemperaturwerten stärker ausgeprägt ist) (Vergleich: Arable-Labs 2020a, S. 8, 11–12; Abimbola et al., 2022, S. 6). Die größten Abweichungen treten bei den Niederschlagswerten auf, doch liefern diese keine direkte Aussage über die Güte der Messungen. Im Fall der vorliegenden Masterarbeit und den Ergebnissen nach Abimbola et al., (2022) wäre es ratsam, einen solchen Vergleich zwischen den Arable-Mark-2 Sensoren und einer nationalen Wetterstation in direkter Nähe zueinander durchzuführen. Überdies wäre es für die Analy-

se der Güte der Messergebnisse, die nicht explizit durchgeführt wurde, ratsam, weitere Datenquellen wie Fernerkundungsdaten zurate zu ziehen (Verweis: Lembrechts, Nijs et al., 2019).

### 5.3 Hypothesenüberprüfung

Die getätigten und soeben geschilderten Schlussfolgerungen bezüglich der Qualität der programmierten R-Skripte, als auch der Messergebnisse der Arable-Mark-2 Sensoren als solche, bilden die Grundlage für die nachfolgende Hypothesenüberprüfung, die zu Beginn der Durchführung dieser Masterarbeit formuliert wurden.

Zuallererst galt es zu überprüfen, *ob der Einsatz von mobilen Wetterstationen wie die Arable-Stationen eine exaktere Abbildung der Realität erlauben, als es anhand von makroklimatischen Daten möglich ist* (Vergleich: Burnett und Anderson 2019; Kearney, Shamakhy et al., 2014; Kollas et al., 2014; Lembrechts, Nijs et al., 2019; Storlie et al., 2014; Zellweger et al., 2019)? Diese Hypothese lässt sich nur schwierig bestätigen. Messsensoren von Wetterstationen weisen unterschiedliche Genauigkeiten und Ungenauigkeiten auf und sind zum Beispiel für räumlich abhängige Klimavariablen wie dem Niederschlag stark von räumlichen Gegebenheiten abhängig und erschweren den Vergleich zwischen mehreren Stationen (Pedersen et al., 2010, 447 ff.). Ferner stellen Messungen nie die direkten, realen Bedingungen dar. Jede Messung ist ein Eingriff in das lokale Klimageschehen und das bloße Aufstellen von Messapparaten kann zu Verzerrungen der Ergebnisse führen (Stoutjesdijk und Barkman 2014, 171 ff.). Diese Messergebnisse sind jedoch auch abhängig von den eingesetzten Messgeräten und dem Objekt, das untersucht werden soll. Wie Maclean, Duffy et al., (2021, S. 1398) aufzeigen, kommt es allein bei der Messung der Lufttemperatur zu Problematiken (unter anderem Rückstrahlung von umliegenden Objekten wie Vegetation oder den Boden), die sich direkt auf das Ergebnis der Messungen auswirken können, je näher man dem Boden kommt. Diese Hypothese kann somit weder bejaht noch verneint werden. Sowohl direkte Messungen im Feld als auch die Modellierung anhand von makroklimatischen Daten stellen immer nur Annäherungen an die Wirklichkeit dar und sind behaftet mit Ungenauigkeiten (Fodor und Kovács 2005, S. 54–56).

Die zweite zu überprüfende Hypothese lautet *Wird die Datenqualität der zu überprüfenden Daten als gut befunden, dann werden die Abweichungen zwischen diesen und anderen Wetterstationen wie jenen des DWD nur geringe Messunterschiede aufweisen* und setzt sich aus den zwei methodischen Teilaspekten dieser Arbeit zusammen, deren Ergebnisse bereits vorgestellt werden. Als erste Bedingung gilt die Überprüfung der Datenqualität, die anhand des R-Skriptes durchgeführt wurde. Wie vorgestellt, wurden mehr als 95 % (Lufttemperatur und Niederschlag), beziehungsweise 91 % (relative Luftfeuchtigkeit) aller Werte der drei Klimavariablen als valide eingestuft. Als zweite Bedin-

gung gilt der Vergleich der Arable-Stationen mit den DWD-Stationen, die vergleichbare Ergebnisse für die Lufttemperatur und die relative Luftfeuchtigkeit erzielen, wohingegen die Ausprägung der Niederschlagsmessungen voneinander abweichen können.

Auf Basis der alleinigen Betrachtung des Vergleiches zwischen Erdebörn und Querfurt würde sich diese Hypothese zumindest für die Lufttemperatur und die relative Luftfeuchtigkeit bestätigen, wohingegen die Einschätzung der Güte der Niederschlagswerte aufgrund der hohen räumlichen Schwankungen schwieriger ausfällt (Abimbola et al., 2022, S. 7). Der Vergleich zwischen Bentzin und Greifswald zeigt weiterführend auf, dass, selbst wenn Messwerte als valide eingestuft werden, es zu deutlichen Abweichungen zwischen diesen und zum Beispiel DWD Wetterstationen kommen kann. Der hier angesprochene Zeitraum vom 04.05. bis 10.05.2022 ist hier zu nennen, dessen Verlauf der Stundenwerte als fragwürdig erscheinen, wenn man diese mit umliegenden Städten und Ortschaften wie Greifswald oder Demmin vergleicht. Anhand solcher Messergebnisse müsste die Hypothese eher verneint werden, da sie zwar von der Datenqualitätsüberprüfung als valide gekennzeichnet wurden, jedoch sehr auffällig von Vergleichsstationen abweichen. Da derlei Messergebnisse jedoch kaum im Arable-Datensatz gefunden werden konnten und sie für die Bentzin Station nur in den ersten Tagen auftreten, kann die Hypothese als solche bejaht werden. Wenn die Tests, die Messergebnisse als valide markieren, weisen sie bezogen auf die Lufttemperatur und die relative Luftfeuchtigkeit nur geringfügige Unterschiede zu anderen Datenquellen auf.

*Wenn die Arable-Mark-2 Sensoren vergleichbare Ergebnisse wie etablierte Wetterstationen generieren, können diese auch dort eingesetzt werden, wo keine großräumigen Daten erhoben werden, um qualitativ gute Daten gewinnen zu können* – Diese Hypothese lässt sich anhand der erzielten Ergebnisse und der vorangegangenen bejahen. Die Vergleiche zwischen den Arable und DWD-Stationen zeigen auf, dass diese besonders bezogen auf die Lufttemperatur und die relative Luftfeuchtigkeit sehr vergleichbare Ergebnisse erzielen. Selbst die Niederschlagswerte weisen, wenn sie von mehreren Arable-Stationen ortsnahe gemessen werden, Ähnlichkeiten untereinander auf (siehe Bentzin Ergebnisse). So werden dieselben Niederschlagsereignisse erfasst, deren unterschiedliche Ausprägung der Messsignale von der hohen Variabilität von Niederschlagsereignissen über bereits wenige hundert Meter erklärt werden kann (Jensen und Pedersen 2005, S. 269, 276–277; Pedersen et al., 2010, S. 449). Wenn die Arable-Stationen korrekt im Feld aufgestellt werden, dann sollten sie über die Zeit gute Messergebnisse der lokalen Bedingungen erheben können. Für die spätere Anwendung der Daten sollten diese immer anhand unterschiedlicher Methoden wie statistischen Analysen ausführlicher untersucht werden, um potenzielle Auffälligkeiten, wie Ausreißer daraufhin zu überprüfen, ob diese durch Extremwetterereignisse hervorgerufen wurden oder Messfehler aufweisen, weswegen sie als fragwürdig gekennzeichnet wurden. Wie die Untersuchung der Bentzin Stationen

gezeigt hat, können auch Auffälligkeiten auftreten, die von der Datenqualitätsüberprüfung nicht erfasst werden, die durch die Anwendung eines Experten schnell auffindig gemacht werden können. Alles in allem eignen sich die Arable-Stationen auf Basis dieser Untersuchung und weiteren für das Erfassen lokaler Klimabedingungen, sowohl in Regionen, die bereits von anderen Erfassungssystemen wie nationale Wetterdaten abgedeckt werden, eine solche Abdeckung nicht vorzuweisen haben oder im Rahmen von wissenschaftlichen Untersuchungen, die darauf abzielen, die lokalen Klima- und Wettergegebenheiten aufzuzeichnen (siehe: Abimbola et al., 2022).

Die letzte zu überprüfende Hypothese besagt: *wenn die Genauigkeit der Messdaten von mobilen Wetterstationen hoch ausfällt, dann können mit diesen mikroklimatischen Bedingungen beobachtet und analysiert werden.*

Die Bewertung dieser Hypothese ist stark von der angesetzten Definition von Mikroklima abhängig. Versteht man mikroklimatische Bedingungen als jene Klimaausprägungen, die lokal und in kleinräumigen Gebieten auftreten (etwa Hanglagen), die maximal bis wenige Meter über der Erdoberfläche auftreten und sich von ihrer Umgebung klimatologisch abheben, so kann diese Hypothese unter Vorbehalt bejaht werden (Wanjohi 2017). Der Vergleich zwischen den Stationen in Bentzin mit der DWD-Station in Greifswald zeigt auf, dass die Messwerte zwischen den beiden Ortschaften einander zwar ähneln, jedoch bezogen auf die Luftfeuchtigkeit und den Niederschlag in ihren Ausprägungen voneinander abweichen. Hingegen erzielen die ortsnah aufgestellten Arable-Stationen stark vergleichbare Ergebnisse mit nahezu perfekten linearen Zusammenhängen zwischen den Lufttemperaturen und den relativen Luftfeuchtigkeitswerten jeweils untereinander. Die bereits zitierte Arbeit von Abimbola et al., (2022, S. 14) arbeitet in diesem Zusammenhang heraus, dass die Verwendung von mit Arable-Mark-2 Sensoren für die Berechnung von Vorhersagemodellen aufgrund der genaueren Beschreibung lokaler Klimabedingungen eine Verbesserung solcher Modelle erzielen können und dieser Einschätzung schließt sich diese Arbeit an. Solange mindestens ein Meter zwischen den Arable-Mark-2 Sensoren und dem zu beobachteten Untersuchungsobjekt wie Nutzpflanzen vorhanden ist, sollten diese die lokalen Wetterausprägungen optimal messen können (Wright 2021, S. 3). Definiert man Mikroklima jedoch als jenen Bereich, der erdbodennah gelegen ist, könnten die Arable-Mark-2 Sensoren unzureichend sein. Nach Maclean, Duffy et al., (2021, S. 1398) ist die Erhebung von akkuraten, mikroklimatischen Daten allein für relevante Klimavariablen wie die Lufttemperatur unterhalb der 1,2 Metermarke stark von potenziellen Verzerrungsquellen betroffen. Die Gründe hierfür wurden bereits geschildert, doch um es kurz in Erinnerung zu rufen, stellt die Hauptproblematik die Abnahme der Luftzirkulation dar, die den Einsatz von abgeschirmten Thermometern erschwert, während gleichzeitig die Auswirkungen der reflektierten Strahlung der umliegenden Objekte zunimmt (Maclean, Duffy et al., 2021, S. 1398). Diese können die Messergebnisse eines nicht

abgeschirmten Thermometers verzerren (Maclean, Duffy et al., 2021, S. 1398).

Wie genau die Arable-Mark-2 Sensoren in Bodennähe arbeiten, wurde im Rahmen dieser Arbeit nicht getestet und könnte eine interessante Forschungsfrage für eine weiterführende Arbeit sein. Die von Arable Labs vorgesehene Nutzungsweise der Sensoren beschränkt sich derzeit auf die Anwendung innerhalb der Landwirtschaft, um zum Beispiel für die Einschätzung der Entwicklung von Nutzpflanzen relevante Klimavariablen generieren zu können (Habte et al., 2019, S. 1). Für diesen Einsatzbereich von Installationshöhen von über einen Meter bis zum Beispiel 4,9 m, wenn die zu untersuchende Vegetation hochgewachsen ist (Abimbola et al., 2022, S. 3), sind die Sensoren getestet und die AI-Algorithmen angepasst (siehe: Arable-Labs 2020a). So gibt der *ARABLE MARK 2 QUICK START GUIDE* an, dass die Sensoren mindestens einen Meter über die zu untersuchende Vegetation angebracht werden soll (Wright 2021, S. 3). Im Rahmen dieser Arbeit wird jedoch nicht die zweite Definition von Mikroklima angesetzt, sondern jene, die einen kleinräumigen Bereich von wenigen Metern über der Erdoberfläche abdeckt, der sich von seiner Umgebung klimatologisch abhebt. Innerhalb dieses Definitionsspielraumes scheinen die Arable-Mark-2 Sensoren durchaus in der Lage zu sein, die besonderen lokalen, also mikroklimatischen Bedingungen messen zu können.

#### 5.4 Literaturdiskussion: Relevanz von in situ Daten

Trotz der vorangegangenen Aussagen und Schlussfolgerungen wurde bisher eine zentrale Fragestellung, die den wissenschaftlichen Diskurs dominiert, nicht thematisiert: braucht es für die Untersuchung oder Anwendung von mikroklimatischen Klimadaten die Erhebung von in situ Daten oder genügt es, wenn diese modelliert werden?

Diese Frage wurde bereits in der Vorstellung des **aktuellen Forschungsstandes (2.3)** aufgeworfen und die drei hauptsächlichen Methoden, die angewendet werden, um im Forschungskontext mikroklimatische Daten zu generieren, vorgestellt. Zur Erinnerung: Diese drei Methodenpakete sind grob gefasst die Modellierung auf Basis von makro- beziehungsweise mesoklimatischen Datensätzen, der Einsatz von Fernerkundungsmethoden oder die Erhebung direkt vor Ort und im Feld. Mögliche Stärken und Schwächen der einzelnen Ansätze wurden bereits aufgezeigt. Unter anderem wurden benannt: der hohe Rechenaufwand für die Modellierung mikroklimatischer Bedingungen anhand von makroklimatischen Datensätzen und der aktuelle Trend innerhalb der Forschung mesoklimatische Daten mit in solche Modellierungen einzubeziehen, um akkurate Stundenwerte und Ähnliches generieren zu können bei gleichzeitiger Reduzierung des Rechenaufwandes (Lembrechts und Lenoir 2019, S. 338; Maclean, Mosedale et al., 2019); die Kombination verschiedener Fernerkundungsmethoden, um punktuelle Aussagen über Klimabedingungen zu erschließen, die über große Räume hinweg interpoliert werden können, und eine Vielzahl an Klimavariablen abdecken, wobei beson-

ders die räumliche (wie zeitliche) Auflösung für die Berechnung mikroklimatischer Bedingungen erschweren Bedingungen mit sich bringen (Maclean, Mosedale et al., 2019, S. 281; Zellweger et al., 2019, 328 ff.); und schlussendlich die Erhebung im Feld, die direkte Messungen mikroklimatischer Bedingungen erlauben, sich jedoch als ein recht komplexes Methodengebiet herausstellt, da die herkömmlichen Messmethoden in Bodennähe mit Verzerrungseffekten konfrontiert werden können, die in Höhen von 1,8 m und mehr in diesen Ausprägungen nicht auftreten (Maclean, Duffy et al., 2021, 1398 ff.).

Eine weitere Problematik der einzelnen Ansätze sind die zur Verfügung stehenden Daten selbst. Meso- und makroklimatische Datensätze von etwa nationalen Wetterstationen weisen zwar eine gute zeitliche Auflösung auf (decken Jahrzehnte und mehr ab), sind dafür in ihrer räumlichen Verteilung eingeschränkt und können für lange Zeitreihen Fehlwerte enthalten, mit denen umgegangen werden muss (Xu et al., 2013, 7452 f. Jeffrey et al., 2001, S. 312–321). Sie sind punktuell verteilt und Gebiete, wo keine Messstation steht, müssen auf Basis anderer Stationen interpoliert werden (Jeffrey et al., 2001, S. 309–310). Fernerkundungsdatensätze hingegen können großräumige, flächenhafte Informationen erfassen (die Sentinel-2 Satelliten besitzen eine Schwadbreite von 290 km) (ESA 2022), weisen dabei jedoch Einschränkungen in ihrer zeitlichen und räumlichen Auflösung auf (siehe: ESA 2022; EOS 2022). So erreichen die europäischen SENTINEL-2 Satelliten je nach Sensor zwischen räumliche Auflösungen von 10 bis 60 m Bildpixel (EOS 2022) und Umlaufbahnen, bis der gleiche Ort wiederholt überflogen wird, von wenigen Tagen (ESA 2022). Diese Problematik ist für in situ Daten nicht gegeben. Je nach betrachteter mikroklimatischer Dimension können direkt im Feld Sensoren verwendet werden, um vor Ort konstante Messungen durchzuführen (Lembrechts, Nijs et al., 2019, S. 1269). Jedoch sind diese Messungen in ihrer zeitlichen Auflösung beschränkt und sind oft nur für kurze Zeitperioden wie Jahreszeiten oder ausgewählte Jahre im Betrieb (Lembrechts, Nijs et al., 2019, S. 1269). Nach Lembrechts, Nijs et al., (2019, S. 1269) messen diese keine mikroklimatischen Bedingungen, als vielmehr die derzeitigen Ausprägungen des Mikrowetters. Neben der zeitlichen Auflösung sind in situ Datensätze stark in ihrer räumlichen Verteilung beschränkt. Anschaffungskosten und Wartung der Gerätschaften verhindern einen großräumigen Einsatz dieser (Lenoir et al., 2013, S. 1471).

Welche dieser unterschiedlichen Datensätze ist für die Anwendung und Analyse des Mikroklimas einzelner Regionen im wissenschaftlichen Kontext am besten geeignet?

Derzeitige Trends zielen auf die Verbesserung bestehender methodischen Ansätze ab. So werden mechanistische Modelle um mesoklimatische Bedingungen erweitert, um mikroklimatische Bedingungen exakter berechnen und um Stundenwerte generieren zu können (siehe: Maclean, Mosedale et al., 2019; Lembrechts und Lenoir 2019). Aufgrund ihrer jeweiligen Stärken und Schwächen

erscheinen Kombinationen dieser methodischen Vorgehensweisen als ratsam. So können in situ Daten mit Satellitendaten aus der Fernerkundung kombiniert werden, um mikroklimatische Bedingungen über den Raum zu interpolieren (Greiser et al., 2018, S. 153–156). Erweitert man diesen Ansatz durch die Hinzunahme von etwa nationalen Wetterstationen, lassen sich Datensätze erzeugen, die mikroklimatische Bedingungen aus den in situ Daten, mit der räumlichen Ausbreitung der fernerkundlichen Daten, mit der zeitlichen Auflösung von nationalen Wetterstationen kombiniert. Ein solcher Ansatz wird von Lembrechts, Nijs et al., (2019, 1274 ff.) für den Einsatz mikroklimatischer Daten für Artenverteilungsmodelle postuliert. Solche Modelle benötigen, um exakte Aussagen für Arten vor dem Hintergrund des globalen Klimawandels und dessen Auswirkungen treffen zu können, Informationen bezüglich des Mikroklima (Storlie et al., 2014, S. 3–4).

Innerhalb der Landwirtschaft zeichnen sich ähnliche Erkenntnisse und Trends ab. Vor dem Hintergrund der Gewinnmaximierung bei gleichzeitiger Kostenreduzierung eignen sich in situ Daten, wie sie mit den Arable-Mark 2 Sensoren aufgezeichnet werden können, um während der Wachstumsphasen der Pflanzen korrekte Entscheidungen, bezüglich deren Bedarf treffen zu können (Abimbola et al., 2022, S. 2). Nachhaltigkeitsinteressen und ressourcenschonende Anbautechniken stehen dabei im Vordergrund moderner Entwicklungen innerhalb der Landwirtschaft (BMEL 2022, S. 11). So werden schon jetzt fernerkundungsgestützte Verfahren verwendet, um Traktoren dank Satellitennavigation bis auf zwei Zentimeter genau zu steuern, wodurch Überlappung bei der Zugabe von Dünger und Pflanzenschutzmittel vermindert wird (BMEL 2022, S. 12). Drohnen werden verwendet, um Tiere oder Objekte wie Steine im Feld erfassen und entfernen zu können (BMEL 2022, S. 12). Zusätzlich dienen sie der Messung von relevanten Daten wie Bestandsdichte, Nährstoffversorgung oder Trockenstress oder für die gezielte Ausgabe von Schädlingsbekämpfungsmitteln (Lutz 2017, S. 438). Wetterdaten von nationalen Wetterstationen werden verwendet, um den Wasserbedarf einschätzen zu können (BMEL 2022, 15 f.). Direkt vor Ort erhobene Daten (in situ) bezüglich der Temperatur, des Niederschlages (Wasserbedarf), der relativen Luftfeuchtigkeit, der Bodenbeschaffenheit sollen diese Entscheidungen weiter verbessern können (Lutz 2017, S. 434).

Zusammengefasst lässt sich festhalten, dass die Bedeutung der Betrachtung von mikroklimatischen Prozessen und Bedingungen in den letzten Jahrzehnten zugenommen hat. Die Darstellung der Abbildung 1 stellte diese Zunahme in grafischer Form anhand von wissenschaftlichen Artikeln, die sich mit diesem Thema auseinandersetzen, grob dar, und wurde im Zuge der Darstellung des **Forschungsstandes (2.3)** und der darauf aufbauenden **Literaturdiskussion (5.4)** erweitert. Mikroklimatische Datensätze werden in den unterschiedlichsten Bereichen benötigt, sei es um die Auswirkungen des globalen Klimawandels auf Ökosysteme und Lebewesen einschätzen zu können (Artenverteilungsmodellen) oder um ressourcenschonende, effiziente Entscheidungen innerhalb der

Landwirtschaft treffen zu können (Lutz 2017, S. 434). Welche Daten und welche Methodenart verwendet wird, ist immer abhängig von der jeweiligen Forschungsfrage, den gegebenen Daten für jene Region und dem Forschungsziel. Keine der genannten Erhebungsarten kann als die Beste betitelt werden; vielmehr sind sie gleichwertige Werkzeuge, die je nach Forschungsobjekt in Tandem verwendet werden sollten (Lembrechts, Nijs et al., 2019, 1274 ff. WMO 2021, S. 18).

## 6 Fazit

Diese Masterarbeit, die den Titel „Anwendung von mobilen Wetterdaten zur Analyse und Einschätzung von Mikroklimaten“ trägt, verfolgt das Ziel, die bereitgestellten Wetterdaten, die mit den Arable-Mark-2 an verschiedenen Standorten für unterschiedlich lange Zeitabschnitte gesammelt wurden, auf ihre Qualität hin zu überprüfen. Während der Phase der Ideensammlung bezüglich der Umsetzung kristallisierten sich zwei grundlegende Vorgehensweisen heraus: zum einen sollte ein Skript geschrieben werden, das mithilfe von selbst geschriebenen Funktionen anhand von vorangegangenen wissenschaftlichen Arbeiten die Daten testet und zum anderen sollte ein Vergleich zwischen diesen und anderen Wetterdaten angestellt werden, um herauszufinden, wie ähnlich sich diese sind. Die Arbeiten an dem Skript führten dazu, dass dieses aufgeteilt wurde in ein R-Paket, das die einzelnen Funktionen gesammelt beinhaltet, und R-Skripte, die verwendet werden, um diese Funktionen auf den Datensatz anzuwenden. Trotz Problematiken wie eingeschränkten SQL-Abfragen und dem Handling eines Datensatzes, der die Daten mehrerer Stationen beinhaltet, welche methodenbedingt getrennt voneinander untersucht werden müssen, konnten zwei Skripte geschrieben werden, die zur Überprüfung des Datensatzes verwendet werden können. Das Erste ermöglicht es, den gesamten Datensatz zu überprüfen, wohingegen das Zweite genutzt werden kann, um lediglich die neuen, nicht überprüften Werte zu testen. Letzteres benötigte viele Korrekturarbeiten, um mit den Besonderheiten der Dateneingabe in die Datenbank umgehen zu können (wie zum Beispiel die bereits erwähnte Tatsache, dass teils weiter in der Vergangenheit liegende Werte eingetragen werden, obwohl neuere hinzugefügt und getestet wurden). Vor diesem Hintergrund ist es ein optimales Ergebnis, dass die Umsetzung der recherchierten Methoden nicht nur als Funktionen innerhalb eines R-Paketes möglich war, sondern auch, dass ein Skript geschrieben werden konnte, das trotz der aufgezeigten Besonderheiten in der Lage ist, lediglich die noch nicht getesteten Werte zu kontrollieren und zu flaggen. Die Tatsache, dass *R* eine Programmiersprache und Entwicklungsumgebung darstellt, die für den Einsatz von statistischen Verfahren und der Erstellung von statistischen Grafiken entwickelt wurde (Tucker et al., 2022, S. 2), ist dies ein sehr zufriedenstellendes Ergebnis. Die Erwähnung der eigentlichen Nutzungsweise *Rs* leitet in den zweiten zentralen Aspekt dieser Arbeit

ein: der Vergleichsanalyse.

Als Vergleichsstation wurde aufgrund der freien Zugänglichkeit und dem direkten Einbinden in *R* die Wetterstationsdaten des DWD ausgewählt. Auf Basis von Recherchen wurden DWD-Stationen ausgewählt, die räumlich betrachtet den zu testenden Arable-Stationen am nächsten sind. Die Arable-Station bei Erdeborn wurde dabei zufällig herausgewählt, wohingegen die vier Bentzin Stationen bewusst während des Schreibens dieser Arbeit Anfang Juli in die Analyse aufgenommen wurden, weil sie die perfekte Ergänzung zum Erdeborn-Querfurt Vergleich darstellen. Auf Basis dieser beiden Vergleiche lässt sich festhalten, dass die Arable Sensoren dazu in der Lage sind, die Lufttemperatur und die relative Luftfeuchtigkeit zu messen. Im Vergleich mit DWD-Stationen fallen die Messwerte stark vergleichbar aus, wohingegen die Stationen untereinander noch ähnlichere Ergebnisse erzielen (siehe: Tabelle 6 (S. 45), Tabelle 7 (S. 58), Tabelle 8 (S. 59) und Tabelle 9 (S. 61)). Dies deutet darauf hin, dass die Messungenauigkeiten von 0,8 °C im kalibrierten Bereich für die Lufttemperatur und 5 % für die relative Luftfeuchtigkeit nur marginale Auswirkungen auf die Messungen haben. Die Überprüfung des Niederschlages gestaltet sich als deutlich schwieriger aufgrund der hohen räumlichen Variabilität dessen (Jensen und Pedersen 2005, S. 269, 276–277; Pedersen et al., 2010, S. 449). Dennoch erscheinen die Messungen als plausibel und realitätsnah; besonders wenn dieser Fakt mit der potenziellen Messungenauigkeit von 6 % bedacht wird.

Kommen wir an dieser Stelle zurück zum Titel der Masterarbeit und ergänzen diesen durch die eingangs gestellte Forschungsfragen: *Eignen sich mobile Wetterstationen wie die Arable-Mark-2, um einen Mehrwert für die Betrachtung, Analyse und Einschätzung von Mikroklimaten zu erzeugen? Wie gut ist die Datenqualität solcher Messsensoren und wie stark weichen die erhobenen in Situ Daten von den makroklimatischen Wetterdaten nationaler Wetterstationen ab? Lässt sich anhand der dargestellten Ergebnisse und Zusammenhänge diese Frage bejahen und inwiefern?*

Der zweite Aspekt dieser Forschungsfragen wurde soeben erst beantwortet, jedoch: auf Basis der angestellten Vergleichsanalysen kann zumindest für die Messungen der Lufttemperatur als auch der relativen Luftfeuchtigkeit der Schluss gezogen werden, dass die Arable-Mark-2 Sensoren stark vergleichbare Messungen zu zum Beispiel makroklimatische Wetterdaten erzielen können. Dabei werden 95 % der Lufttemperatur- und der Niederschlags- und 90 % der relativen Luftfeuchtigkeitswerte als valide klassifiziert. Wenn Werte als fragwürdig deklariert wurden, dann häufig auf Basis der Grenzwerte der *zeitlichen Konsistenzprüfung*. Diese Werte werden innerhalb der Literatur für klassische Wetterstationen angegeben und nicht für Messsensoren, wie sie für die Arable-Stationen Anwendung finden (siehe: Cerlini et al., 2020, S. 6; Zahumenský 2002, S. 6). Wie jedoch dargestellt werden konnte, gewinnen in situ Daten erst langsam und seit einigen Jahren mehr und mehr an Bedeutungszuwachs, wenn es darum geht, sie in die Analyse von wissenschaftlichen Fragestellungen

einzubinden (Storlie et al., 2014, S. 1–3; Zellweger et al., 2019, S. 329; Kearney, Shamakhy et al., 2014, S. 273; Burnett und Anderson 2019, S. 1). Diese Grenzwerte könnten und sollten mit Veränderungen des derzeitigen Klimas und anderen Arbeiten im Laufe der Zeit Anpassungen erfahren (WMO 2021, S. 39).

Für die Beantwortung des ersten und zentralen Teils der Forschungsfragen muss auf die **Hypothesenüberprüfung (5.3)** und die **Literaturdiskussion (5.4)** verwiesen werden. Je nach Definition des Begriffes *Mikroklima* variiert die Beantwortung dieser Forschungsfrage. Versteht man Mikroklima wie jene Klimabedingungen, denen Kleinstiere und Pflanzen in Bodennähe ausgesetzt sind (Broll et al., 2017), so kann sie auf Basis der vorgestellten Ergebnisse und angestellten Untersuchungen nicht beantwortet werden. Die Arable Wetterstationen sind für den Einsatz der Beobachtung und Messung von Bedingungen, denen Nutzpflanzen innerhalb der Landwirtschaft ausgesetzt sind, gedacht, um zielgerichtete Entscheidungen der Landwirte zu ermöglichen (Arable-Labs 2020a). Dabei müssen die Sensoren mindestens einen Meter zum Untersuchungsobjekt aufweisen (Wright 2021, S. 3). Diese Arbeit überprüft jedoch nicht, wie korrekt die Messungen ausfallen, wenn dieser Abstand nicht eingehalten wird oder wie genau die Messungen sind, wenn statt Pflanzen der Boden betrachtet wird. Wenn das Mikroklima als ein Raum von wenigen Metern über den Boden, der besonders durch seine Verschiedenheit von seiner direkten Umgebung, als solches definiert wird (Vanwalleghem und Meentemeyer 2009, S. 1158; Gebhardt et al., 2007, S. 192; Broll et al., 2017), dann kann die Forschungsfrage auf Basis der ermittelten Ergebnisse bejaht werden (Vergleich: Abimbola et al., 2022, S. 6–7).

Darüber hinaus scheint die Güte der Messungen gut genug zu sein, um, wie nachgewiesen wurde, sehr vergleichbare Messungen wie etablierte Systeme wie national Wettersysteme des DWD zu erzielen (siehe: **4.2.1** und **4.2.2**). Dies führt zu der Schlussfolgerung, dass die Arable-Mark-2 Sensoren nicht nur in der Lage sind, lokale Gegebenheiten über mindestens 1 m zu erfassen (Wright 2021, S. 3), sondern auch darüber hinaus eingesetzt werden können, um an Orten, wo keine nationalen Wetterstationen oder vergleichbares vorzufinden sind, aufgestellt werden können. Dieser Aspekt der Güte der Messungen, gepaart mit der simplen Installation der Stationen im Feld (siehe: Wright 2021), macht diese Art der mobilen Wetterstation für die Erforschung mikroklimatischer Fragestellung relevant. Zum einen können Daten an Orte gesammelt werden, wo keine etablierten Systeme vorkommen, und zum anderen ermöglichen diese die vertiefende Betrachtung und Analyse von Forschungsvorhaben, wenn weitere Datenquellen gegeben sind (Lembrechts, Nijs et al., 2019, S. 1274).

Wenn es das Ziel einer Untersuchung sein soll, lokale Bedingungen zu messen, wo keine anderen etablierten Systeme vorkommen, die zum Vergleich herangezogen werden könnten, wäre es sinn-

voll, die in dieser Arbeit vorgestellte Methodik zu erweitern. Nach der automatisierten Datenqualitätsüberprüfung sollen Messwerte immer der Betrachtung durch einen Experten unterzogen werden (WMO 2021, S. 5). Wenn dies nicht möglich ist, dann sollten zumindest die fragwürdigen Werte kritisch hinterfragt werden (WMO 2021, S. 5). Diese Überprüfung der Datensätze kann selbst anhand von Vergleichsanalysen, wie die anhand der Stationen des DWD durchgeführten sehr schnell sehr zeitaufwendig werden. So konnten im Rahmen dieser Arbeit lediglich 5 der 48 enthaltenden Arable-Stationen im Detail betrachtet werden. Ferner liegen solche Vergleichsdatsätze nicht für jeden potenziellen Einsatzort für mobile Wetterstationen vor. Darauf aufbauend kann der weiterführende Forschungsansatz erdacht werden, Methoden der Fernerkundung hinzuzunehmen, um die lokal durchgeführten Messungen zu überprüfen (Lembrechts, Nijs et al., 2019, S. 1274). Damit nicht genug wäre es sinnvoll, wenn eine solche Überprüfung anhand von Fernerkundungsdaten zumindest teils automatisch durchgeführt werden könnte, um den Arbeitsaufwand der manuellen Kontrolle zu minimieren. Nach der WMO (2021, S. 18) könnten auf diese Weise Klimavariablen wie die tägliche Durchschnittstemperatur, die tägliche Niederschlagssumme, die Sonneneinstrahlung, die Wolkenbedeckung und die Schneedecke validiert werden (WMO 2021, S. 18–19).

Ein weiterer aufbauender Forschungsansatz könnte experimenteller Natur sein und versuchen, zu untersuchen, was die Grenzen der Arable-Mark-2 Sensoren sind. Es wurde bereits geschlossen, dass sich diese Sensoren sehr wahrscheinlich nicht dazu eignen, um mikroklimatische Bedingungen nahe dem Boden zu messen. So schreibt Arable Labs selbst, dass die Sensoren mindestens 1 m über Pflanzenwuchs aufgestellt werden sollen und für den Einsatz innerhalb der Landwirtschaft ausgelegt sind (Wright 2021, S. 3; Habte et al., 2019, S. 1; Arable-Labs 2020a). *Unter welchen Bedingungen liefern mobile Wetterstationen wie Arable-Mark-2 die genauesten Ergebnisse? Wie sehr lässt sich ins Detail bei mikroklimatischen Bedingungen gehen? Und braucht es das oder reicht die Güte der Sensoren für mikroklimatische Fragestellungen aus?* - wären Fragen, von denen man sich leiten lassen könnte.

Neben diesen Ansätzen lassen sich noch weitere herausarbeiten, wie zum Beispiel eine detaillierte Vergleichsanalyse von mikroklimatischen Datensätzen, die auf verschiedene Weise generiert wurden (siehe: **2.3.3 Forschungsansätze zur Erforschung von Mikroklimaten**, doch mit den genannten wird es an dieser Stelle belassen. Die zukünftigen Einsatzfelder von Mikroklimadaten und den unterschiedlichen Erhebungsweisen wie durch mobile Wetterstationen sind mannigfaltig und werden im Zuge der Digitalisierung der Landwirtschaft und weiterer Bereiche zunehmen (BMEL 2022, S. 11). Diese Masterarbeit lieferte auf Basis der Arable-Mark-2 Wetterdaten einen relevanten Schritt in diese Richtung. Mobile Wetterstationsdaten sind in der Lage, ein besseres Verständnis über lokale Klimabedingungen zu ermöglichen. Ihr Einsatz kann bestehende Datensätze und Erkenntnisse erweitern (Lembrechts, Nijs et al., 2019, S. 1274) oder neue für Orte generieren, wo diese nicht

vorliegen. Die Stärke dieser Klimadaten und mobiler Wetterstationen als solche wird jedoch in dem gemeinsamen Einsatz mit weiteren Datenquellen wie jene aus der Fernerkundung oder ähnlichem gesehen. Eine Kombination der derzeit verfügbaren Methoden (Modellierung aus Makroklimadaten, Fernerkundung und in situ Messungen) kann detailreiche Klimaforschungen ermöglichen, die eine hohe zeitliche und flächenhafte Ausbreitung aufweisen, die dank lokaler Messungen stärker und genauer mikroklimatischer Bedingungen berücksichtigen kann. Solche Klimaforschungen und -ansätze sind vonnöten, wenn den direkten wie indirekten Folgen des anthropogenen, globalen Klimawandels nachhaltig begegnet werden soll und hier kommt die Betrachtung, Erhebung und Analyse des Mikroklimas ins Spiel, denn es sind diese kleinräumigen, lokalen Klima-, Witterungs- und Wetterausprägungen, die Pflanzen- und Tierarten und damit auch den Menschen direkt beeinflussen.

## Literatur

- Abimbola, Olufemi P., Trenton E. Franz, Daran Rudnick, Derek Heeren, Haishun Yang, Adam Wolf, Abia Katimbo, Hope N. Nakabuye und Anthony Amori (2022). "Improving crop modeling to better simulate maize yield variability under different irrigation managements". In: *Agricultural Water Management* 262, S. 107429. DOI: <https://doi.org/10.1016/j.agwat.2021.107429>.
- Arable-Labs (2020a). *Arable Mark 2 Core Measurements – An Accuracy Comparison*. URL: <https://www.arable.com/blog/how-arable-achieves-world-class-accuracy/> (besucht am 07.12.2021).
- (2020b). *Arable Mark 2 Product Specifications*. URL: [https://www.arable.com/wp-content/uploads/2021/10/Arable-Mark-2-w\\_-Solar-Product-Specifications-20\\_10.pdf](https://www.arable.com/wp-content/uploads/2021/10/Arable-Mark-2-w_-Solar-Product-Specifications-20_10.pdf) (besucht am 07.12.2021).
- (2021). *Arable Mark 2 Measurements*. URL: [https://learn.arable.com/hubfs/Arable-Mark-2-Measurements-21\\_05.pdf?hsCtaTracking=6cff6856-fa77-46f7-b2fc-15f870f2e9b5%7C0fd950ea-4651-4fe3-9bc8-889b6e585c41](https://learn.arable.com/hubfs/Arable-Mark-2-Measurements-21_05.pdf?hsCtaTracking=6cff6856-fa77-46f7-b2fc-15f870f2e9b5%7C0fd950ea-4651-4fe3-9bc8-889b6e585c41) (besucht am 06.03.2022).
- (2022a). *About Arable*. URL: <https://www.arable.com/company/> (besucht am 05.03.2022).
- (2022b). *The Arable Mark 2*. URL: <https://www.arable.com/products/sensing/> (besucht am 05.03.2022).
- Bahrenberg, G., E. Giese, N. Mevenkamp und J. Nipper (2017). *Statistische Methoden in der Geographie - Band 1: Univariate und bivariate Statistik*. Stuttgart: Borntraeger.
- Bellard, C., C. Bertelsmeier, P. Leadley, W. Thuiller und F. Courchamp (2012). "Impacts of climate change on the future of biodiversity". In: *Ecology Letters* 15.4, S. 365–377. DOI: <https://doi.org/10.1111/j.1461-0248.2011.01736.x>.
- Best, M. J., M. Pryor, D. B. Clark, G. G. Rooney, R. L. H. Essery, C. B. Ménard, J. M. Edwards, M. A. Hendry, A. Porson, N. Gedney, L. M. Mercado, S. Sitch, E. Blyth, O. Boucher, P. M. Cox, C. S. B. Grimmond und R. J. Harding (2011). "The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes". In: *Geoscientific Model Development* 4, S. 677–699. DOI: <https://doi.org/10.5194/gmd-4-677-2011>.
- BMEL (2022). *Digitalisierung in der Landwirtschaft*. URL: <https://www.bmel.de/SharedDocs/Downloads/DE/Broschueren/digitalpolitik-landwirtschaft.html> (besucht am 20.07.2022).
- Bowgen, K.M., E.F. Kettel, S.H.M. Butchart, J.A. Carr, W.B. Foden, G. Magin, M.D. Morecroft, R.K. Smith, B.A. Stein, W.J. Sutherland, C.B. Thaxter und J.W. Pearce-Higgins (2022). "Conservation interventions can benefit species impacted by climate change". In: *Biological Conservation* 269, S. 1–11. DOI: <https://doi.org/10.1016/j.biocon.2022.109524>.

- Broll, G., H. Egner, H. Leser, E. Rothfuß und M. Vetter (2017). *Diercke Wörterbuch Geographie*. Braunschweig: Bildungshaus Schulbuchverlage.
- BRRY (2022). *rdwd*. URL: <https://github.com/brry/rdwd> (besucht am 06.06.2022).
- Burnett, J. D. und P. D. Anderson (2019). "Using generalized additive models for interpolating microclimate in dry-site ponderosa pine forests". In: *Agricultural and Forest Meteorology* 279, S. 1–11. DOI: <https://doi.org/10.1016/j.agrformet.2019.107668>.
- Cai, L. und Y. Zhu (2015). "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era". In: *Data Science Journal* 14.2, S. 1–10. DOI: <http://doi.org/10.5334/dsj-2015-002>.
- Cerlini, P. B., L. Silvestri und M. Saraceni (2020). "Quality control and gap-filling methods applied to hourly temperature observations over central Italy". In: *Meteorological Applications* 27.3, S. 1–15. DOI: <https://doi.org/10.1002/met.1913>.
- Clarivate (2022). *Web of Science*. URL: <https://www.webofscience.com/> (besucht am 01.06.2022).
- DWD (2022a). *Warnkriterien*. URL: [https://www.dwd.de/DE/wetter/warnungen\\_aktuell/kriterien/warnkriterien.html](https://www.dwd.de/DE/wetter/warnungen_aktuell/kriterien/warnkriterien.html) (besucht am 07.06.2022).
- (2022b). *Wetter- und Klimalexikon: Klima*. URL: <https://www.dwd.de/DE/service/lexikon/Functions/glossar.html?lv2=101334&lv3=101462> (besucht am 10.05.2022).
- (2022c). *Wetter- und Klimalexikon: Platzregen*. URL: <https://www.dwd.de/DE/service/lexikon/Functions/glossar.html?lv3=102076&lv2=101996> (besucht am 07.06.2022).
- (2022d). *Wetter- und Klimalexikon: Witterung*. URL: <https://www.dwd.de/DE/service/lexikon/Functions/glossar.html?lv2=102936&lv3=103212> (besucht am 10.05.2022).
- EOS (2022). *Sentinel-2*. URL: <https://eos.com/find-satellite/sentinel-2> (besucht am 18.06.2022).
- ESA (2022). *Sentinel-2*. URL: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2> (besucht am 29.07.2022).
- Falkner, R. (2016). "The Paris Agreement and the new logic of international climate politics". In: *International Affairs* 92.5, S. 1107–1125.
- Fanelli, C. (2014). "Climate Change: 'The Greatest Challenge of Our Time'". In: *Alternate Routes: A Journal of Critical Social Research* 25, S. 15–32.
- Fick, S. E. und R. J. Hijmans (2017). "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas". In: *International Journal of Climatology* 37.12, S. 4302–4315. DOI: <https://doi.org/10.1002/joc.5086>.
- Field-Studies-Council (2022). *Microclimate*. URL: <https://www.field-studies-council.org/resources/16-18-biology/fieldwork-techniques/abiotic-factors/microclimate/> (besucht am 20.07.2022).

- Fleming, J. R. (1999). "Joseph Fourier, the 'greenhouse effect', and the quest for a universal theory of terrestrial temperatures". In: *Endeavour* 23.2, S. 72–75.
- Fodor, N. und G. J. Kovács (2005). "Sensitivity of crop models to the inaccuracy of meteorological observations". In: *Physics and Chemistry of the Earth, Parts A/B/C* 30.1, S. 53–57. DOI: <https://doi.org/10.1016/j.pce.2004.08.020>.
- Frenne, P. De, J. Lenoir, M. Luoto, B. R. Scheffers, F. Zellweger, J. Aalto, M. B. Ashcroft, D. M. Christiansen, G. Decocq, K. De Pauw, S. Govaert, C. Greiser, E. Gril, A. Hampe, T. Jucker, D. H. Klinges, I. A. Koelemeijer, J. J. Lembrechts, R. Marrec, C. Meeussen, J. Ogée, V. Tyystjärvi, P. Vangansbeke und K. Hylander (2021). "Forest microclimates and climate change: Importance, drivers and future research agenda". In: *Global Change Biology* 27.11, S. 2279–2297. DOI: <https://doi.org/10.1111/gcb.15569>.
- Freund, R. J. und W. J. Wilson P. Sa (2006). *Regression analysis: statistical modeling of a response variable*. London, Oxford, Boston, New York und San Diego: Academic Press.
- Galarnyk, M. (2018). *Understanding Boxplots*. URL: <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51> (besucht am 03.06.2022).
- Gardner, A.S., I.M.D. Maclean, K.J. Gaston und L. Bütikofer (2021). "Forecasting future crop suitability with microclimate data". In: *Agricultural Systems* 190, S. 1–9. DOI: <https://doi.org/10.1016/j.agsy.2021.103084>.
- Gebhardt, H., R. Glaser, U. Radtke und P. Reuber (2007). *Geographie - Physische Geographie und Humangeographie*. München: Spektrum Akademischer Verlag.
- Geiger, R. (1950). *The Climate Near The Ground*. Cambridge: Harvard University Press.
- Ghent-University (2022). *Forest & Nature Lab (ForNaLab)*. URL: <https://www.ugent.be/bw/environment/en/research/fornalab> (besucht am 18.06.2022).
- Gillingham, P. K., B. Huntley, W. E. Kunin und C. D. Thomas (2012). "The effect of spatial resolution on projected responses to climate warming". In: *Diversity and Distributions* 18.10, S. 990–1000. DOI: <https://doi.org/10.1111/j.1472-4642.2012.00933.x>.
- Gillis, J. (2013). *U.N. Climate Panel Endorses Ceiling on Global Emissions*. URL: <https://www.nytimes.com/2013/09/28/science/global-climate-change-report.html> (besucht am 02.05.2022).
- Gray, P. C., J. T. Ridge, S. K. Poulin, A. C. Seymour, A. M. Schwantes, J. J. Swenson und D. W. Johnston (2018). "Integrating Drone Imagery into High Resolution Satellite Remote Sensing Assessments of Estuarine Environments". In: *Remote Sensing* 10.8, S. 1–24. DOI: <https://doi.org/10.3390/rs10081257>.

- Greiser, Caroline, Eric Meineri, Miska Luoto, Johan Ehrlén und Kristoffer Hylander (2018). "Monthly microclimate models in a managed boreal forest landscape". In: *Agricultural and Forest Meteorology* 250-251, S. 147–158. DOI: <https://doi.org/10.1016/j.agrformet.2017.12.252>.
- Habte, A., A. Andreas, M. Sengupta, R. Narasappa, A. Hoke, P. Gotseff, R. Thiagarajan, A. Wolf, L. Carranza und D. Wattsirst (2019). "Low-Cost Multiparameter Device for Solar Resource Applications". In.
- Hendl, M. und H. Liedtke (1997). *Lehrbuch der Allgemeinen Physischen Geographie*. Gotha: Justus Perthes Verlag GmbH.
- IPCC (1992). *Climate Change: The 1990 and 1992 IPCC Assessments*. Genf: IPCC.
- (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]* Genf: IPCC.
- Jeffrey, S. J., J. O. Carter, K. B. Moodie und A. R. Beswick (2001). "Using spatial interpolation to construct a comprehensive archive of Australian climate data". In: *Environmental Modelling & Software* 16, S. 309–330.
- Jensen, N. E. und L. Pedersen (2005). "Spatial variability of rainfall: Variations within a single radar pixel". In: *Atmospheric Research* 77, S. 269–277. DOI: <https://doi.org/10.1016/j.atmosres.2004.10.029>.
- Johannsen, C. J. und C. S. Daughtry (2009). "Surface reference data collection". In: *The SAGE handbook of remote sensing (Hrsg.: Warner, T. A. Nelis, M. D. und G. M. Foody)*, S. 244–256.
- Kaspar, F., G. Müller-Westermeier, E. Penda, H. Mächel, K. Zimmermann, A. Kaiser-Weiss und T. Deuschländer (2013). "Monitoring of climate change in Germany - data, products and services of Germany's National Climate Data Centre". In: *Advances in Science & Research* 10, S. 99–106. DOI: <https://doi.org/10.5194/asr-10-99-2013>.
- Kearney, M. R. und W. P. Porter (2017). "NicheMapR – an R package for biophysical modelling: the microclimate model". In: *Ecography* 40, S. 664–674. DOI: <https://doi.org/10.1111/ecog.02360>.
- Kearney, M. R., A. Shamakhy, R. Tingley, D. J. Karoly, A. A. Hoffmann, P. R. Briggs und W. P. Porter (2014). "Microclimate modelling at macro scales: a test of a general microclimate model integrated with gridded continental-scale soil and weather data". In: *Methods in Ecology and Evolution* 5, S. 273–286. DOI: <https://doi.org/10.1111/2041-210X.12148>.
- Kollas, C., C. F. Randin, Y. Vitasse und C. Körner (2014). "How accurately can minimum temperatures at the cold limits of tree species be extrapolated from weather station data?" In: *Agricultural and Forest Meteorology* 14, S. 257–266. DOI: <https://doi.org/10.1016/j.agrformet.2013.10.001>.

- Lathiya, K. (2021). *R-Lang*. URL: <https://r-lang.com/setwd-function-in-r-with-example/> (besucht am 06.06.2022).
- Lefsky, M. A., W. B. Cohen, G. G. Parker und D. J. Harding (2002). "Lidar Remote Sensing for Ecosystem Studies". In: *BioScience* 52.1, S. 19–30.
- Lembrechts, J. J. und J. Lenoir (2019). "Microclimatic conditions anywhere at any time!" In: *Global Change Biology* 26, S. 337–339. DOI: <https://doi.org/10.1111/gcb.14942>.
- Lembrechts, J. J., I. Nijs und J. Lenoir (2019). "Incorporating microclimate into species distribution models". In: *Ecography* 42, S. 1267–1279. DOI: <https://doi.org/10.1111/ecog.03947>.
- Lenoir, J., B. J. Graae, P. A. Aarrestad, I. G. Alsos, W. S. Armbuster, G. Austrheim, C. Bergendorff, H. J. B. Birks, K. A. Bråthen, J. Brunet, H. H. Bruun, C. J. Dahlberg, G. Decocq, M. Diekmann, M. Dynesius, R. Ejrnæs, J.-A. Grytnes, K. Hylander, K. Klanderud, M. Luoto, A. Milbau, M. Moora, B. Nygaard, A. Odland, V. T. Ravolainen, S. Reinhardt, S. M. Sandvik, F. H. Schei, J. D. M. Speed, L. U. Tveraabak, V. Vandvik, L. G. Velle, R. Virtanen, M. Zobel und J.-C. Svenning (2013). "Local temperatures inferred from plant communities suggest strong spatial buffering of climate warming across Northern Europe". In: *Global Change Biology* 19.5, S. 1470–1481. DOI: <https://doi.org/10.1111/gcb.12129>.
- Liang, X., V. Kankare, J. Hyyppä, Y. Wang, A. Kukko, H. Haggrén, X. Yu, H. Kaartinen, A. Jaakkola, F. Guan, M. Holopainen und M. Vastaranta (2016). "Terrestrial laser scanning in forest inventories". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 115, S. 63–77. DOI: <https://doi.org/10.1016/j.isprsjprs.2016.01.006>.
- Lutz, K. J. (2017). "Digitalisierung der Landwirtschaft: Revolution mit evolutionärem Charakter". In: *CSR und Digitalisierung (Hrsg.: Hildebrandt, A. und W. Landhäußer)*, S. 429–442. DOI: <https://doi.org/10.1007/978-3-662-53202-7>.
- Maclean, I. M., J. P. Duffy, S. Haesen, S. Govaert, P. De Frenne, T. Vanneste, J. Lenoir, J. J. Lembrechts, M. W. Rhodes und K. Van Meerbeek (2021). "On the measurement of microclimate". In: *Methods in Ecology and Evolution* 12, S. 1397–1410. DOI: <https://doi.org/10.1111/2041-210X.13627>.
- Maclean, I. M., J. R. Mosedale und J. J. Bennie (2019). "Microclima: An r package for modelling meso- and microclimate". In: *Methods in Ecology and Evolution* 10, S. 280–290. DOI: <https://doi.org/10.1111/2041-210X.13093>.
- Muldowney, J., J. Mounsey und L. Kinsella (2013). "Agriculture in the climate change negotiations; ensuring that food production is not threatened". In: *Animal* 7, S. 206–211. DOI: <https://doi.org/10.1017/S175173111300089X>.
- Nakagawa, S., P. C. D. Johnson und H. Schielzeth (2017). "The coefficient of determination  $R^2$  and intra-class correlation coefficient from generalized linear mixed-effects models revisited and

- expanded". In: *Journal of the Royal Society Interface* 14, S. 1–11. DOI: <http://dx.doi.org/10.1098/rsif.2017.0213>.
- Nattke, M. (2021). "Effizientere Flächennutzung in der Landwirtschaft mit vernetzten Drohnen". In: *INFORMATIK 2021*, S. 389–400. DOI: <https://doi.org/10.18420/informatik2021-032>.
- Pedersen, L., N. E. Jensen, L. E. Christensen und H. Madsen (2010). "Quantification of the spatial variability of rainfall based on a dense network of rain gauges". In: *Atmospheric Research* 95.4, S. 441–454. DOI: <https://doi.org/10.1016/j.atmosres.2009.11.007>.
- PRODATA (2022). *Comparing observed readings with other stations*. URL: [https://www.weatherstations.co.uk/comparing\\_data.htm](https://www.weatherstations.co.uk/comparing_data.htm) (besucht am 07.06.2022).
- R-Project (2022). *Getting Help with R*. URL: <https://www.r-project.org/help.html#:~:text=The%20help%20%28%29%20function%20and%20%3F%20help%20operator,the%20standard%20R%20distribution%20and%20for%20contributed%20packages.> (besucht am 01.08.2022).
- Rafferty, J. P. (2022). *Microclimate*. URL: <https://www.britannica.com/science/microclimate> (besucht am 10.05.2022).
- Ray, C., E. A. Beever und T. J. Rodhouse (2016). "Distribution of a climate-sensitive species at an interior range margin". In: *Ecosphere* 7.6, S. 1–22.
- RDocumentation (2022a). *cor: Correlation, Variance and Covariance (Matrices)*. URL: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cor> (besucht am 03.06.2022).
- (2022b). *dim: Dimensions of an Object*. URL: <https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/dim> (besucht am 20.05.2022).
- (2022c). *plot: Generic X-Y Plotting*. URL: <https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/plot> (besucht am 03.06.2022).
- (2022d). *summary.lm: Summarizing Linear Model Fits*. URL: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/summary.lm> (besucht am 03.06.2022).
- Rebaudo, F., E. Faye und O. Dangles (2016). "Microclimate Data Improve Predictions of Insect Abundance Models Based on Calibrated Spatiotemporal Temperatures". In: *Frontiers in Physiology* 7, S. 1–12. DOI: <https://doi.org/10.3389/fphys.2016.00139>.
- Reckien, D., M. Salvia, O. Heidrich, J. M. Church, F. Pietrapertosa, S. De Gregorio-Hurtado, V. D'Alonzo, A. Foley, S. G. Simoes, E. Krkoška Lorencová, H. Orru, K. Orru, A. Wejs, J. Flacke, M. Olazabal, D. Geneletti, E. Feliu, S. Vasilie, C. Nador, A. Krook-Riekkola, M. Matosović, P. A. Fokaides, B. I. Ioannou, A. Flamos, N.-A. Spyridaki, M. V. Balzan, O. Fülöp, I. Paspaldzhiev, S. Grafakos und R. Dawson (2018). "How are cities planning to respond to climate change? Assessment of local climate plans from 885 cities in the EU-28". In: *Journal of Cleaner Production* 191, S. 207–219. DOI: <https://doi.org/10.1016/j.jclepro.2018.03.220>.

- RStudio (2022a). *Building, Testing, and Distributing Packages*. URL: <https://support.rstudio.com/hc/en-us/articles/200486508-Building-Testing-and-Distributing-Packages> (besucht am 30.04.2022).
- (2022b). *PostgreSQL*. URL: <https://db.rstudio.com/databases/postgresql/#using-the-rpostgres-package> (besucht am 20.05.2022).
- Scherrer, D. und C. Körner (2010). "Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming". In: *Journal of Biogeography* 38.2, S. 406–416. DOI: <https://doi.org/10.1111/j.1365-2699.2010.02407.x>.
- Schmidt, A., A. Ivanova und M. S. Schäfer (2013). "Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries". In: *Global Environmental Change* 23, S. 1233–1248. DOI: <https://doi.org/10.1016/j.gloenvcha.2013.07.020>.
- Schneider, S. H. (1989). "The Greenhouse Effect: Science and Policy". In: *Science* 243.(4892), S. 771–781. DOI: <https://doi.org/10.1126/science.243.4892.771>.
- Semprebello, A., S. Magazù und M. T. Caccamo (2021). "JOSEPH FOURIER: A DIDACTIC PATH CROSSING LIFE AND WORKS". In: *ATTI DELLA ACCADEMIA PELORITANA DEI PERICOLANTI - CLASSE DI SCIENZE FISICHE, MATEMATICHE E NATURALI* 99.S1, A8 1–28. DOI: <https://doi.org/10.1478/AAPP.99S1A8>.
- Slavich, E., D. I. Warton, M. B. Ashcroft, J. R. Gollan und D. Ramp (2014). "Topoclimate versus macroclimate: how does climate mapping methodology affect species distribution models and climate change projections?" In: *Diversity and Distributions* 20.8, S. 952–963. DOI: <https://doi.org/10.1111/ddi.12216>.
- Spengler, R. (2002). "The new Quality Control- and Monitoring System of the Deutscher Wetterdienst". In: *INSTRUMENTS AND OBSERVING METHODS - REPORT No. 75 WMO/TD-No. 1123*, S. 245–248.
- Stephenson, J., K. Newman und S. Mayhew (2010). "Population dynamics and climate change: what are the links?" In: *Journal of Public Health* 32.2, S. 150–156. DOI: <https://doi.org/10.1093/pubmed/fdq038>.
- STHDA (2022). *ggplot2 histogram plot : Quick start guide - R software and data visualization*. URL: <http://www.sthda.com/english/wiki/ggplot2-histogram-plot-quick-start-guide-r-software-and-data-visualization> (besucht am 29.05.2022).
- Storlie, C., A. Merino-Viteri, B. Phillips, J. VanDerWal, J. Welbergen und S. Williams (2014). "Stepping inside the niche: microclimate data are critical for accurate assessment of species' vulnerability to climate change". In: *Biology Letters* 10, 20140576 (1–4). DOI: <https://doi.org/10.1098/rsbl.2014.0576>.

- Stoutjesdijk, Ph. und J. J. Barkman (2014). *Microclimate, vegetation and fauna*. Zeist: KNNV Publishing.
- Strahler, A. H. und A. N. Strahler (2005). *Physische Geographie*. Stuttgart: Verlag Eugen Ulmer KG.
- Suggitt, A. J., P. K. Gillingham, J. K. Hill, B. Huntley, W. E. Kunin, D. B. Roy und C. D. Thomas (2011). "Habitat microclimates drive fine-scale variation in extreme temperatures". In: *Oikos* 120, S. 1–8. DOI: <https://doi.org/10.1111/j.1600-0706.2010.18270.x>.
- Sykes, A. O. (1993). "An Introduction to Regression Analysis". In: *Coase-Sandor Institute for Law & Economics Working Paper No. 20*, S. 1–33.
- Taylor, R. (1990). "Interpretation of the Correlation". In: *Journal of Diagnostic Medical Sonography* 6, S. 35–39.
- Timeanddate (2022). *Wetter-Rückblick Demmin, Mecklenburg-Vorpommern, Deutschland – gestern, letzte Wochen & Monate*. URL: <https://www.timeanddate.de/wetter/deutschland/demmin/rueckblick?month=5&year=2022> (besucht am 21.06.2022).
- Tucker, M. C., S. T. Shaw, J. Y. Son und J. W. Stigler (2022). "Teaching Statistics and Data Analysis with R". In: *Journal of Statistics and Data Science Education* 00, S. 1–15. DOI: <https://doi.org/10.1080/26939169.2022.2089410>.
- UNFCCC (2015). *Paris Agreement*. URL: [https://unfccc.int/sites/default/files/english\\_paris\\_agreement.pdf](https://unfccc.int/sites/default/files/english_paris_agreement.pdf) (besucht am 03.05.2022).
- (2022). *The Paris Agreement*. URL: <https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement> (besucht am 03.05.2022).
- Vanwalleghem, T. und R. K. Meentemeyer (2009). "Predicting Forest Microclimate in Heterogeneous Landscapes". In: *Ecosystems* 12, S. 1158–1172. DOI: <https://doi.org/10.1007/s10021-009-9281-1>.
- Wang, X.-j., J.-y. Zhang, S. Shahid, E.-h. Guan, Y.-x. Wu, J. Gao und R.-m. He (2016). "Adaptation to climate change impacts on water demand". In: *Mitigation and Adaptation Strategies for Global Change volume 21*, S. 81–99. DOI: <https://doi.org/10.1007/s11027-014-9571-6>.
- Wanjohi, W. (2017). *What Is A Microclimate?* URL: <https://www.worldatlas.com/articles/what-is-a-microclimate.html> (besucht am 10.05.2022).
- Wasof, S., J. Lenoir, P. A. Aarrestad, I. G. Alsos, W. S. Armbruster, G. Austrheim, V. Bakkestuen, H. J. B. Birks, K. A. Bråthen, O. Broennimann, J. Brunet, H. H. Bruun, C. J. Dahlberg, M. Diekmann, S. Dullinger, M. Dynesius, R. Ejrnæs, J.-C. Gégout, B. J. Graae, J.-A. Grytnes, A. Guisan, K. Hylander, I. S. Jónsdóttir, J. Kapfer, K. Klanderud, M. Luoto, A. Milbau, M. Moora, B. Nygaard, A. Odland, H. Pauli, V. Ravolainen, S. Reinhardt, S. M. Sandvik, F. H. Schei, J. D. M. Speed, J.-C. Svenning, W. Thuiller, L. U. Tveraabak, V. Vandvik, L. G. Velle, R. Virtanen, P. Vittoz, W. Willner, T. Wohlgemuth, N. E. Zimmermann, M. Zobel und G. Decocq (2015). "Disjunct populati-

- ons of European vascular plant species keep the same climatic niches". In: *Global Ecology and Biogeography* 24.12, S. 1401–1412. DOI: <https://doi.org/10.1111/geb.12375>.
- Al-Wassai, F. A. und N. V. Kalyankar (2013). "MAJOR LIMITATIONS OF SATELLITE IMAGES". In: *Journal of Global Research in Computer Science* 4.5, S. 51–59.
- Williamson, D. F., R. A. Parker und J. S. Kendrick (1989). "The Box Plot: A Simple Visual Method to Interpret Data". In: *Annals of Internal Medicine* 110.11, S. 916–921.
- WMO (1993). *Guide on the Global Data-Processing System*. Bd. WMO-No. 305. Genf: World Meteorological Organization.
- (2018). *Guide to Instruments and Methods of Observation: Volume I –Measurement of Meteorological Variables*. Bd. WMO-No. 8. Genf: World Meteorological Organization.
- (2019). *Manual on the WMO Integrated Global Observing System*. Bd. WMO-No. 1160. Genf: World Meteorological Organization.
- (2021). *Guidelines on Surface Station Data Quality Control and Quality Assurance for Climate Applications*. Bd. WMO-No. 1269. Genf: World Meteorological Organization.
- Wright, Eliza (2021). *Mark 2 Quickstart Guide*. URL: <https://support.arable.com/hc/en-us/articles/1500002735642-Mark-2-Quickstart-Guide> (besucht am 02.07.2022).
- Xu, C.-D., J.-F. Wang, M.-G. Hu und Q.-X. Li (2013). "Interpolation of Missing Temperature Data at Meteorological Stations Using P-BSHADE". In: *Journal of Climate* 26, S. 7452–7463. DOI: <https://doi.org/10.1175/JCLI-D-12-00633.1>.
- Xuejie, G., Z. Zongci, D. Yihui, H. Ronghui und F. Giorgi (2001). "Climate Change due to Greenhouse Effects in China as Simulated by a Regional Climate Model". In: *Advances in Atmospheric Sciences* 18.6, S. 1224–1230.
- Yang, Y., E. Gatto, Z. Gao, R. Buccolieri, T. E. Morakinyo und H. Lan (2019). "The "plant evaluation model" for the assessment of the impact of vegetation on outdoor microclimate in the urban environment". In: *Building and Environment* 159, S. 1–13. DOI: <https://doi.org/10.1016/j.buildenv.2019.05.029>.
- Zahumenský, I. (2002). *Guidelines on Quality Control Procedures for Data from Automatic Weather Stations*. Genf: World Meteorological Organization.
- Zellweger, F., P. De Frenne, J. Lenoir, D. Rocchini und D. Coomes (2019). "Advances in Microclimate Ecology Arising from Remote Sensing". In: *Trends in Ecology & Evolution* 34.4, S. 327–341. DOI: <https://doi.org/10.1016/j.tree.2018.12.012>.
- Zimmermann, K., J. Behrendt und E. Penda (2015). *Konzept zur Qualitätskontrolle historischer Klimadaten*. URL: [https://www.dwd.de/DE/klimaumwelt/klimaueberwachung/klimadatenverarbeitung/qualitaetssicherung/download\\_quali\\_hist\\_doku.pdf?\\_\\_blob=publicationFile&v=5](https://www.dwd.de/DE/klimaumwelt/klimaueberwachung/klimadatenverarbeitung/qualitaetssicherung/download_quali_hist_doku.pdf?__blob=publicationFile&v=5) (besucht am 20.03.2022).

Zimmermann-Janschitz, S. (2014). *Statistik in der Geographie*. Berlin Heidelberg: Springer-Verlag.

## Anhang

Im Zusammenhang mit der Ergebnisdarstellung in den Kapiteln **4.2.1 Vergleich der Arable-Station Erdeborn mit der DWD-Station Querfurt** und **textbf4.2.2 Vergleich der Arable-Stationen Bentzin mit der DWD-Station Greifswald** wurde auf ausgewählte Abbildungen eingegangen, die innerhalb des Fließtextes nicht dargestellt wurden. Diese Abbildungen werden im nachfolgenden der Reihe nach aufgelistet.

### Erläuterung zu den Liniendiagrammen

#### Stationsvergleich: Erdeborn-Querfurt

Neben der Darstellung der Durchschnitts- beziehungsweise Summenwerte der Klimavariablen wurden im Zuge des **Unterkapitels 4.2.1** ein kürzerer Zeitraum basierend auf den Stundenwerten betrachtet. Dieser Zeitraum deckt die Tage 14. - 21.07.2021 ab und soll einen Eindruck über die Unterschiede der Stationen anhand der Klimavariablen liefern. In Orange wird jeweils die Arable-Station und in Grün die DWD-Station dargestellt. Diese drei Abbildungen 15, 16 , und 17 finden sich ab Seite 90 und bis 91.

#### 7 tägiger Vergleich der Stundenwerte der Temperatur für Erdeborn und Querfurt

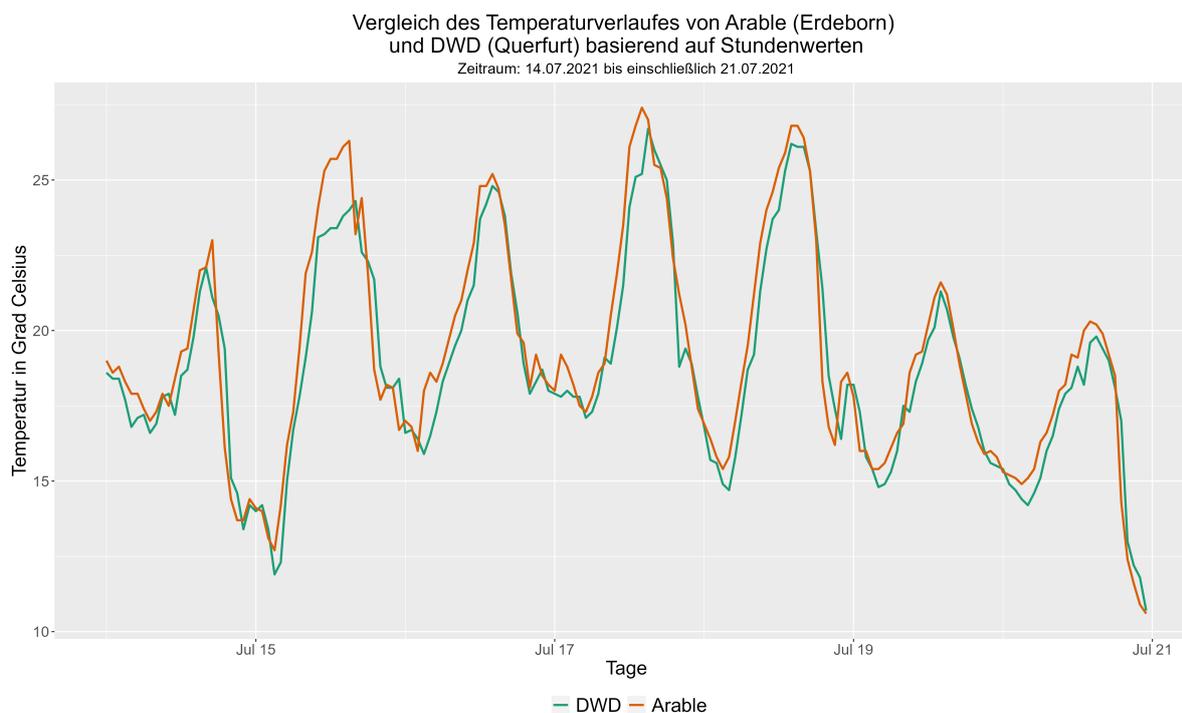


Abbildung 15: 7 tägiger Vergleich der Stundenwerte der Temperatur für Erdeborn und Querfurt (Eigene Darstellung).

**7 tägiger Vergleich der Stundenwerte des Niederschlages für Erdeborn und Querfurt**

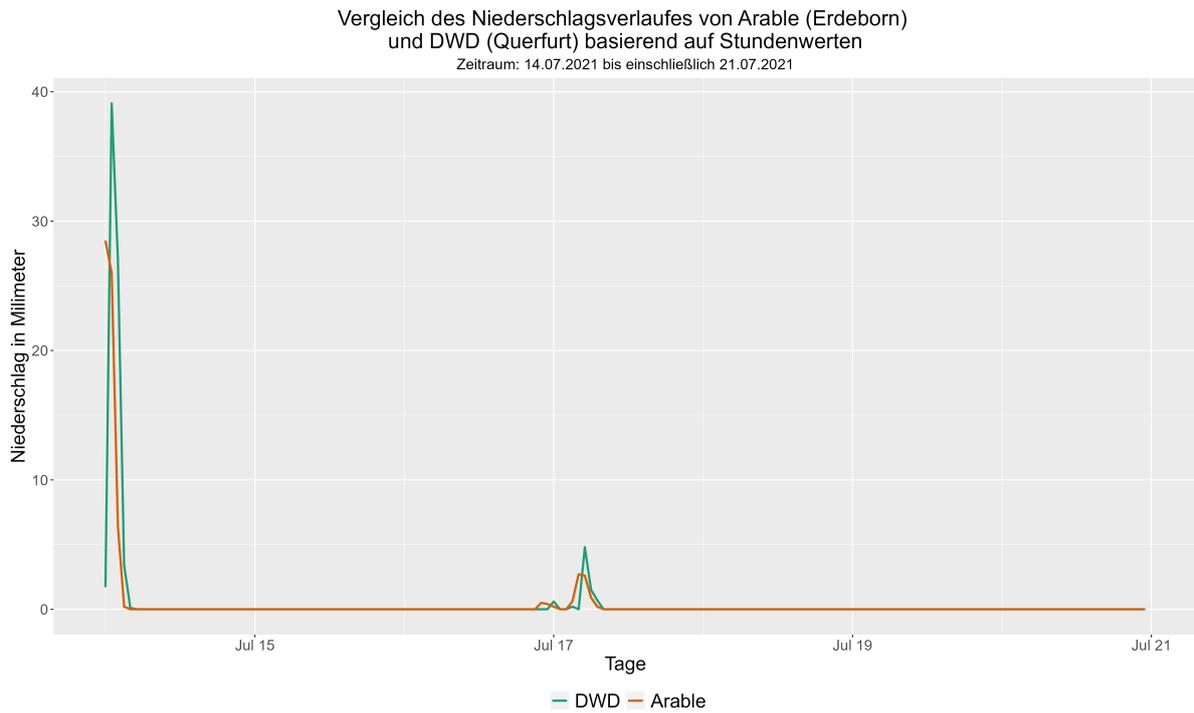


Abbildung 16: 7 tägiger Vergleich der Stundenwerte des Niederschlages für Erdeborn und Querfurt (Eigene Darstellung).

**7 tägiger Vergleich der Stundenwerte der rel. Luftfeuchtigkeit für Erdeborn und Querfurt**

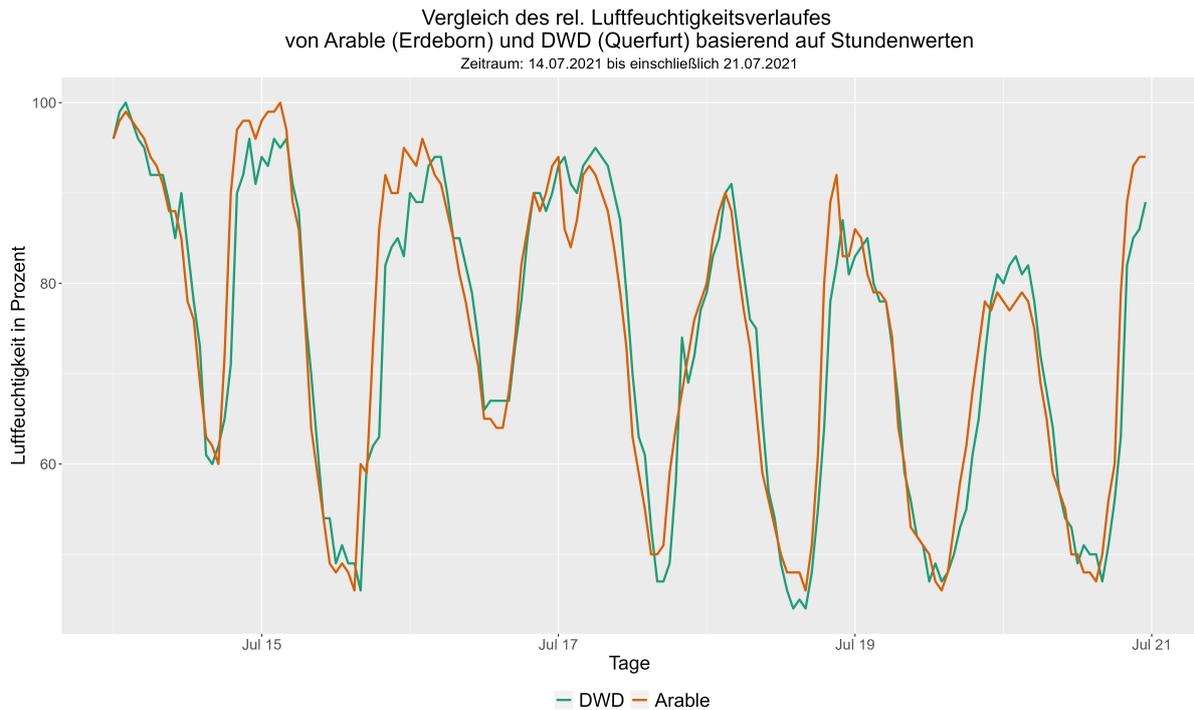


Abbildung 17: 7 tägiger Vergleich der Stundenwerte der relativen Luftfeuchtigkeit für Erdeborn und Querfurt (Eigene Darstellung).

## Erläuterung zu den Histogrammen

### Stationsvergleich: Erdeborn-Querfurt

Die Abbildungen 18, 19, 20, 21, 22 und 23, beginnend mit Seite 93 und endend mit Seite 95 basieren auf den unter **4.2.1 Vergleich der Arable-Station Erdeborn mit der DWD-Station Querfurt** dargestellten Vergleich zwischen den Stationen Erdeborn (Arable) und Querfurt (DWD). Dargestellt werden jeweils die drei Klimavariablen *Lufttemperatur*, *Niederschlag* und *Relative Luftfeuchtigkeit*, die jeweils für die Arable- (in orange) und einmal für die DWD-Station (in grün) in abwechselnder Folge dargestellt werden.

### Stationsvergleich: Bentzin-Greifswald

Die Abbildungen 24, 25, 26, 27, 28 und 29, beginnend mit Seite 96 und endend mit Seite 101 entsprechen den unter **4.2.2 Vergleich der Arable-Stationen Bentzin mit der DWD-Station Greifswald** thematisierten Darstellungen der Histogramme des Stationsvergleiches Bentzin-Greifswald. Diese sechs Abbildungen stellen Zusammenstellungen von jeweils fünf Histogrammen dar, wobei ein jedes Histogramm für eine der fünf untersuchten Stationen des Bentzin-Greifswald-Vergleiches stehen. Für jede Klimavariablen werden nacheinander zwei Abbildungen aufgeführt, die vom Aufbau her identisch sind. Der Unterschied zwischen diesen liegt in dem betrachteten Zeitraum der Daten. Abbildungen 24, 26 und 28 repräsentieren den gesamten Untersuchungszeitraum vom 04.05. bis 04.07.2022. Die Abbildungen 25, 27 und 29 hingegen stellen den Zeitraum vom 10.05. bis 04.07.2022 dar. Diese Gegenüberstellung soll es den Betrachter ermöglichen, sich die Auswirkungen der fragwürdigen Werte vom 04.05. bis 10.05.2022 auf die Verteilung der Messdaten zu erschließen.

**Histogramm: Lufttemperatur für Erdeborn (Arable)**

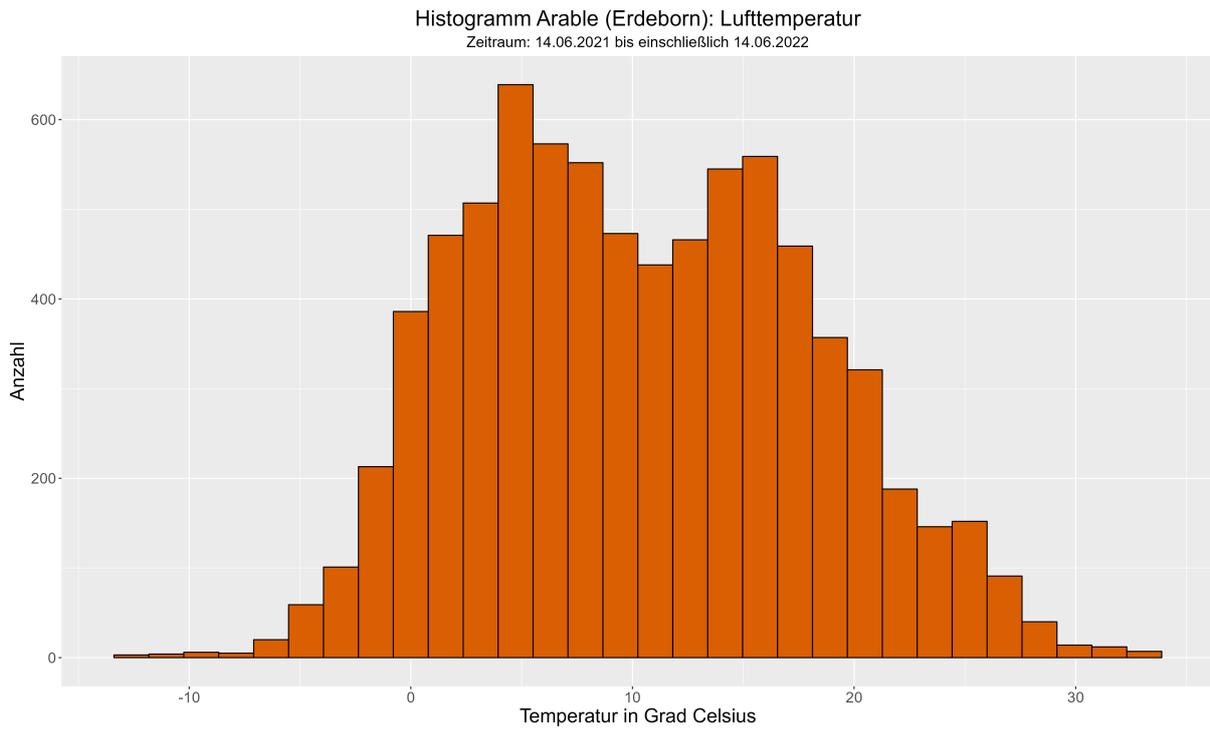


Abbildung 18: Histogramm der Lufttemperatur für Erdeborn (Arable) (Eigene Darstellung).

**Histogramm: Lufttemperatur für Querfurt (DWD)**

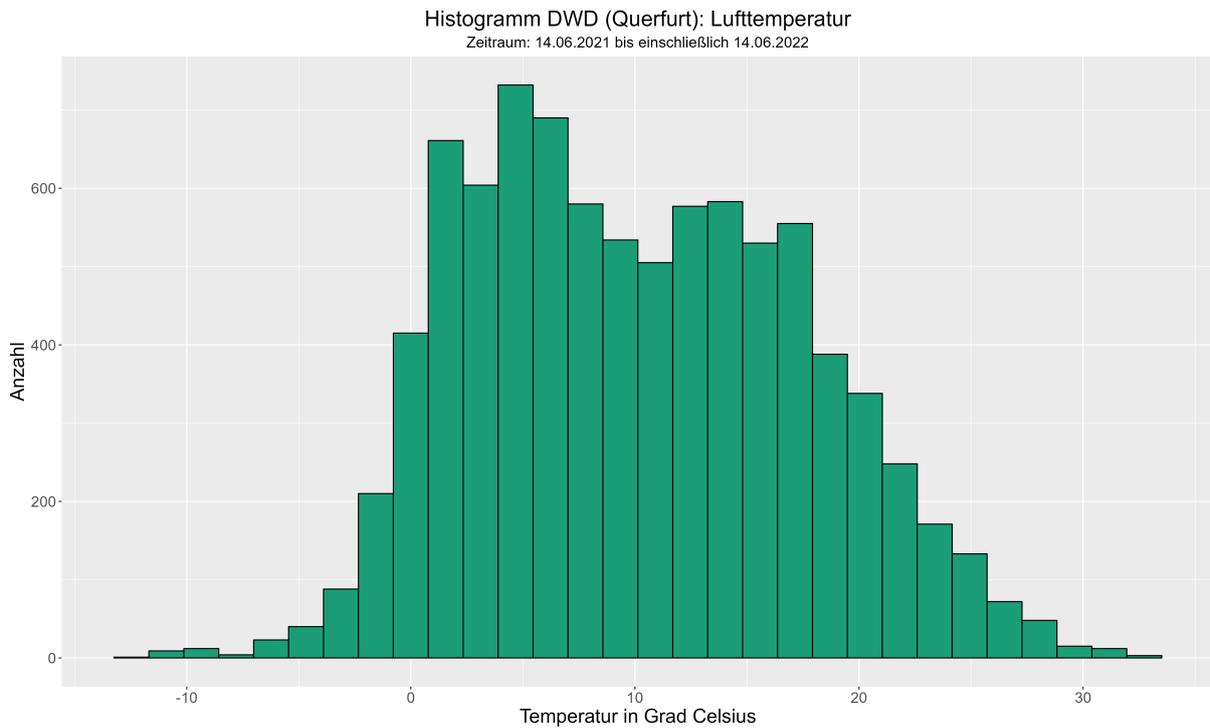


Abbildung 19: Histogramm der Lufttemperatur für Querfurt (DWD) (Eigene Darstellung).

**Histogramm: Niederschlag für Erdeborn (Arable)**

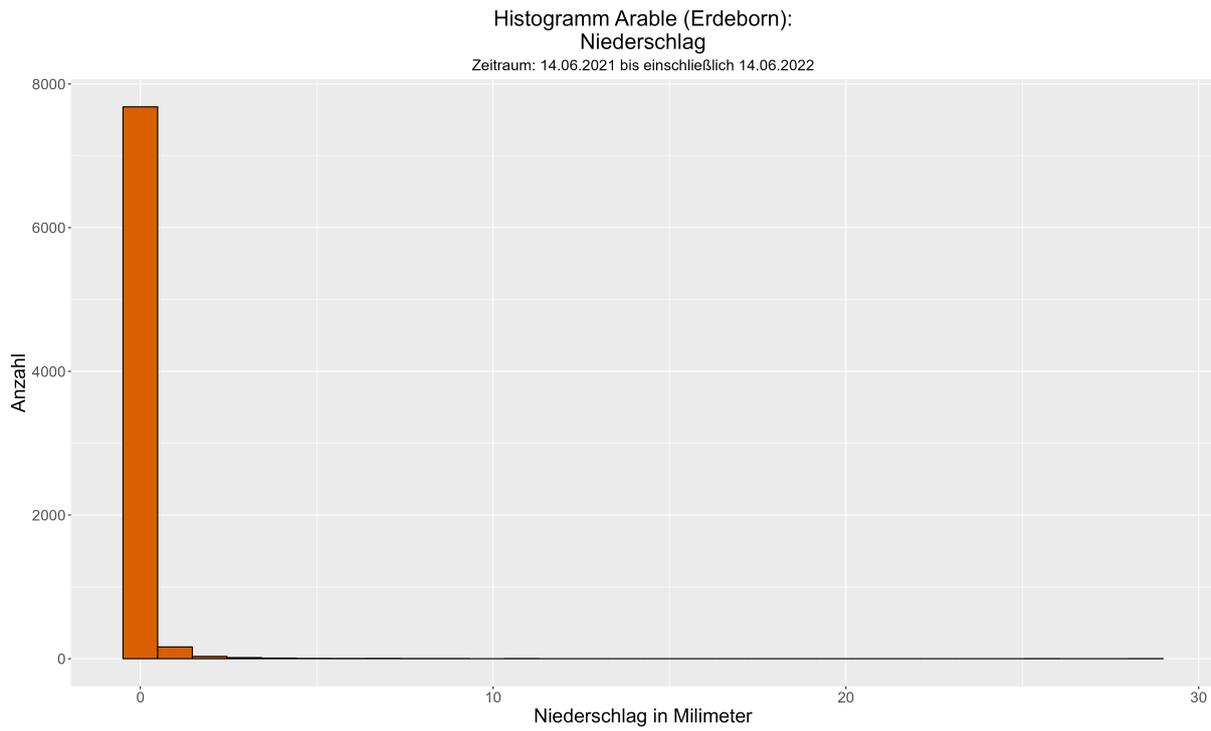


Abbildung 20: Histogramm des Niederschlages für Erdeborn (Arable) (Eigene Darstellung).

**Histogramm: Niederschlag für Querfurt (DWD)**

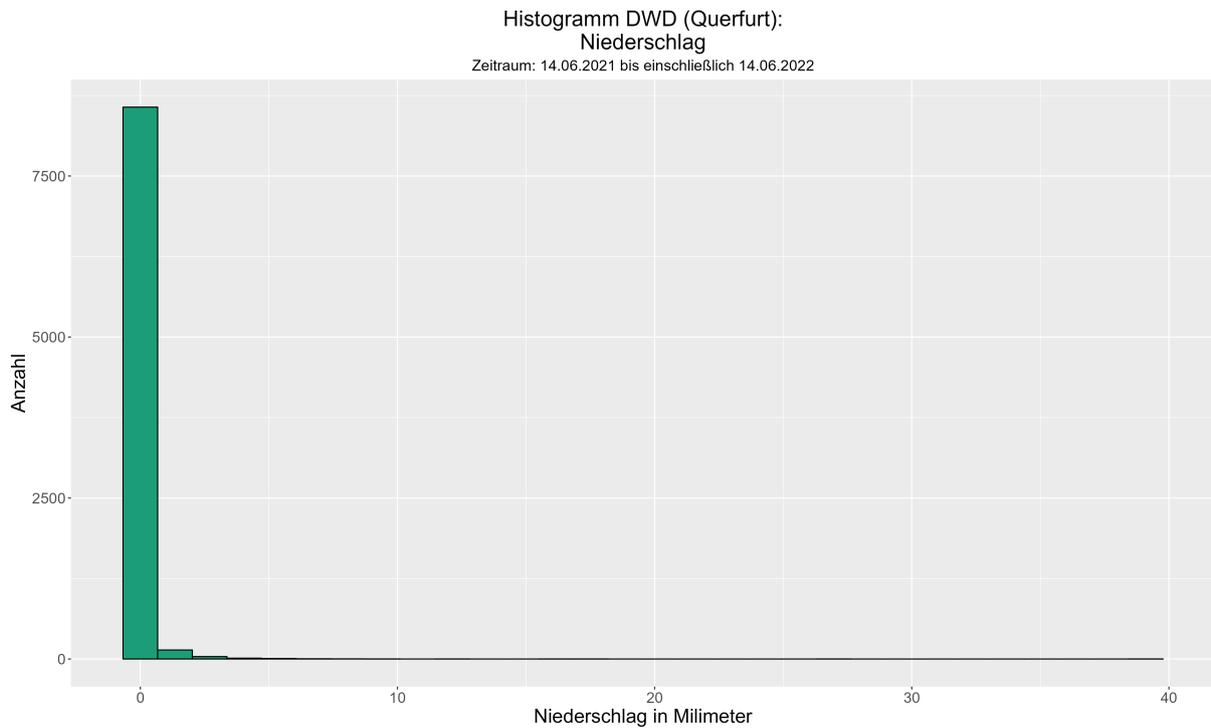


Abbildung 21: Histogramm des Niederschlages für Querfurt (DWD) (Eigene Darstellung).

**Histogramm: Relative Luftfeuchtigkeit für Erdeborn (Arable)**

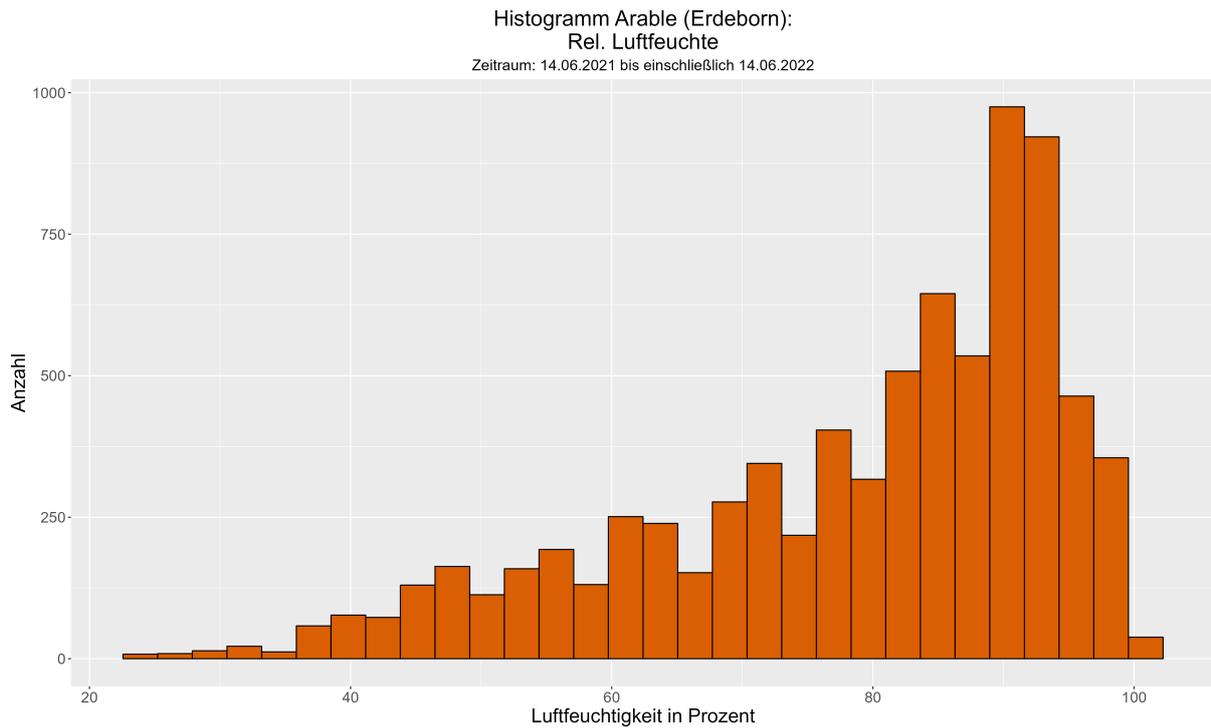


Abbildung 22: Histogramm der relativen Luftfeuchtigkeit für Erdeborn (Arable) (Eigene Darstellung).

**Histogramm: Relative Luftfeuchtigkeit für Querfurt (DWD)**

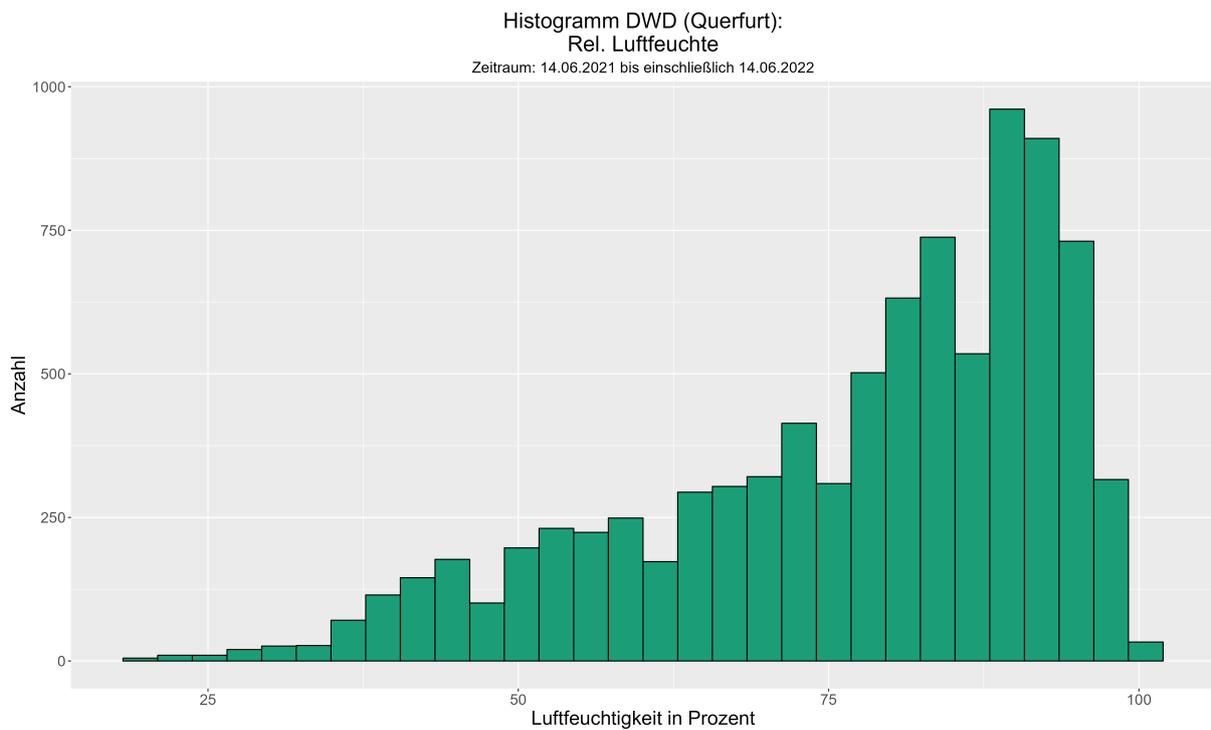


Abbildung 23: Histogramm der relativen Luftfeuchtigkeit für Querfurt (DWD) (Eigene Darstellung).

**Histogramm: Lufttemperaturwerte Bentzin und Greifswald mit 4.-10.05.2022**

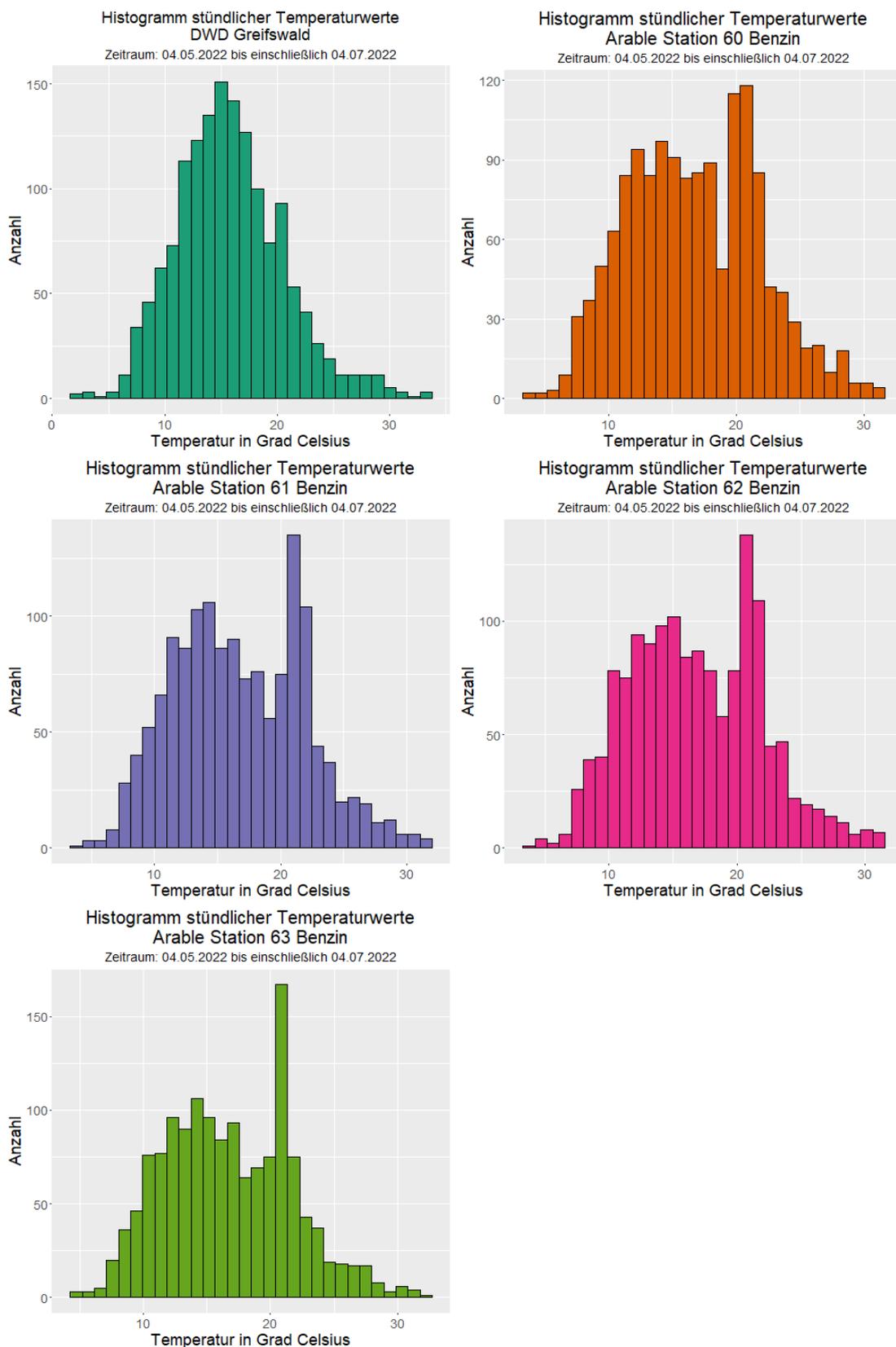


Abbildung 24: Histogramme: Lufttemperatur Bentzin (Arable) und Greifswald (DWD) ab 04.05.2022 (Eigene Darstellung).

**Histogramm: Lufttemperaturwerte Bentzin und Greifswald ohne 4.-10.05.2022**

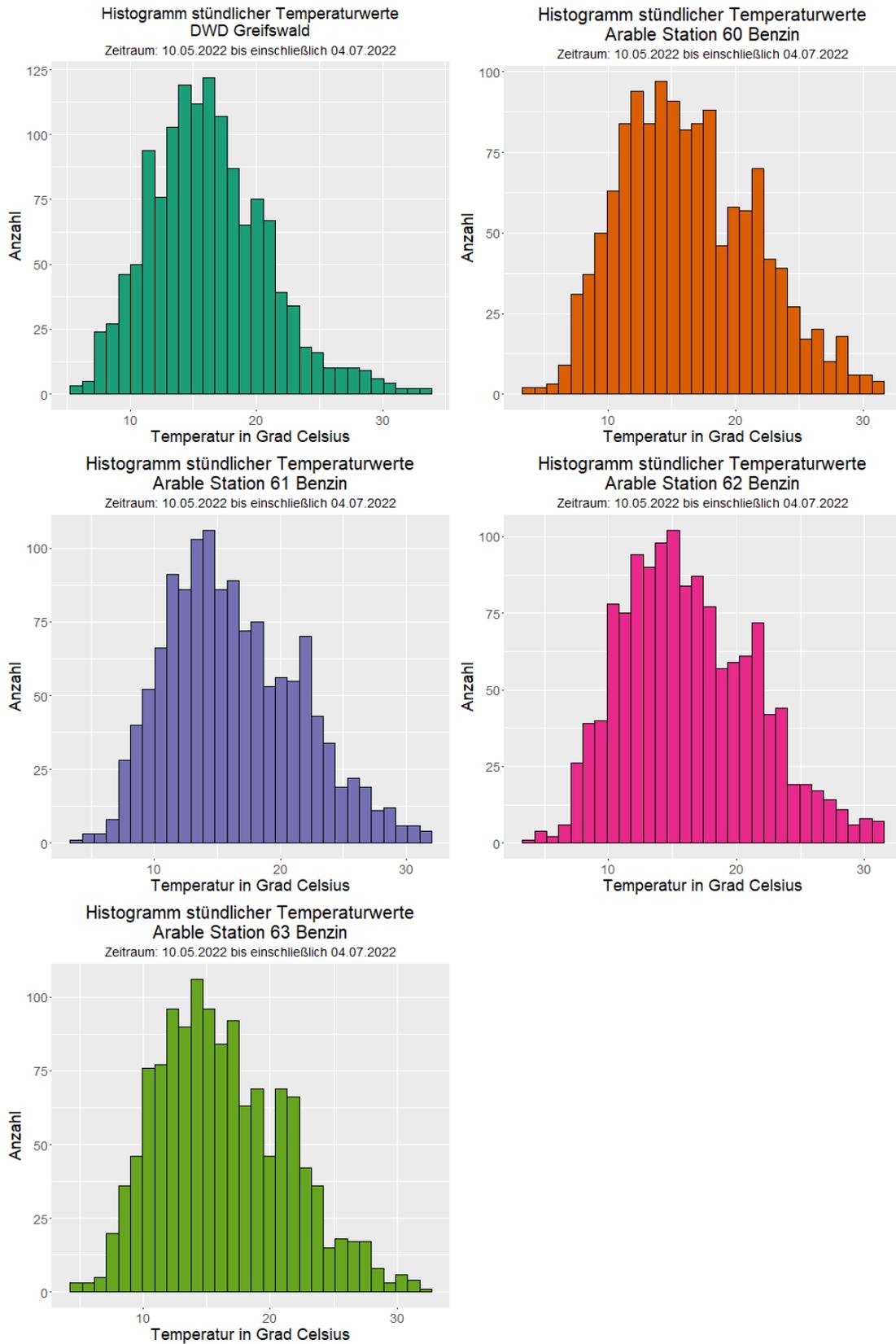


Abbildung 25: Histogramme: Lufttemperatur Bentzin (Arable) und Greifswald (DWD) ab 10.05.2022 (Eigene Darstellung).

**Histogramm: Niederschlagswerte Bentzin und Greifswald mit 4.-10.05.2022**

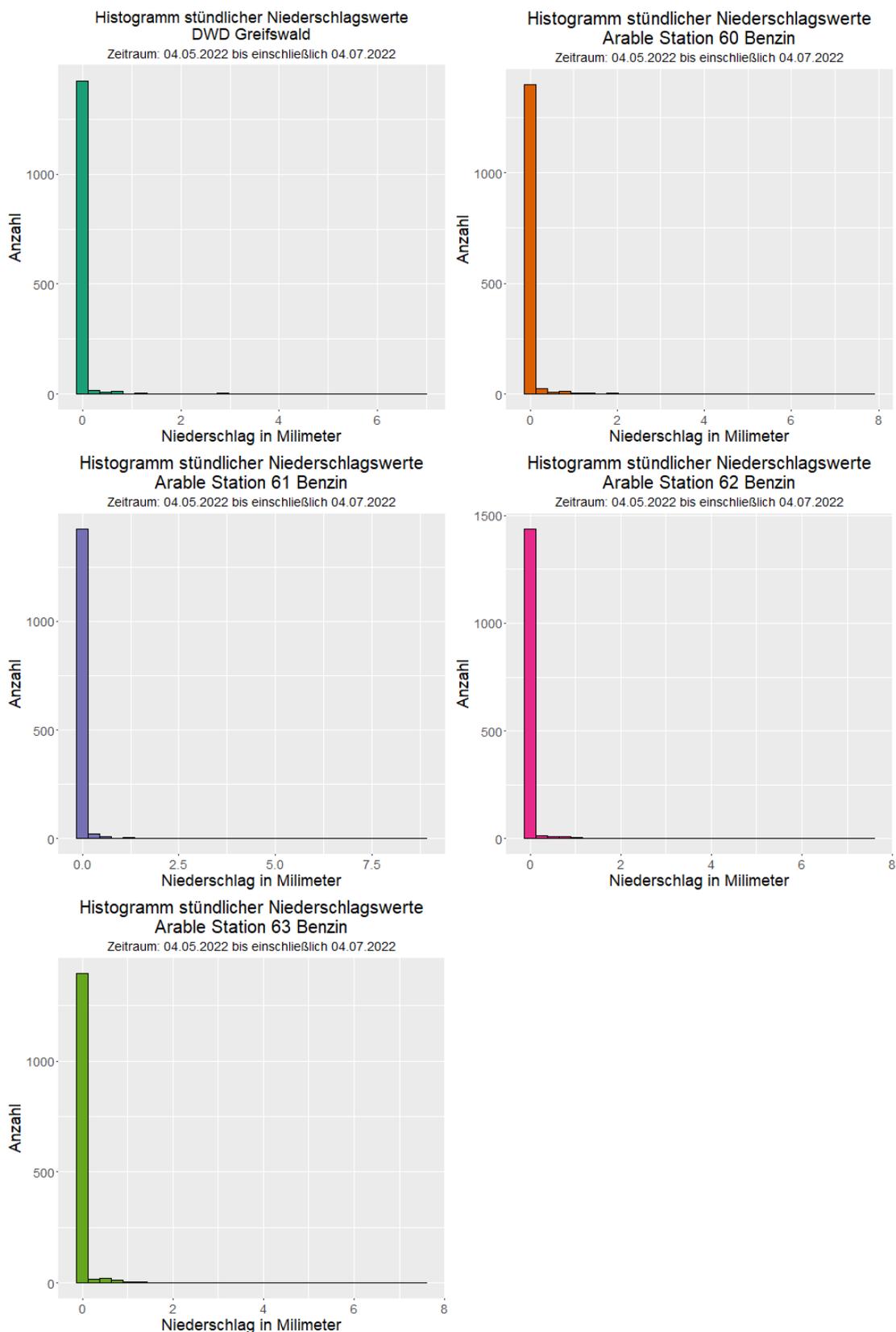


Abbildung 26: Histogramme: Niederschlag Bentzin (Arable) und Greifswald (DWD) ab 04.05.2022 (Eigene Darstellung).

**Histogramm: Niederschlagswerte Bentzin und Greifswald ohne 4.-10.05.2022**

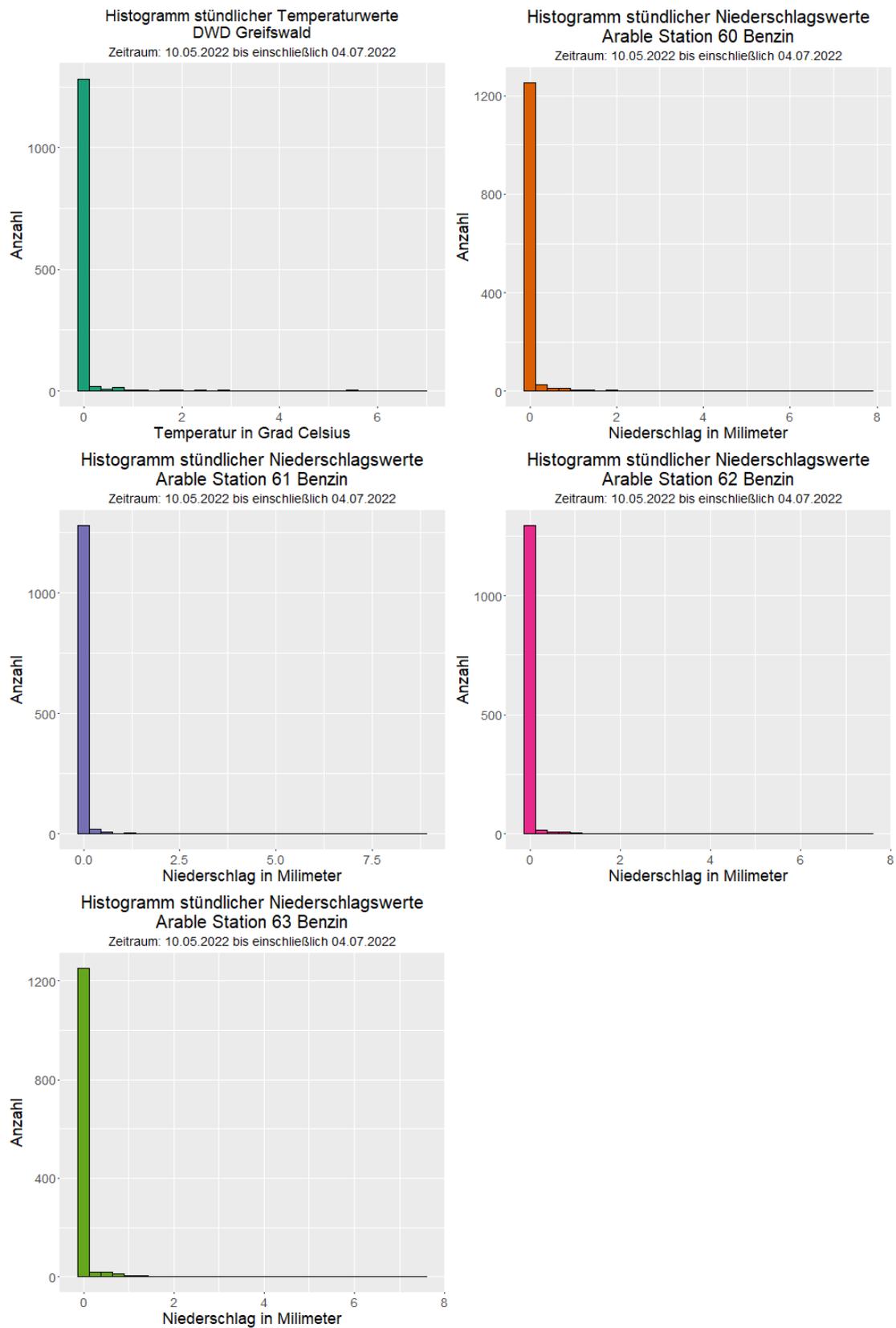


Abbildung 27: Histogramme: Niederschlag Bentzin (Arable) und Greifswald (DWD) ab 10.05.2022 (Eigene Darstellung).

**Histogramm: Luftfeuchtwerte Bentzin und Greifswald mit 4.-10.05.2022**

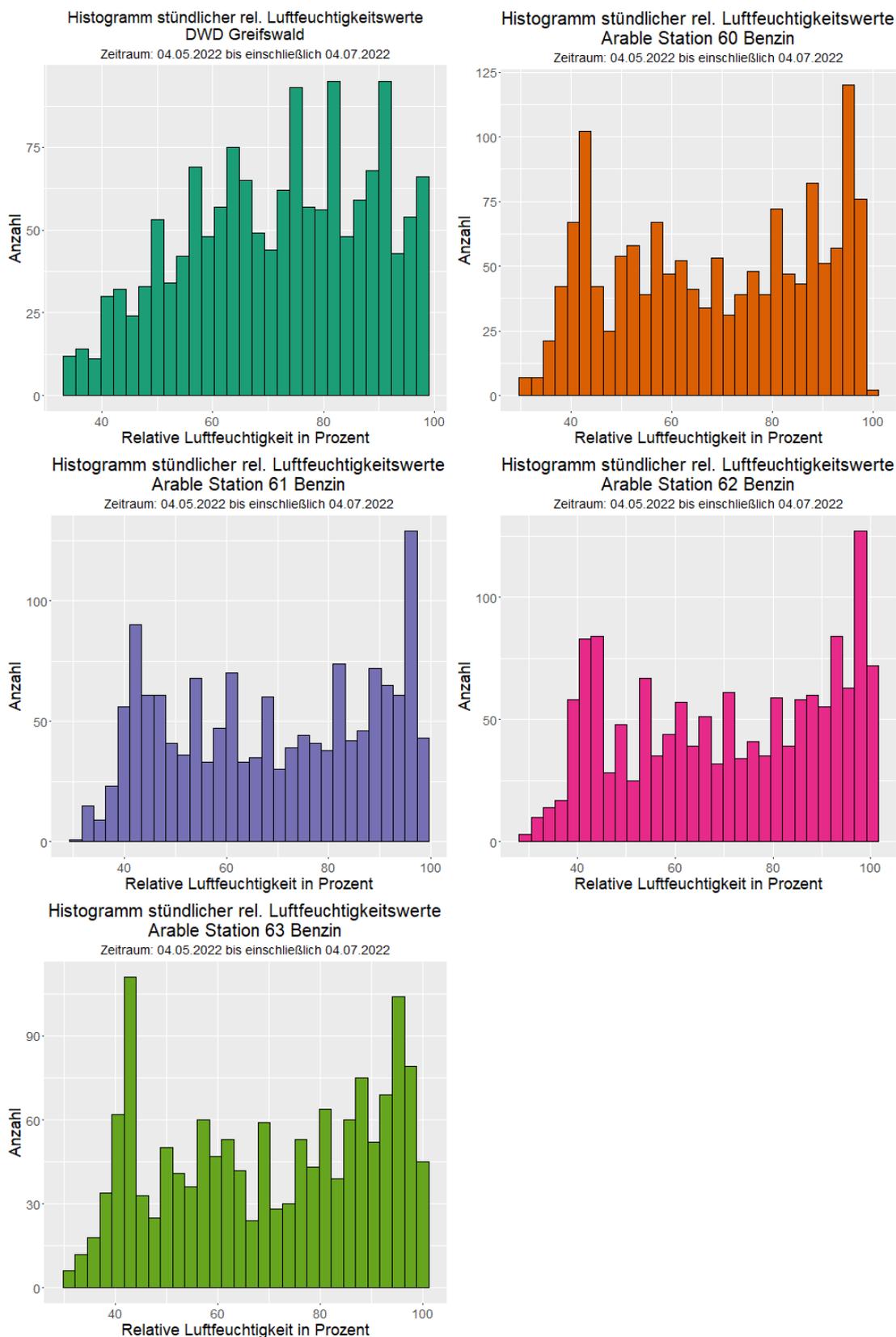


Abbildung 28: Histogramme: Rel. Luftfeuchtigkeit Bentzin (Arable) und Greifswald (DWD) ab 04.05.2022 (Eigene Darstellung).

**Histogramm: Luftfeuchtwerte Bentzin und Greifswald ohne 4.-10.05.2022**

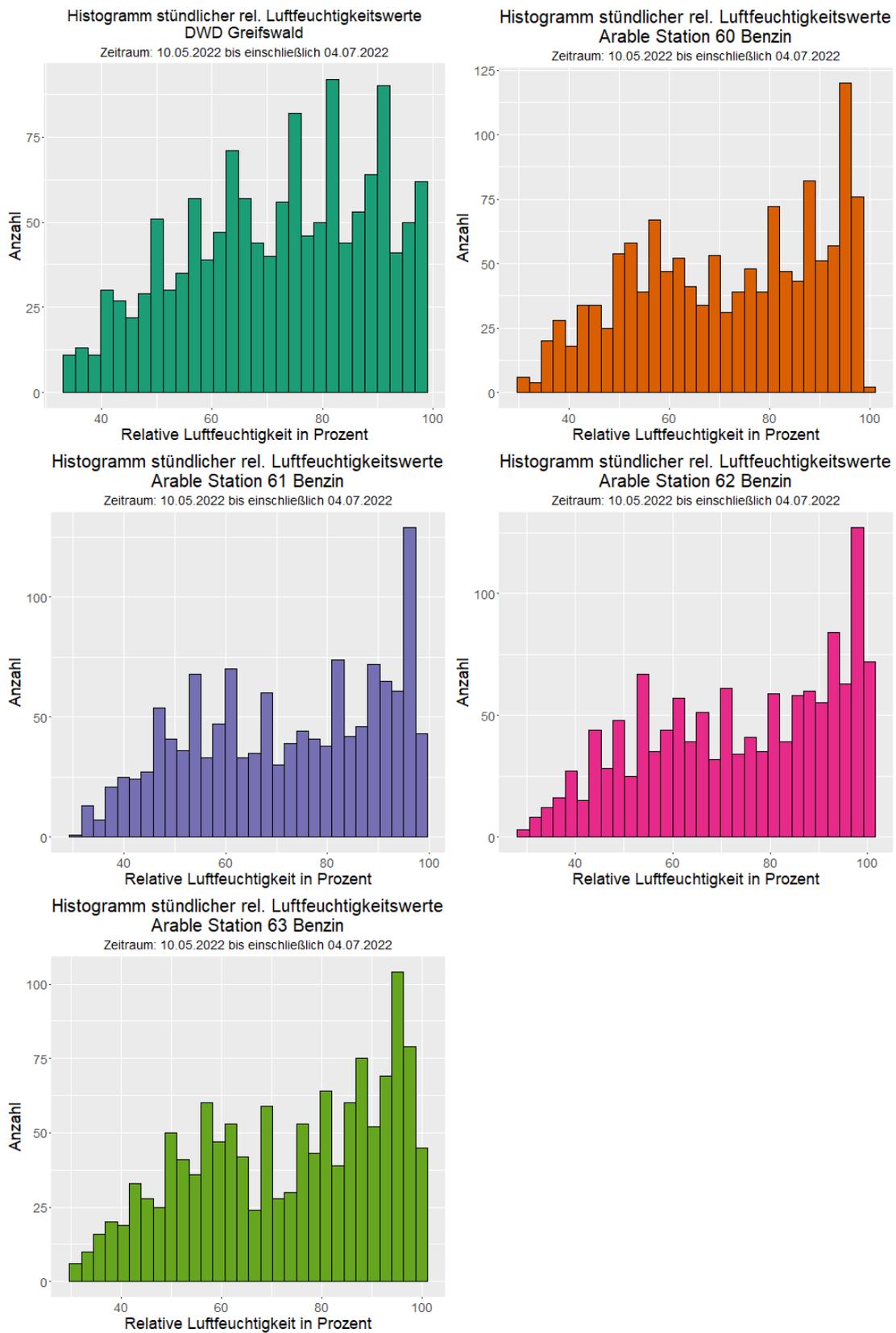


Abbildung 29: Histogramme: Rel. Luftfeuchtigkeit Bentzin (Arable) und Greifswald (DWD) ab 10.05.2022 (Eigene Darstellung)

### Regression Bentzin

Den Tabellen 7, 8 und 7 folgend, wurden drei Regressionsdiagramme erstellt, die die Regressionen zwischen den einzelnen Stationen für jede Klimavariablen gemeinsam darstellen.

Diese Abbildungen 30, 31 und 32 sind dabei wie folgt zu lesen: die Grafiken in der ersten Zeile rechts neben *DWD* in Abbildung 30 stellen die Regressionen der Regressionsanalysen zwischen Arable 60 und DWD, Arable 61 und DWD, und so weiter dar.

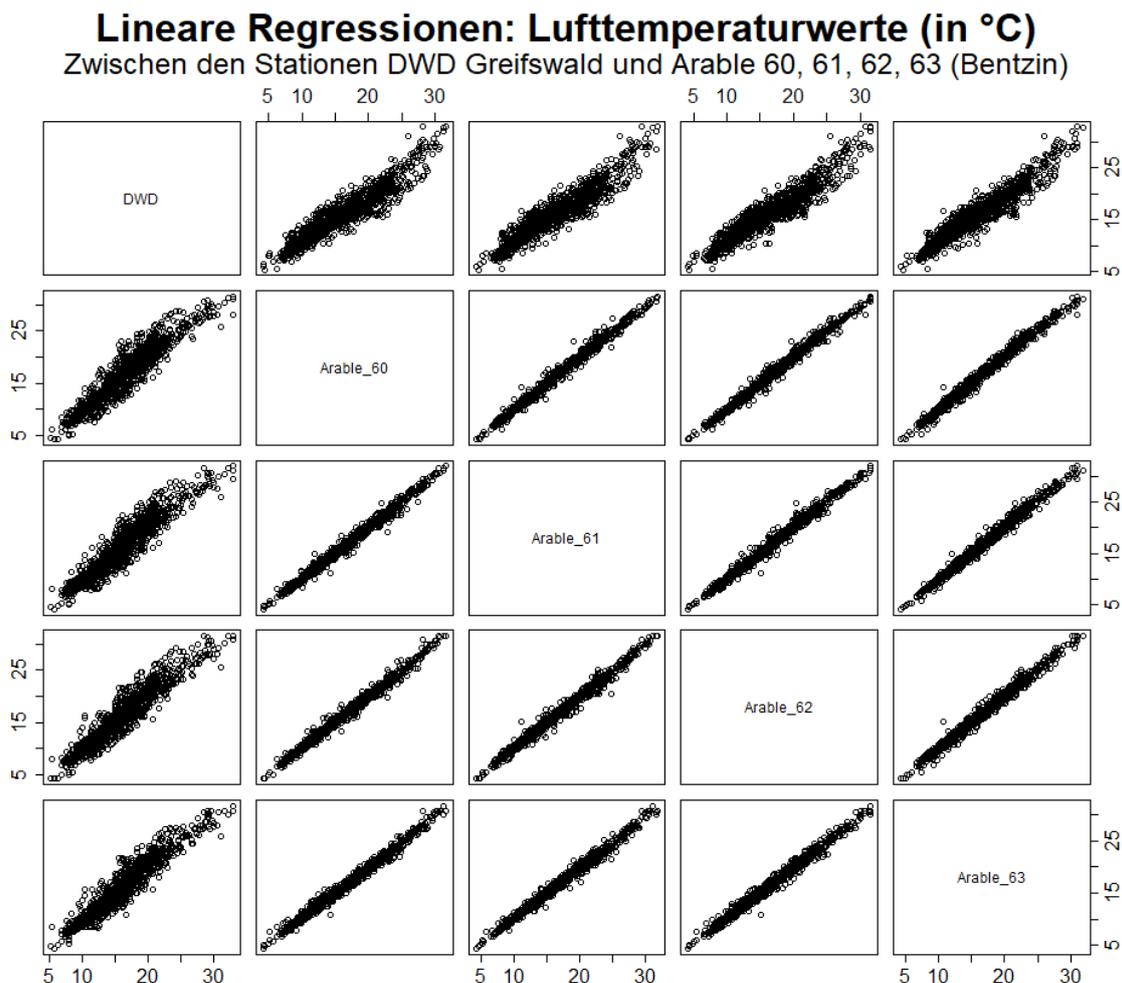


Abbildung 30: Lineare Regressionen der Lufttemperaturwerte für Bentzin (Arable) und Greifswald (DWD) (Eigene Darstellung).

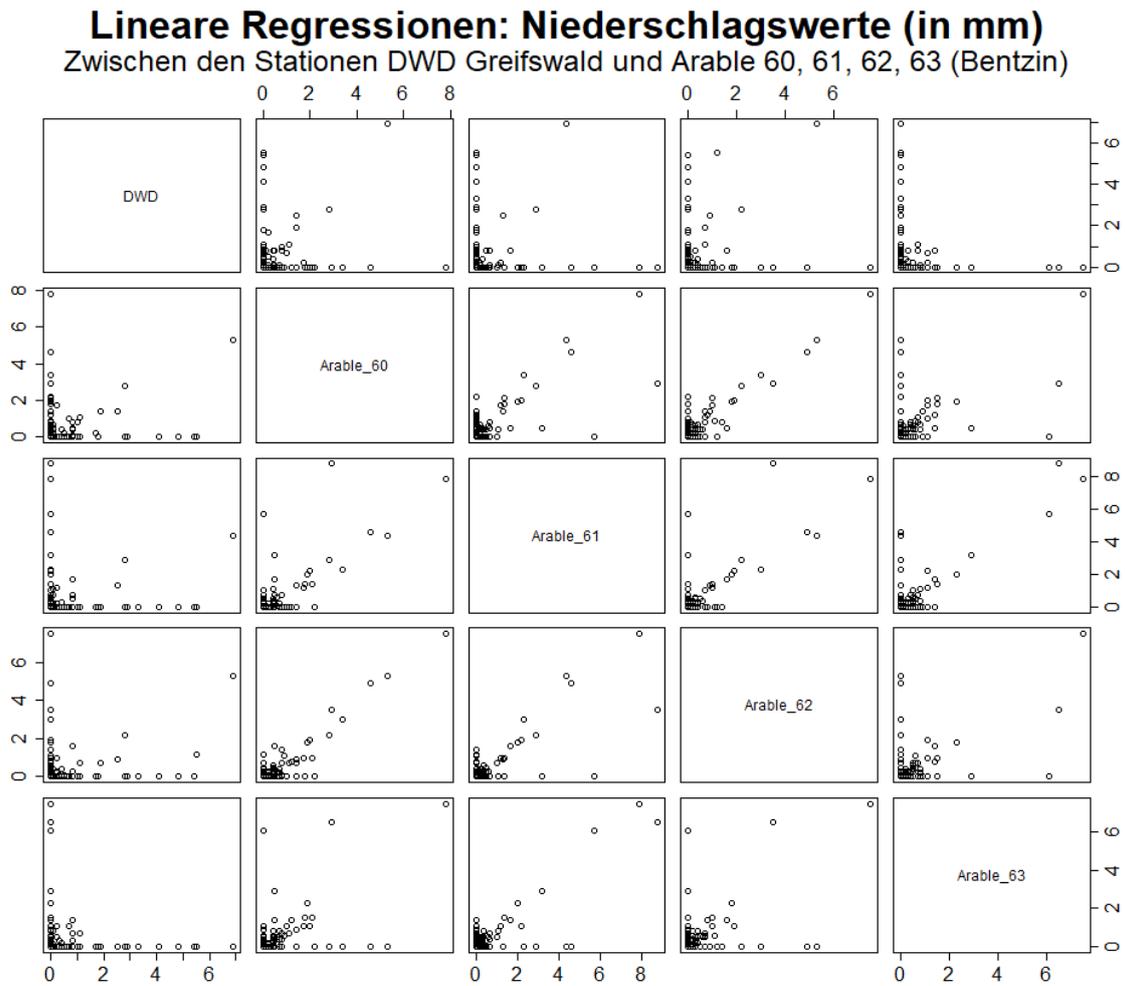


Abbildung 31: Lineare Regressionen der Niederschlagswerte für Bentzin (Arable) und Greifswald (DWD) (Eigene Darstellung).

**Lineare Regressionen: Relativer Luftfeuchtwerte (in %)**  
 Zwischen den Stationen DWD Greifswald und Arable 60, 61, 62, 63 (Bentzin)

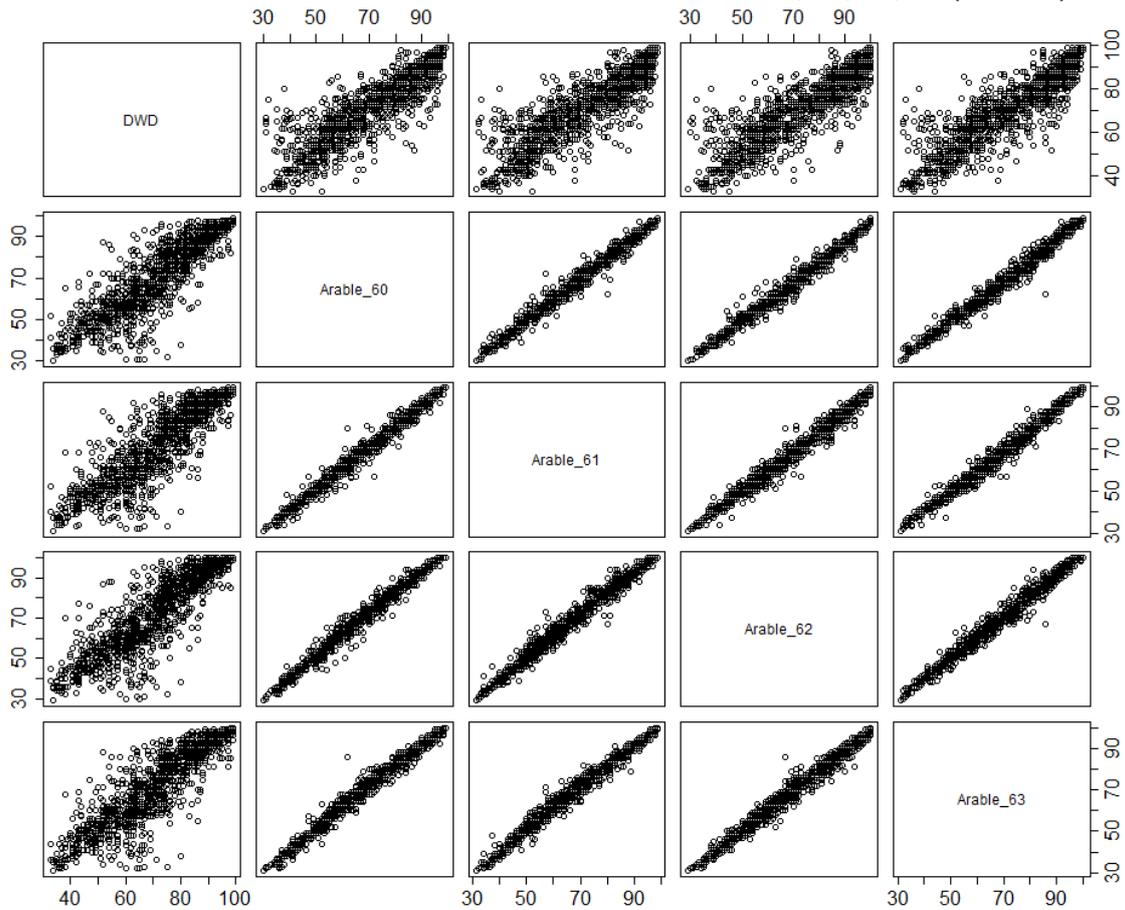


Abbildung 32: Lineare Regressionen der relativen Luftfeuchtwerte für Bentzin (Arable) und Greifswald (DWD) (Eigene Darstellung).

## Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit mit dem Titel „Anwendung von mobilen Wetterdaten zur Analyse und Einschätzung von Mikroklimaten“ selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Stellen sind als solche kenntlich gemacht.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Die digitalisierte Fassung der Arbeit stimmt mit der in Schriftform vorgelegten Fassung wörtlich überein.

**Name, Vorname:** Lorenz, Christopher Günter

**Matrikelnummer:** 219226310

---

Ort, Datum

Unterschrift