



Meteorologisches Bürgermessnetz aus dem Projekt OpenUCO Berlin 2022

Datenpublikation

C. Böttcher, V. Sobottke, H. W. Rust, F. Meier, M. Otto,

J. Ulrich, J. Lehmke, D. Schoster

22. Dezember 2024



1 Messkampagne

Citizen Science, zu deutsch Bürgerwissenschaft, bezeichnet eine Form der wissenschaftlichen Forschung, bei der Bürgerinnen und Bürger aktiv am wissenschaftlichen Prozess teilnehmen [Vohland et al., 2021]. Eine Form der Teilnahme ist die aktive Unterstützung beim Sammeln von Daten. Die gesammelten meteorologischen Daten (Lufttemperatur, Luftdruck, Niederschlag, Luftfeuchtigkeit, Beleuchtungsstärke) aus dem Citizen Science Projekt *Open Urban Climate Observatory (OpenUCO)*, insbesondere aus einer Messkampagne im Sommer 2022 werden hier genauer beschrieben und zur Verfügung gestellt.

Das *meteorologische Bürgermessnetz* war der Kern des Citizen Science Projekts *Open Urban Climate Observatory Berlin (OpenUCO Berlin)*, welches in den Jahren 2022 und 2023 durchgeführt wurde. Die Messungen fokussierten jeweils auf die Sommermonate, wobei vereinzelte Stationen auch über einen längeren Zeitraum gemessen haben. Das Projekt *OpenUCO* und damit auch die Messkampagne wurde von der *Berlin University Alliance* (BUA, <https://www.berlin-university-alliance.de/>) finanziert und von der *Freien Universität Berlin* (FUB) und der *Technischen Universität Berlin* (TUB) in Berlin durchgeführt. Die Messstationen wurden dabei vornehmlich in Kleingärten von teilnehmenden Bürger:innen platziert. Ziel des Projektes war die Erweiterung des Verständnisses des sich verändernden Stadtklimas Berlins aufgrund des Bevölkerungszuwachses und der räumlichen Heterogenität. Dabei wurden zwei Ansätze zur Einbeziehung der Kleingärtner:innen verfolgt:

ko-produktiv Bürger:innen messen atmosphärische Größen im Kleingarten

ko-kreativ Bürger:innen entwickeln eigene, für die Stadtgesellschaft relevante Forschungsfragen

Für die Messkampagne kamen zwei Gerätetypen zum Einsatz:

- Die an der *Freien Universität Berlin* entwickelten sogenannten *MESSIs* (Akronym für „Mein eigenes Subskalen Messinstrument für atmosphärische Größen“, siehe Abb. 1) und
- die von der Firma *Pessl Instruments* entwickelten und vertriebenen LoRains (<https://metos.global/de/about-us/>, <https://metos.global/de/lorain/>).

Beide Gerätetypen werden im Abschnitt 3 näher beschrieben.

2 Wissenschaftliche Fragestellung

Die Berlin University Alliance (BUA) fördert Open Science Projekte im Rahmen des Objective 3 „Advancing Research Quality and Value“. Dazu zählt das Citizen Science Projekt „Open Urban Climate Observatory Berlin (OpenUCO)“. Dabei werden wichtige Fragen zur Auswirkung von Grünflächen auf die Lufttemperatur und zu räumlichen Unterschieden bei Starkregen untersucht. Ein dichtes Netzwerk von Messstationen ist dafür von entscheidender Bedeutung. Da die Installation und Wartung einer großen Anzahl von Stationen in der gesamten Stadt jedoch sehr aufwändig ist, hat OpenUCO-Berlin die Unterstützung der Berlinerinnen und Berliner gesucht und gefunden. Die meisten Stationen senden ihre Messdaten über das freie *LoRa*-Funk-Netzwerk¹, um eine zeitnahe Datenverarbeitung zu ermöglichen. Zusätzlich zu den Wetterstationen wurden auch zahlreiche LoRa-Gateways an verschiedenen Standorten in Berlin installiert, um die Datenverfügbarkeit in Echtzeit zu erhöhen.

Ein Großteil der Messstationen wurde in den Kleingärten der Stadt aufgebaut, denn mit den Messdaten soll unter anderem die Wirkung von Kleingärten auf das Stadtklima untersucht werden. Außerdem ermöglicht die große Anzahl von Niederschlagsmessstationen zusammen mit dem von FUB und TUB gemeinsam betriebenen Wetterradar die Erforschung der Verteilung von Niederschlagsfeldern über der gesamten Stadt (<https://www.klima.tu-berlin.de/radar/>). Nicht zuletzt können die Bürgerwissenschaftler:innen die Messdaten ihrer Station nutzen, um zum Beispiel die Häufigkeit und Menge der Bewässerung in ihrem Garten abzuschätzen. Das Interesse am Projekt und die Nachfrage nach den Wetterstationen war so groß, so dass nur etwa die Hälfte der Interessenten eine *OpenUCO*-Wetterstation in ihrem Garten in Betrieb nehmen konnten. Die Standorte und Live-Messwerte aller derzeit aktiven Messstationen sind auch für alle, die nicht direkt beteiligt sind, öffentlich einsehbar unter https://messi.openuco.berlin/public_devices.

3 Verwendete Messgeräte

3.1 MESSI: Allgemeines

Der MESSI ist eine kostengünstige automatische Wetterstation, die speziell für Citizen Science Projekte an der Freien Universität entwickelt wurde². Die Station wurde mit Technologien entwickelt, die das sogenannte „Internet-of-things“ mit sich bringt, das heißt es wurden kostengünstige und energiesparende Sensoren und Mikrocontroller verbaut,

¹<https://resources.lora-alliance.org/technical-specifications>

²Der MESSI wurde von Jonas Lehmke entwickelt jonas@lehmke.xyz und in weiten Teilen von der Firma Lehmke Prototype gefertigt

um möglichst viele Geräte herstellen zu können, die auch unabhängig vom Stromnetz sind. Desweiteren erlaubt der Funkstandard *LoRa* eine weitreichende und energiesparende Möglichkeit die Messdaten in Echtzeit zu übertragen. Das Gerät misst Luftdruck, Luftfeuchtigkeit, Lufttemperatur (mit drei verschiedenen Sensoren) und Beleuchtungsstärke, siehe Abb. 1. Mithilfe einer zusätzlich an den MESSI montierten Niederschlagswippe, konnte der MESSI zudem Niederschlagsdaten aufzeichnen.



Abbildung 1: Automatische Wetterstation *MESSI*



Abbildung 2: Erweiterung des *MESSIs* mit einer kostengünstigen Niederschlagswippe

Entwicklungsziele für diese Messplattform waren unter anderem die folgenden, siehe auch die Darstellung in [Kox et al., 2021]:

Günstig Um ein dichtes Messnetz zu erreichen, soll eine große Anzahl Geräte gebaut

werden.

Autonom Der MESSI soll weder vom Stromnetz noch vom W-LAN abhängig sein, damit eine Standortwahl ohne Einschränkungen durch Nähe zum Gebäude getroffen werden kann.

Vergleichbarkeit Es sollte möglich sein, die Messwerte sinnvoll in Beziehung zu solchen von professionellen Messgeräten zu setzen.

Attraktivität Das Gerät und die Nutzung soll für die Teilnehmer:innen ansprechend sein.

Einfachheit Um mögliche Einflüsse durch falsche Bedienung zu vermeiden, soll das Gerät möglichst einfach zu bedienen sein.

Diese autonome Messplattform MESSI wurde für atmosphärische Messungen konzipiert und basiert auf dem Mikrocontroller *SAML21*³ der Firma *Microchip*. Klassische meteorologische Netze sind aufgrund hoher Kosten einer einzelnen Messstation in der Anzahl der Einheiten stark begrenzt. Dies beschränkt die möglichen Forschungsfragen, die mit den von diesen Stationen gelieferten Daten behandelt werden können. Die *MESSI*-Plattform schafft hier Abhilfe und zielt auf den Aufbau eines kostengünstigen, dichten und offenen meteorologischen Messnetzes unter Einbeziehung der Bürger:innen ab. Meteorologische Phänomene, die im submesoskaligen und städtischen Kontext auftreten, können mit diesen günstigen Methoden beobachtet werden. Bei einem konventionellen grobmaschigen Messnetz fallen diese Phänomene durch das Raster und entziehen sich der Beobachtung.

Die oben beschriebenen Designanforderungen führen zu folgenden Hauptmerkmalen des MESSIs:

- unabhängige Energieversorgung mit Solarzelle und LiFePo4-Akku
- Funktechnologie LoRaWAN ermöglicht Datenübertragung und Überwachung in nahezu Echtzeit und eine energiesparende Zeit- und Datumssynchronisation
- interner 32 MB Flash-Speicher (Auslesen über USB möglich)
- GPS ermöglicht Standortbestimmung und eine präzise Zeitsynchronisation
- modularer Aufbau
 - Sensorplatinen können kostengünstig und einfach ausgetauscht werden

³<https://www.microchip.com/en-us/products/microcontrollers-and-microprocessors/32-bit-mcus/sam-32-bit-mcus/sam-1>

- weitere Erweiterungen können entworfen und dem MESSI hinzugefügt werden, um andere Größen zu messen, um neue Fragestellungen zu beantworten
- Gehäuse teilweise 3D-gedruckt oder im Spritzgussverfahren hergestellt
- PCB-Design (eigens gefertigte Platine) ermöglicht
 - hohe Stückzahlen
 - kosteneffiziente Produktion
 - reduziert Fehler und Komplexität (plug-and-measure)
 - vereinfachter Zusammenbau durch wenige Komponenten (keine hohe Anzahl von Breakout-Boards)

3.2 MESSI: Sensorik

In Abb. 3 ist die Hauptplatine des MESSIs abgebildet. Hier befinden sich alle wichtigen Komponenten zur Datenerfassung, Abspeicherung und die Funkeinheit. Neben der Solarzelle zur Stromversorgung ist der Sensor zur Messung der Beleuchtungsstärke (E) von Silicon Labs angebracht (Si115x). Außerdem ist das GPS-Modul auf dieser Platine zu finden.

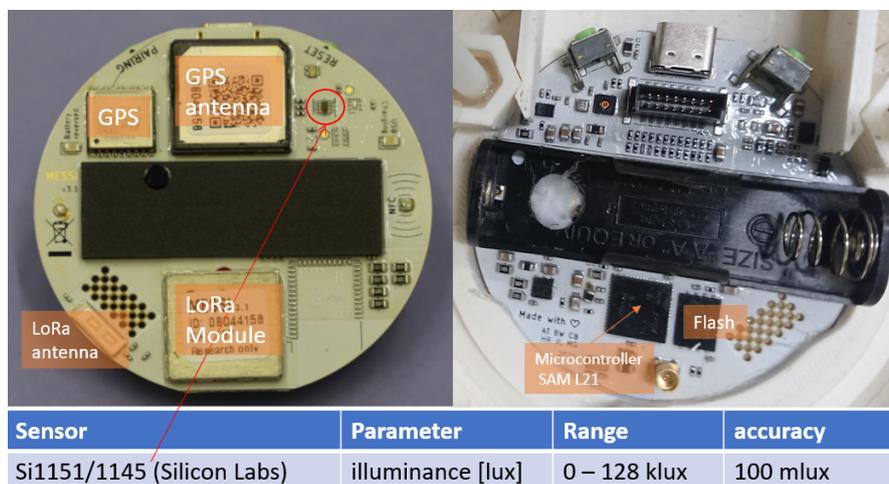


Abbildung 3: Hauptplatine des MESSIs mit Sensorik

Der modulare Aufbau beinhaltet auch eine weitere Platine zur Temperaturmessung

(siehe Abb. 4). Es können 2 Platin-Widerstandsthermometer (PT1000) mit dieser Platine ausgelesen werden. Der erste Messwiderstand ist im inneren des Gehäuses installiert (T_{in}) und besitzt eine geringe Trägheit (bedrahteter Dünnschicht-Sensor). Desweiteren befindet sich ein gekapselter PT1000 Messwiderstand außerhalb des Gehäuses (T_{out}), um eine Messung ohne den Einfluss des Gehäuses auf die Trägheit der Messung zu erhalten. Somit kann mit statistischen Methoden der Einfluss des nicht aktiv belüfteten Gehäuses unter Nutzung der Information von T_{out} und E minimiert werden.

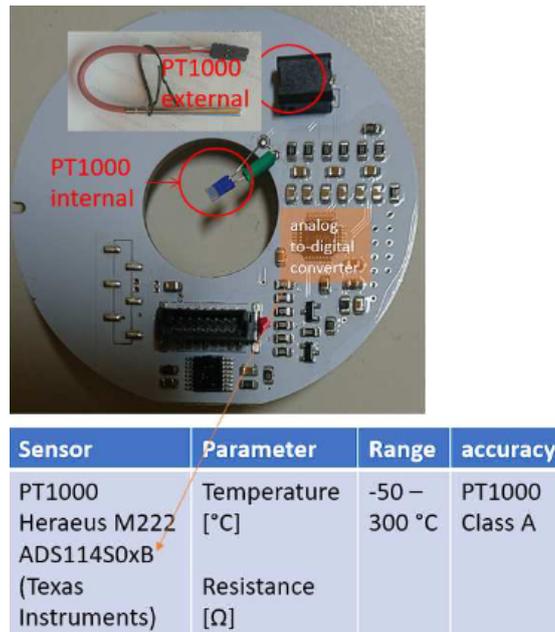


Abbildung 4: Erweiterung zur Temperaturmessung mit 2 PT1000 Thermometern

Auf einer dritten Platine (Abb. 5) sind 2 digitale Sensoren platziert und auch ein Anschluss für die Kippwaage zur Messung von Niederschlag. Zur Bestimmung des Luftdrucks (P) wird der Drucksensor **BMP38X** von Bosch Sensortec verwendet. Die Luftfeuchte (RF) und die digitale Temperatur (T_{dig}) wird mit dem Sensor **SHT31** von Sensirion erfasst. Die Messwerte des Feuchtesensors sind generell als plausibel einzustufen, weil durch eine Schutzmembran das Polymer des kapazitiven Sensors vor Staub und Schmutz geschützt ist. Außerdem ist die Platine im Gegensatz zur Hauptplatine vor Sonnenlicht abgeschirmt, sodass keine Drifts aufgrund einer abweichenden Sensortemperatur zu erwarten sind. Die Daten des digitalen Temperatursensors (T_{dig}) wurden im Jahr 2022 noch nicht verwendet (erst im Datensatz aus dem Jahr 2023 verfügbar).

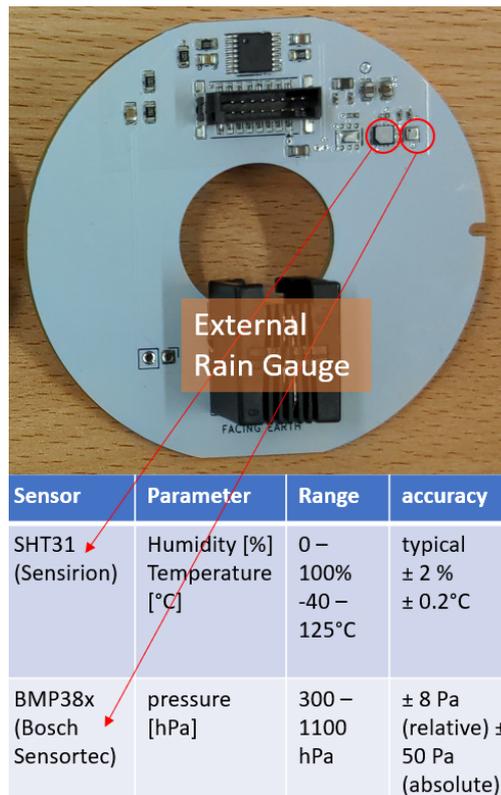


Abbildung 5: Erweiterung zum Anschluss einer Kippwaage zur Niederschlagsbestimmung und zur Messung von Druck, Feuchte und Temperatur

3.3 LoRain

Da im Zuge der Coronakrise bis weit in das Jahr 2022 hinein nur sehr beschränkt die für die *MESSI*-Produktion benötigten Bauteile verfügbar waren, wurde in dem Projekt zusätzlich die von *Pessl Instruments* hergestellte Wetterstation *LoRain* in das Messnetz integriert, um engmaschige Temperaturdaten erfassen zu können. Dieses Gerät misst Lufttemperatur, Luftfeuchtigkeit und mittels einer fest integrierten Niederschlagswippe die Niederschlagssumme. Die Daten werden zunächst in der Datenbank einer von *Pessl* angebotenen Webapp gespeichert, wurden zudem aber mit einer REST-API auf den Server der FU übertragen, auf dem auch die *MESSI*-Daten gespeichert wurden, damit alle Daten des Messnetzes gemeinsam betrachtet werden können.

3.4 Standorte

Die Standorte der MESSIs und LoRains sind weiträumig über die Stadt Berlin verteilt, eine Übersicht ist in Abb. 6 dargestellt. Hinweis: Im Jahr 2023 wurden deutlich mehr Stationen betrieben (98 Standorte in 2023 gegenüber 38 in 2022), insbesondere auch im Norden der Stadt.

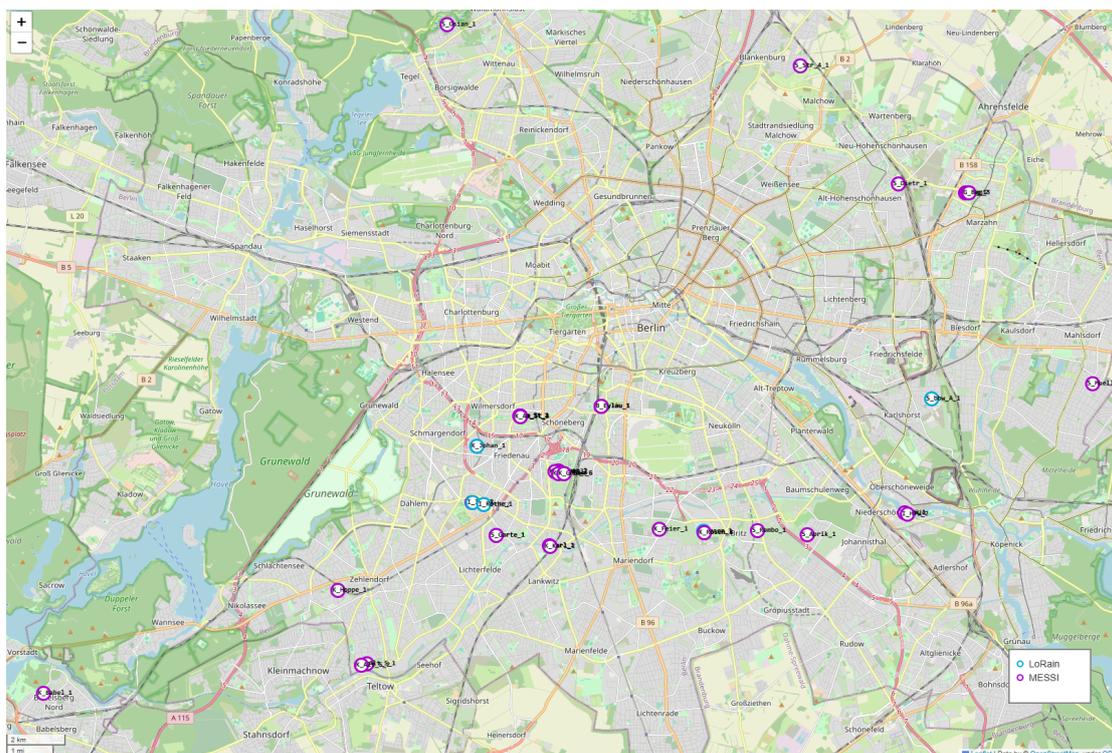


Abbildung 6: Standorte der MESSIs

Die Karte ist als interaktive HTML-Datei auch in den Metadaten verfügbar. Zusätzliche Informationen über die Standorte aller Stationen gibt es in der Datei `Stationen.csv`

Die jeweiligen Gerätestandorte wurden zu Beginn der Messung manuell notiert und nach dem Abbau überprüft, ob die aufgenommene Zeitreihe auf eine unerwünschte Bewegung des Geräts hindeutet, ggf. wurde die Zeitreihe dann gekürzt. Es ist möglich, dass an einem Standort verschiedene Messgeräte eingesetzt wurden, weil durch Ausfälle ein Auswechseln der Geräte nötig wurde. Es wurden allerdings nur Messungen von Geräten des gleichen Typs (*MESSI* oder *LoRain*) zu einem Standort zusammengefügt. Außerdem wurde der Beginn und das Ende der Messreihe visuell begutachtet, um einen plausiblen Messzeitraum zu bestimmen, auf den dann die in 4.1 und 5.1 beschriebene Qualitätsprüfung angewendet

wurde.

Bei den *MESSIs* wurden zudem noch die vom *MESSI* gespeicherten GPS-Daten wie in [Rust, 2024] beschrieben verwendet, um sicherzustellen, dass die Geräte während der Messperiode nicht vom vorgesehenen Standort bewegt wurden. Eventuell vorhandene Unsicherheiten zum Standort wurden in der Datei `Stationen.csv` beschrieben.

Da die Messstationen bauartbedingt ein unterschiedliches Messverhalten haben und das Postprocessing für die *MESSIs* deutlich umfangreicher war, werden nachfolgend die Mechanismen, die zu den veröffentlichten Daten geführt haben, für alle eingesetzten Gerätetypen einzeln beschrieben.

4 Beschreibung der Messdaten des *MESSIs*

Bei allen Variablen handelt es sich um Mittelwerte über eine Minute. Der *MESSI* führt intern alle 10 Sekunden eine Messung durch, die dann gemittelt werden. Der Zeitstempel ist in UTC angegeben und gibt das Ende des Mittelungsintervalls an. Die Messwerte des Zeitstempel 02:01 UTC beinhalten die Mittelwerte der Messungen von 02:00:10 UTC bis 02:01:00 UTC. Die Differenz der mitteleuropäischen Zeit (MEZ) zur koordinierten Weltzeit UTC beträgt +1 Stunde. Die Differenz der mitteleuropäischen Sommerzeit (MESZ) zur UTC beträgt +2 Stunden. In Tab. 1 ist eine Übersicht der gemessenen und abgeleiteten Messgrößen des *MESSIs* abgebildet.

Variablenname	Beschreibung	Einheit
T_{in}	Lufttemperatur innerhalb des Gehäuses	$^{\circ}C$
T_{out}	Lufttemperatur außerhalb des Gehäuses	$^{\circ}C$
T_{mod}	Lufttemperatur modelliert (s. Kap. 4.2.1)	$^{\circ}C$
T_{cor}	Lufttemperatur innerhalb des Gehäuses - Bias-korrigiert (s. Kap. 4.2.2)	$^{\circ}C$
P	Luftdruck in Höhe des <i>MESSIs</i>	hPa
P_0	Luftdruck auf Meeresniveau reduziert	hPa
RF	relative Feuchte	%
E	Beleuchtungsstärke	Lux
RR	Niederschlagssumme	mm

Tabelle 1: Übersicht über die gemessenen Variablen

4.1 Qualitätsmanagement *MESSI*

Bei einem bürgerwissenschaftlichen Projekt sind neben Abweichungen durch die preisgünstige Sensorik bzw. die günstige Konstruktion, auch Abweichungen von den Aufstellbedingungen durch die Teilnehmenden nicht auszuschließen. Möglicherweise wurden vereinzelt Vorgaben zum Aufstellort nicht eingehalten oder zeitweise Daten an nicht vorgesehenen Orten (z.B. im Geräteschuppen, Auto, etc.) erhoben.

Für die Sicherung der Datenqualität wurden einige automatische Methoden zur Datenüberprüfung entwickelt, um Probleme an den erhobenen Messdaten zu ermitteln. Auffälligkeiten in den Daten, beispielsweise Inkonsistenzen der verschiedenen Sensoren eines Gerätes oder Messungen, die sehr stark von Messungen an anderen Messstandorten abweichen, werden deswegen mit einem niedrigen Qualitätsniveau gekennzeichnet. Messungen, die alle Qualitätssicherungen erfolgreich bestanden haben, werden hingegen als vertrauenswürdig eingestuft und mit einem hohen Qualitätsniveau markiert. In Tab. 2 werden alle Qualitätsstufen (Flags) dargestellt, die nach Durchlauf der Prüfungen möglich sind.

Flag	Bedeutung	Beschreibung
g	gut	alle Qualitätsprüfungen bestanden
d	dubios	die automatische Prüfung hat den Messwert als fehlerhaft erkannt
p	potenziell dubios	reduzierter Luftdruck potenziell fehlerhaft, da ein fehlerhaft erkannter Temperaturwert in die Berechnung einging
m	manuell	auffälliger Wert, wurde manuell als unplausibel eingestuft
e	ersetzt	ersetzter Messwert (Beleuchtungsstärke in der Nacht auf 0 gesetzt)
f	fehlt	kein Messwert verfügbar

Tabelle 2: Übersicht über die Qualitätsstufen (Flags)

Bemerkung zum Flag „f“ Das Flag wurde gesetzt sobald ein Wert einer physikalischen Größe fehlte. Wenn etwa T_{in} fehlte, aber T_{out} verfügbar war, so wurde dennoch das Flag QT auf „f“ gesetzt. Wenn P_0 fehlte und P verfügbar war, wurde ebenso QP auf „f“ gesetzt.

Bemerkung zum Flag „e“ Teilweise kam es nachts zu fehlenden Werten der Beleuchtungsstärke. Da dies auch in der Justierung der Temperatur zu fehlenden Werten führt, wurden solche Fehlwerte durch 0 ersetzt. In der Nacht gemessene Werte der Beleuchtungsstärke liegen immer sehr nahe bei 0 und künstliches Licht würde keinen Strahlungsfehler verursachen.

Bemerkung zum Flag „m“ Zwei verschiedene Listen werden bei diesem manuellen Verfahren gepflegt. Eine sogenannte *Blacklist*, die Zeiträume angibt, in denen bestimmte Variablen unplausibel erscheinende Werte annehmen und manuell geflaggt werden („m“) sowie eine sogenannte *Whitelist*, die Zeiträume kennzeichnet die fälschlicherweise als dubios („d“) eingestuft wurden und wieder als plausibel („g“) eingestuft werden sollen. Das manuelle Flaggen wurde nur stichprobenartig bei auffälligen Werten durchgeführt bzw. dann, wenn unten beschriebene Kriterien auffälliger Werte gegeben waren. Bereits automatisch geflaggte Werte wurden nicht manuell geflaggt.

Bei der Messstation *MESSI* konnten für die Messgrößen Lufttemperatur, Luftdruck, Beleuchtungsstärke und Niederschlag Qualitätsflags ermittelt werden. Die Luftfeuchtigkeit wurde keiner Qualitätsprüfung unterzogen.

4.1.1 Lufttemperatur

Bei der Qualitätsprüfung der Temperaturwerte (QT) wurde zunächst die Differenz $\Delta T = T_{in} - T_{out}$ ermittelt. Da kurzzeitig größere Differenzen durch unterschiedliche Trägheit der Sensorik möglich sind, wurde ein gleitendes Mittel der ΔT -Werte über eine Stunde gebildet (zentriert). Nur wenn der Betrag dieses gleitenden Mittels am Tage größer als 3 K bzw. in der Nacht größer als 2 K war, wurde der Wert mit einem „d“ (dubios) gekennzeichnet und als auffällig eingestuft. Ein bekanntes Problem bei den *MESSIs* ist ein ungleichmäßiger Auftrag des Schutzlackes auf den Platinen. Gerade bei hoher Luftfeuchtigkeit kam es dann zu erkennbaren Temperaturdrifts durch eine Änderung der Widerstände einzelner Bauteile. Das Qualitätsflag QT bezieht sich auf alle Temperaturmessungen und insbesondere auch auf die in Kap. 4.2 beschriebene modellierte Temperatur T_{mod} und die bias-korrigierte Temperatur T_{cor} .

Neben der automatischen Detektion solcher Probleme wurden manuell-visuelle Begutachtungen durchgeführt, um eventuell nicht erkannte oder zu viel erkannte Datenfehler zu kennzeichnen. So wurden etwa starke Abweichungen von Referenzdaten des Deutschen Wetterdienstes oder starke Temperaturdifferenzen zweier aufeinanderfolgender Messungen kontrolliert. Durch eine visuelle Begutachtung kann dann festgestellt werden, ob die Werte herausgefiltert werden sollten oder plausibel sind, weil der Standort durch die mikroklimatischen Gegebenheiten diese Unterschiede verursacht. Diese Fehlerkontrolle wird mit einem Flag „m“ (manuell) gekennzeichnet.

4.1.2 Beleuchtungsstärke

Mit dem Sensor, der die Beleuchtungsstärke E misst, wurden weitere Qualitätsprüfungen (QE) durchgeführt. Es wurde geprüft, ob tagsüber unplausibel geringe Werte ($E < 200 \text{ lux}$) bzw. nachts unplausibel hohe Werte ($E > 1000 \text{ lux}$) auftraten (angegebene Schwellwerte wurden durch manuelle Begutachtung ermittelt). Dafür wurden die Zeitreihen zunächst in Tages- und Nachtzeiträume unterteilt, die mit dem Python-Paket **Astral** aus den Attributen für die Morgen- bzw. Abenddämmerung (dawn/dusk) bestimmt wurden⁴.

Diese Qualitätsprüfungen können z.B. bei einem undokumentierten Austausch der Station oder einer Reparatur anschlagen.

Wenn die automatische Qualitätsprüfung Unplausibilitäten bei der Beleuchtungsstärke aufzeigte, dann wird das QE -Flag und auch das QT -Flag auf „d“ gesetzt. Bei starken Gewittern am Tage wurde das QT -Flag teilweise fälschlicherweise auf „d“ gesetzt, dies wurde durch manuelle Einträge in der oben erwähnten Whitelist korrigiert.

Bei einigen Messstationen wurde der Sensor nicht richtig ausgelesen und es kam zu Messausfällen der Beleuchtungsstärke. Deswegen wurden fehlende Nachtwerte zum Teil auch ersetzt (Flag e), damit bei der modellierten Temperatur in der Nacht keine Fehlwerte auftreten. Generell wird immer eine Verfügbarkeitsquote an Messwerten der Beleuchtungsstärke berechnet. Nur wenn eine hohe Verfügbarkeit von mindestens 90% an Messwerten vorliegt wird dieser Sensor für die modellierte Temperatur genutzt (siehe Kap. 4.2).

⁴Nach manueller Begutachtung wurden dabei die Tagphasen durch einen 45-minütigen Puffer vor und nach der Dämmerung verkürzt - dies hat zu einer besseren Filterung fehlerhafter Werte geführt

4.1.3 Luftdruck

Um auch beim gemessenen Luftdruck P eine Qualitätsprüfung durchzuführen (QP), wurde zunächst aus P und der *geschätzten Höhe h der jeweiligen Messstationen über dem Meeresspiegel*⁵ der auf den Meeresspiegel reduzierte Luftdruck P_0 bestimmt⁶. Eine explorative Analyse der P_0 -Zeitreihen der einzelnen Stationen des Messnetzes hat ergeben, dass diese zumeist sehr ähnlich verlaufen und nur vereinzelt auftretende Sensorfehler zu stärkeren Abweichungen führen. Aus diesem Grund wurde zur Markierung dieser Fehlwerte ein geläufiges Verfahren zur Erkennung von sogenannten „Ausreißern“ angewendet: *Tukey’s Fence*. Dabei wurden zu jedem Messzeitpunkt t die Werte $P_0(t)$ aller Stationen betrachtet und alle Werte außerhalb des folgenden Intervalls als Ausreißer markiert

$$[Q_1 - k \cdot (Q_3 - Q_1), Q_3 + k \cdot (Q_3 - Q_1)], \quad (1)$$

wobei Q_1 und Q_3 das erste bzw. dritte Quartil der Messwerte bezeichnen. Die Konstante k nimmt im Normalfall den Wert $k = 1.5$ an und dieser Wert wurde auch hier verwendet. Dieses Vorgehen hatte allerdings noch eine Schwäche: Etwa bei Coldpools [Kirsch et al., 2021] die zum raschen Anstieg des Luftdrucks führen können, wurden die Stationen als Ausreißer markiert, die zuerst vom Coldpool erreicht werden. Um diesem Problem zu begegnen, wurde der gleitende Minimalwert und Maximalwert der Tukey-Fence-Intervallgrenzen über ein 25-Minuten-Zeitfenster verwendet, um die untere bzw. obere Grenze für die Markierung als Ausreißer zu bestimmen. In Abbildung 7 sieht man, wie am 09.07.2021 um ca. 14:00 UTC ein Coldpool zum Anstieg des Luftdrucks P_0 einiger Stationen führt. Die Station, deren Luftdruck als erstes ansteigt, würde bei der Anwendung von Tukey’s Fence pro Messzeitpunkt (rot) als Ausreißer markiert werden. Durch die Anwendung des gleitenden Minimal-/Maximalwertes (grün) jedoch bleibt die plausible Messung innerhalb des Toleranzbereiches. Im Datensatz werden sowohl P als auch P_0 veröffentlicht. Das Qualitätsflag QP bezieht sich auf beide Variablen. Immer dann, wenn das QP -Flag durch diese Prüfung auf **d** gesetzt wurde, fand zudem eine manuelle Prüfung der Temperatur statt und ggf. wurde das QT -Flag auf **m** gesetzt. Außerdem wurde das QP -Flag immer dann auf „p“ (potenziell dubios) gesetzt, wenn das

⁵Die Höhe h der Messstation über dem Meeresspiegel wurde wie folgt geschätzt: Zunächst wurden Höhen aus einem Digital Terrain Model vom OPENDEM-Server verwendet, um aus den Stationskoordinaten die jeweilige Höhe h_E des Erdbodens über dem Meeresspiegel zu bestimmen (Download-Link: https://www.opendem.info/downloads/srtm_germany_dtm.zip; Informationen bzgl. Auflösung: https://www.opendem.info/srtm_processing.html). Da den Projektteilnehmer:innen empfohlen wurde, die Station 2 m über dem Erdboden zu platzieren, ergibt sich die geschätzte Stationshöhe h über dem Meeresspiegel durch $h = h_E + 2m$.

⁶Es wurde die temperaturkorrigierte Variante der barometrischen Höhenformel verwendet:

$$P_0(h) = \frac{P(h)}{\exp(-g \cdot h / (R \cdot T_m))}$$

wobei h die Stationshöhe, g die Erdbeschleunigung, R die Gaskonstante und T_m die geschätzte mittlere Temperatur der Luftschicht bezeichnen. Die mittlere Temperatur der Luftschicht wurde durch die in Kap. 4.2 erklärte modellierte Temperatur T_{mod} geschätzt.

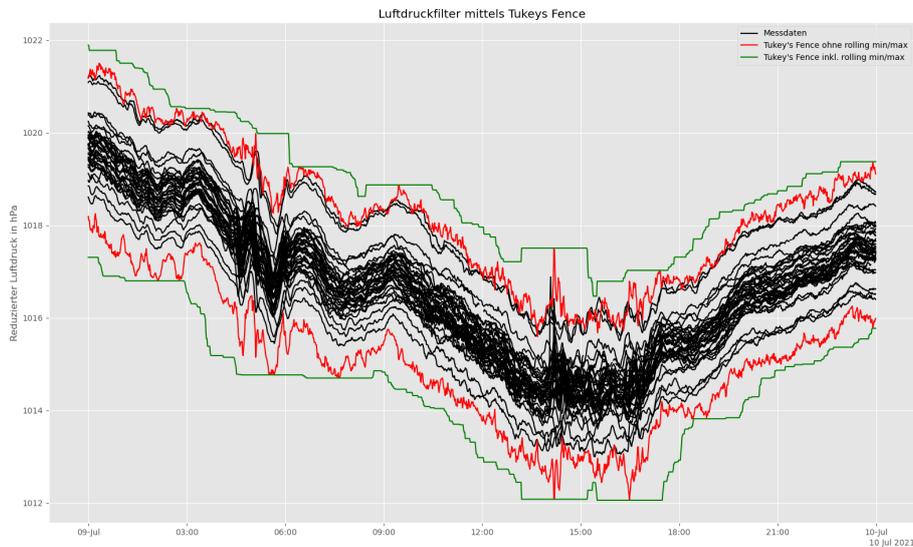


Abbildung 7: Luftdruckfilter

QT-Flag auf „d“ (dubios) gesetzt wurde, weil der reduzierte Luftdruck in diesem Fall mithilfe einer als fehlerhaft erkannten Temperatur berechnet wurde (dieses Flag bezieht sich also nur auf den reduzierten Luftdruck P_0).

4.1.4 Niederschlag

Auch für den Niederschlag wurde eine automatische Prüfung durchgeführt (*QRR*). Die MESSIs haben mit einer zeitlichen Auflösung von 1 Minute den Niederschlag gemessen. Eine Untersuchung der gemessenen Niederschlagswerte hat ergeben, dass in den Jahren 2022 und 2023 kein MESSI einen echten Niederschlag über 5 mm pro Minute gemessen hat. Alle Niederschlagsmessungen $> 5\text{mm}/\text{min}$ resultierten aus Wippenumschlägen wegen Bewegung der Geräte oder ähnlichen Fehlern. Es gab aber durchaus reale Messungen von $> 4\text{mm}/\text{min}$ bei starken Niederschlagsereignissen. Aus diesem Grund wurde für MESSIs ein Maximalwert von $5\text{mm}/\text{min}$ festgelegt. Wann immer dieser Wert überschritten wurde, wurde das *QRR*-Flag auf d gesetzt. Auch alle Messungen aus dem Zeitraum von 20 Minuten vor der Überschreitung bis 20 Minuten nach Überschreitung wurden mit *QRR* = d geflaggt.

Die Teilnehmer:innen wurden angewiesen die Niederschlagswippen vor "künstlichem Niederschlag" (Blumen gießen, Rasen sprengen) zu schützen. Dies wurde in Einzelfällen je-

doch nicht umgesetzt, was zu Fehlmessungen führen kann. Wenn "künstlicher Niederschlag" durch manuelle Begutachtungen und den Vergleich mit nahegelegenen Referenzstationen oder anderen Stationen des Messnetzes erkannt wurde, wurde QRR auf „ m “ gesetzt. Diese manuelle Prüfung konnte allerdings nicht vollumfänglich für alle Stationen vorgenommen werden.

Anders herum gibt es vereinzelt auch Fälle, in denen gefallener Niederschlag nicht erfasst werden konnte. Dies kann insbesondere dadurch verursacht werden, dass die Datenübertragung per LoRaWAN nicht möglich war. Die Messdaten wurden zwar auch auf dem internen Gerätespeicher gesichert, jedoch konnte dieser in Einzelfällen nicht nach der Messkampagne ausgelesen werden, so dass einige Datenlücken bestehen blieben. Dieses Problem betrifft zwar auch andere gemessene Variablen, jedoch kann beim Niederschlag nicht unterschieden werden, ob kein Niederschlag gemessen wurde oder ob ein Datenpaket nicht in die Forschungsdatenbank übertragen werden konnte - in beiden Fällen wird ein Niederschlag von 0 mm ausgewiesen. Dies hat den Hintergrund, dass bei der Datenübertragung mit LoRaWAN immer möglichst wenig Zeit bei der Übertragung der Daten verwendet werden sollte. Das spart Energie und reduziert mögliche Kollisionen mit anderen LoRaWAN-Geräten. Deswegen werden nur Niederschlagsdatenpakete versendet, wenn Niederschlag gemessen wurde.

4.1.5 Luftfeuchte

Für die Luftfeuchte wurden keine Qualitätsprüfungen durchgeführt.

4.2 Modellierung

Die für viele unserer Untersuchungen relevanteste Messgröße ist die Lufttemperatur. Aus diesem Grund veröffentlichen wir neben den zwei direkt gemessenen Temperaturen T_{in} und T_{out} noch zwei weitere Temperaturwerte: Eine mittels eines statistischen Modells adjustierte Temperatur T_{mod} und eine um den in einer Klimakammer ermittelten gerätespezifischen Bias korrigierte Temperatur T_{cor} . Diese adjustierte Temperatur T_{mod} wurde wie im Folgenden beschrieben bestimmt.

4.2.1 Modellierung T_{mod}

Neben einem möglicherweise inherent vorhandenen Bias der verwendeten Temperatursensoren ergeben sich eine Beeinflussung der Messwerte des T_{in} Sensors insbesondere durch die Tatsache, dass der Sensor auf einer Platine innerhalb eines nicht aktiv belüfteten

Plastik-Gehäuses angebracht ist. Dieses Gehäuse heizt sich, trotz der weißen Farbe, insbesondere bei starker Sonneneinstrahlung auf, was zu einer Erhöhung der T_{in} -Werte führt. Dieser Effekt tritt nach unserer Erwartung auf, wenn die vom *MESSI* gemessene Lichtstärke einen hohen Wert annimmt. Der Sensor außerhalb des Gehäuses (T_{out}) hingegen wird bei Niederschlag nass, was zu fehlerhaft reduzierten Messwerten dieses Sensors führen kann. Dieser Effekt tritt nach unserer Erwartung auf, wenn die Differenz zwischen T_{in} und T_{out} einen hohen Wert annimmt. Diese Überlegung und die Tatsache, dass wir vom Sensor T_{in} die bessere Approximation der „wahren Lufttemperatur“ erwarten, führten uns zu folgendem Modell, das wir mit der Programmiersprache R [R Core Team, 2021] und dem Paket `mgcv` [Wood, 2011] für generalisierte additive Modelle (GAM) trainiert haben (Modell A)

$$T_{ref} \sim \text{te}(T_{in}, \text{I}(E/1000)) + \text{s}(\text{I}(T_{out} - T_{in})) \quad (2)$$

Die Zielvariable `T_ref` aus obiger Formel stammt dabei von einem Datensatz einer belüfteten Messstation des Instituts für Meteorologie der Freien Universität Berlin mit minütlicher Auflösung (nicht final kontrolliert). Die Prädiktanten `T_in`, `T_out` und `E` sind die von einem *MESSI* gemessenen Werte der Innentemperatur, Außentemperatur bzw. Beleuchtungsstärke. Der Operator `te(.)` erzeugt ein volles Tensorprodukt mit direkten Effekten und Interaktionen; der Operator `s(.)` erzeugt einen Spline-Glättungsterm, siehe hierzu auch Wood [2017].

Dieser *MESSI* war bei der Messung direkt neben der belüfteten Station platziert. Mit diesen Daten wurde das Modell trainiert und dann auf alle anderen *MESSIs* angewendet, um einen nachjustierten modellierten Temperaturwert T_{mod} zu erhalten.

Das hier angegebene Modell berücksichtigt die oben getätigten Überlegungen zu Beeinflussungen der Messungen der Innen- und Außensensoren der Lufttemperatur und hat im Vergleich zu anderen linearen und generalisierten additiven Modellen unter einer 10-fachen Kreuzvalidierung den geringsten RMSE erzielt (0.41 K gegenüber 0.87 K bei den Rohdaten). Um dieses Modell zu testen, haben wir es auf einen Vergleichsdatensatz eines früheren Projektes angewendet, bei dem ebenfalls ein *MESSI* neben einer belüfteten Messstation stand. Hier liegt der RMSE des *MESSIs* gegenüber der belüfteten Station bei 0.64 K und der RMSE bei Anwendung der oben beschriebenen Justierung liegt bei 0.42 K. Leider weisen einige *MESSIs* zahlreiche Fehlwerte bei der Beleuchtungsstärke auf, was in obigem Modell zu Fehlwerten von T_{mod} führt. Aus diesem Grund haben wir für alle *MESSIs*, mit mehr als 10% fehlenden Werten der Beleuchtungsstärke ein Modell ohne Verwendung dieser Variable verwendet (Modell B)

$$T_{ref} \sim \text{s}(T_{in}) + \text{s}(\text{I}(T_{out} - T_{in})) \quad (3)$$

Das Modell ohne Verwendung der Beleuchtungsstärke hatte für die Trainingsdaten einen RMSE von 0.60 K, gegenüber einem RMSE von 0.87 K der Rohdaten. Beim Vergleichsfall aus dem früheren Projekt hatte dieses Modell einen RMSE von 0.42 K gegenüber einem RMSE von 0.64 K der Rohdaten.

In den einzelnen Datensätzen wird angegeben, ob die Modellierung mit (Modell A) oder ohne Verwendung der Beleuchtungsstärke (Modell B) durchgeführt wurde.

4.2.2 Modellierung T_{cor}

Um den gerätespezifischen Bias $B_{T_{in}}$ des Sensors zu ermitteln, wurden die Geräte über einen gewissen Zeitraum mit verschiedenen Einstellungen in eine Klimakammer gestellt und Temperaturwerte des Gerätes (T_{in}) und der Referenzstation (T_{ref}) der Klimakammer erfasst. Um Probleme durch Konvektionseffekte innerhalb der Kammer während der Kühl- und Heizphasen zu vermeiden, wurden diese Zeiträume ausgeschlossen. Der Bias $B_{T_{in}}$ wurde dann aus den mittleren Temperaturen des Gerätes (\bar{T}_{in}) und des Referenzsensors (\bar{T}_{ref}) während des Messzeitraumes ermittelt:

$$B_{T_{in}} = \bar{T}_{in} - \bar{T}_{ref}.$$

Die Temperatur T_{cor} errechnet sich schließlich durch

$$T_{cor} = T_{in} - B_{T_{in}}.$$

Für manche Geräte wurde diese Untersuchung nicht durchgeführt und T_{cor} ist nicht verfügbar.

5 Beschreibung der Messdaten der *LoRains*

Im Gegensatz zu den *MESSIs*, bei denen die Firmware von Mitarbeitern der FU Berlin geschrieben wurde und somit volle Kontrolle über die Mittelung bestand, gab es bei den *LoRains* nur begrenzte Möglichkeiten die zeitliche Auflösung und Mittelung zu beeinflussen. Es wurden verschiedene einstellbare zeitliche Auflösungen ausprobiert. In den meisten Fällen waren das 5 Minuten (die höchste verfügbare Auflösung), 10 Min oder 15 Minuten. In Einzelfällen war die Auflösung fehlerbedingt aber auch deutlich höher. Es wurden immer die aus der Field-Climate-App stammenden Messzeitpunkte unverändert übernommen. Die genaue Mittelungsprozedur ist nicht näher bekannt. Der Zeitstempel ist wieder in UTC angegeben. In Tab. 3 ist eine Übersicht der gemessenen Messgrößen des LoRain abgebildet.

Variablenname	Beschreibung	Einheit
T	Lufttemperatur innerhalb des Gehäuses	$^{\circ}C$
T_{cor}	Lufttemperatur innerhalb des Gehäuses - Bias-korrigiert (s. Kap. 5.2)	$^{\circ}C$
RF	relative Feuchte	%
RR	Niederschlagssumme	mm

Tabelle 3: Übersicht über die gemessenen Variablen

5.1 Qualitätsmanagement *LoRains*

Der einzige automatisiert Filter, der bei *Pessl*-Stationen angewandt wurde, betrifft den Niederschlag und resultiert aus dem Umgang mit fehlenden Werten seitens der Field-Climate-API. Wie oben beschrieben gab es für die *Pessl*-Stationen einige fehlerbedingte Datenlücken. Wenn es zu Datenlücken kommt, die auf Seiten der Field-Climate-API begründet sind, weist die Field-Climate-API den gesamten Niederschlag, der im betroffenen Zeitraum gemessen wurde, dem nächsten vorhandenen Messzeitpunkt zu. Auf diese Weise kann zumindest über einen großen Zeitraum die Niederschlagssumme konsistent erhoben werden ⁷. Unsere Messungen sind jedoch auf hohe zeitliche Auflösungen fokussiert und in dem beschriebenen Fall ist nicht nachvollziehbar, wann genau der Niederschlag gefallen ist, der dem Ende der Datenlücke zugewiesen wurde. Aus diesem Grund wurden bei *Pessl*-Stationen für alle Niederschlagsmessungen $> 0mm$ am Ende einer Datenlücke von mehr als 45 Minuten das *QRR*-Flag automatisch auf „d“ gesetzt.

Darüber hinaus wurde ein genereller visueller Plausibilitätscheck der Temperatur- und Niederschlagszeitreihen vorgenommen und die Zeitreihe ggf. gekürzt. Für ggf. aufgefallene offensichtliche Fehlmessungen wurde das *QRR*-Flag oder das *QT*-Flag auf „m“ gesetzt (manuelle Einschätzung als unplausibel, siehe Tab. 2).

5.2 Modellierung

Für die *LoRains* wurde keine statistische Modellierung wie in 4.2.1 für die *MESSIs* beschrieben durchgeführt. Die in 4.2.2 beschriebene Bias-Korrektur, wurde allerdings für die *LoRains* genau so wie für die *MESSIs* durchgeführt und auch hier mit T_{cor} bezeichnet.

⁷Es gibt auch Datenlücken, die nicht in der Field-Climate-API begründet sind, sondern andere Ursachen haben. In solchen Fällen enthält die Messung am Ende der Datenlücke nicht den Gesamtniederschlag, der während des Zeitraums der Datenlücke gefallen ist.

Literatur

- Bastian Kirsch, Felix Ament, and Cathy Hohenegger. Convective cold pools in long-term boundary layer mast observations. *Monthly Weather Review of the American Meteorological Society*, 149(3):811–820, 2021.
- Thomas Kox, Henning W Rust, Bianca Wentzel, Martin Göber, Christopher Böttcher, Jonas Lehmke, Elisabeth Freundl, and Matthias Garschagen. Build and measure: Students report weather impacts and collect weather data using self-built weather stations. *Australasian Journal of Disaster and Trauma Studies*, 25(3):79–86, 2021.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL <https://www.R-project.org/>.
- Henning Rust. Meteorologisches bürgermessnetz im landkreis oder-spreewald 2021. *Refubium - Repositorium der Freien Universität Berlin*, 2024. doi: 10.17169/refubium-42772. URL <https://refubium.fu-berlin.de/handle/fub188/43056>.
- Katrin Vohland, Anne Land-Zandstra, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, Roeland Samson, and Katherin Wagenknecht. The science of citizen science, 2021.
- Simon N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1):3–36, 2011.
- Simon N Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2 edition, 2017.