

# Meteorologisches Bürgermessnetz im Landkreis Oder-Spree 2021

Datenpublikation

C. Böttcher, V. Sobottke, H. W. Rust

Version von 5. April 2024

# 1 Messkampagne

Citizen Science, zu deutsch Bürgerwissenschaft, bezeichnet eine Form der wissenschaftlichen Forschung, bei der Bürgerinnen und Bürger aktiv an wissenschaftlichen Projekten teilnehmen [Vohland et al., 2021]. Die gesammelten Daten aus dem Citizen Science Teilprojekt einer Messkampagne von 2021 werden hier genauer beschrieben und zur Verfügung gestellt.

Das *meteorologische Bürgermessnetz* gehörte zu einer groß angelegten Messkampagne, die unter dem Akronym *FESSTVaL* (Field Experiment on submesoscale spatio-temporal variability in Lindenberg) [Hohenegger et al., 2023] <https://fesstval.de/> stattfand. Die Messkampagne wurde im Rahmen des vom Deutschen Wetterdienst (DWD) finanzierten *Hans-Ertel-Zentrums für Wetterforschung* (HErZ) [Simmer et al., 2016] <https://www.hans-ertel-zentrum.de/> initiiert. Sie fand in den Sommermonaten des Jahres 2021 am Meteorologischen Observatorium Lindenberg – Richard-Aßmann-Observatorium (MOL-RAO) des DWD südöstlich von Berlin statt. Um die Quellen sub-mesoskaliger Variabilität zu betrachten, konzentrierte sich die Messkampagne auf drei Hauptaspekte: Strukturen der atmosphärischen Grenzschicht, Cold Pools und Windböen.

Ergänzend zu dem am Observatorium operationell betriebenen Messnetz und weiteren Messungen mittels professioneller Instrumente und dem von der Universität Hamburg entwickelten, gebauten und gewarteten temporären Stationsmessnetz (APOLLO) wurde im Rahmen der Kampagne der Mehrwert eines bürgerwissenschaftlichen Messnetzwerks untersucht. Eine ausführliche Übersicht über die verschiedenen Instrumentierungen der Kampagne ist in Hohenegger et al. [2023] zu finden.

Die Geräte aus dem Bürgermessnetz wurden mit Technologien entwickelt, die das sogenannte „Internet-of-things“ mit sich bringt. Desweiteren wurden kostengünstige Sensoren verbaut, um möglichst viele Geräte herstellen zu können. Die Geräte wurden als Bausätze ausgeteilt, die von Bürger\*innen in wenigen Schritten zusammengebaut, aufgestellt und gewartet wurden. Für eine weitere Beschreibung der Geräte, siehe Abschnitt 3 und Kox et al. [2021].

## 2 Wissenschaftliche Fragestellung

Die sub-mesoskalige Variabilität bezieht sich auf meteorologische Phänomene, die auf einer räumlichen Skala kleiner als 2 Kilometer auftreten. Die mesoskalige Variabilität bezieht sich auf meteorologische Phänomene, die auf einer räumlichen Skala ab 2 Kilometer auftreten. Einzelne Gewitterzellen können Durchmesser von etwa 2 bis 10 Kilometern haben und gehören somit zu den kleinsten mesoskaligen Phänomenen. Die Erforschung mesoska-

liger und sub-mesoskaliger Variabilität ist aufgrund ihrer kleinen räumlichen Skala und schnellen zeitlichen Änderungen eine Herausforderung. Dennoch sind diese Phänomene von entscheidender Bedeutung für das Verständnis von lokalen Wetterereignissen.

Cold Pools sind spezifische Phänomene im Zusammenhang mit konvektiven Gewittern. Wenn Regen oder Hagel in einem Gewitter abregnen, verdunstet ein Teil des Niederschlags in den unteren Luftschichten. Dieser Verdunstungsprozess kühlt die Luft in der Nähe des Bodens ab. Diese kühlen Luftmassen sind dichter als die umgebende Luft und neigen dazu, sich am Boden auszubreiten und in Mulden oder Senken abzusinken. Diese Ausbreitung von kalter Luft wird als „Cold Pool“ bezeichnet. Ein illustratives Video zu diesem Phänomen gibt es auf <https://www.hans-ertel-zentrum.de/festval/cold-pools/index.html>.

Cold Pools können verschiedene Auswirkungen haben. Sie können die atmosphärische Stabilität beeinflussen und die Entwicklung neuer Gewitterzellen fördern oder hemmen. Die Kaltluft, die sich am Boden ausbreitet, kann auch Windböen verursachen und die lokalen Wetterbedingungen beeinflussen. Die Wechselwirkungen zwischen Cold Pools und anderen atmosphärischen Phänomenen sind komplex und können die Intensität und Entwicklung von Gewittern beeinflussen, siehe auch die Referenzen dazu in [Hohenegger et al. \[2023\]](#).

### 3 MESSI

Der MESSI (Mein eigenes Subskalen Messinstrument für atmosphärische Größen, siehe Abb. 1) ist eine kostengünstige Wetterstation für Citizen Science Projekte, die an der Freien Universität entworfen wurde<sup>1</sup>. Er misst Luftdruck, Luftfeuchtigkeit, Lufttemperatur (mittels zweier verschiedener Sensoren) und Beleuchtungsstärke.

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

**Aneignung** Die Teilnehmer\*innen sollen über den Prozess des Zusammenbauens einen Bezug zu dem Gerät entwickeln.

**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

---

<sup>1</sup>Der MESSI wurde von Jonas Lehmke entwickelt [jonas@lehmke.xyz](mailto:jonas@lehmke.xyz) und in weiten Teilen von der Firma Lehmke Prototype gefertigt



Abbildung 1: meteorologische Messstation MESSI

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 SAM L21<sup>2</sup> 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.

---

<sup>2</sup><https://www.microchip.com/en-us/products/microcontrollers-and-microprocessors/32-bit-mcus/sam-32-bit-mcus/sam-1>

Die oben beschriebenen Designanforderungen führen zu folgenden Hauptmerkmalen des MESSIs:

- unabhängige Energieversorgung mit Solarzelle und LiFePo4-Akku
- Funktechnologie LoRaWAN<sup>3</sup> 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
  - MESSI kann auch von den Bürger:innen selbst montiert werden, basierend auf einen Bausatz
  - Sensorplatinen können kostengünstig und einfach ausgetauscht werden
  - 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)

### 3.1 Standorte

Die Standorte der MESSIs sind um das Meteorologische Observatorium Lindenberg – Richard-Aßmann-Observatorium (MOL-RAO) des Deutschen Wetterdienstes südöstlich

---

<sup>3</sup><https://resources.lora-alliance.org/technical-specifications>

von Berlin verteilt. Eine Übersicht ist in Abb. 2 dargestellt. Die Karte ist als interaktives HTML File auch in den Metadaten verfügbar. Zusätzliche Informationen über die Standorte aller Stationen gibt es in der Datei *Stationen.csv*

Die Koordinaten der Standorte wurden aus gemittelten GPS-Daten gewonnen, die der MESSI einmal pro Tag misst. Dabei wurden Start- und Endzeitpunkt der Messungen an den jeweiligen Standorten durch manuelle Begutachtung der GPS-Datenreihen und auch der Temperatur-Datenreihen festgelegt, um sicherzustellen, dass das Gerät während der Messung nicht vom jeweiligen Standort entfernt wurde. Da die GPS-Daten nur über LoRaWAN übertragen wurden, gab es insbesondere an Standorten mit schlechter LoRaWAN-Infrastruktur größere Lücken in den GPS-Datenreihen. In diesen Fällen kann eine gewisse Unsicherheit des angegebenen Standorts nicht vermieden werden. Um diesem Problem zu begegnen, wurde für solche Fälle zum einen ein entsprechender Hinweis in der Datei *Stationen.csv* eingefügt. Zum anderen dienen die in Kapitel 4.1 beschriebenen Datenfilter teilweise auch als automatische Detektoren ungetrackter Bewegungen der MESSIs.

## 3.2 Sensorik

In Abb. 3 ist die Sensorik des MESSIs auf der Hauptplatine abgebildet. Der Drucksensor und der Sensor zur Beleuchtungsstärke sind auf dieser Platine untergebracht, weil diese kaum temperaturabhängig sind (bzw. wird bei der Druckmessung der Einfluss der Temperatur durch interne Routinen des Sensors und der mitgelieferten Bibliothek ausreichend korrigiert).

Auch der Feuchtesensor befand sich bei dieser eingesetzten Konfiguration auf der Hauptplatine. Die Messwerte dieses kapazitiven Sensors werden aus zweierlei Gründen von uns als fragwürdig eingestuft: Zunächst ist die Messung der Luftfeuchtigkeit stark von der Temperatur abhängig. Hier wäre eine Platzierung auf der vor Hitzeentwicklung besser geschützten Erweiterungsplatine sinnvoller gewesen. Die Hauptplatine kann sich bei Sonnenschein mit der Solarzelle auf 50° C oder mehr erwärmen. Desweiteren war in der Zeit der Liefer- und Chipkrise, auch wegen der Coronapandemie, kein adäquater Sensor mit Schutzmembran lieferbar. Das Polymer des kapazitiven Sensors ohne Schutzmembran kann seine Eigenschaften z.B. durch Schmutz verändern, was zu einem Drift der Werte führt. Die Daten des Feuchtesensors werden trotz dieser Mängel veröffentlicht, jedoch gibt es keine Qualitätsprüfung, weil generell mit einer hohen Unsicherheit zu rechnen ist.

Der modulare Aufbau beinhaltet auch eine weitere Platine zur Temperaturmessung (siehe Abb. 4). Es ist einmal ein PT1000 Filmthermometer im inneren des Gehäuses installiert ( $T_{in}$ ). Außerdem befindet sich ein gekapseltes PT1000 außerhalb des Gehäuses ( $T_{out}$ ), um eine Messung ohne den Einfluss des Gehäuses auf die Trägheit der Messung zu erhalten.

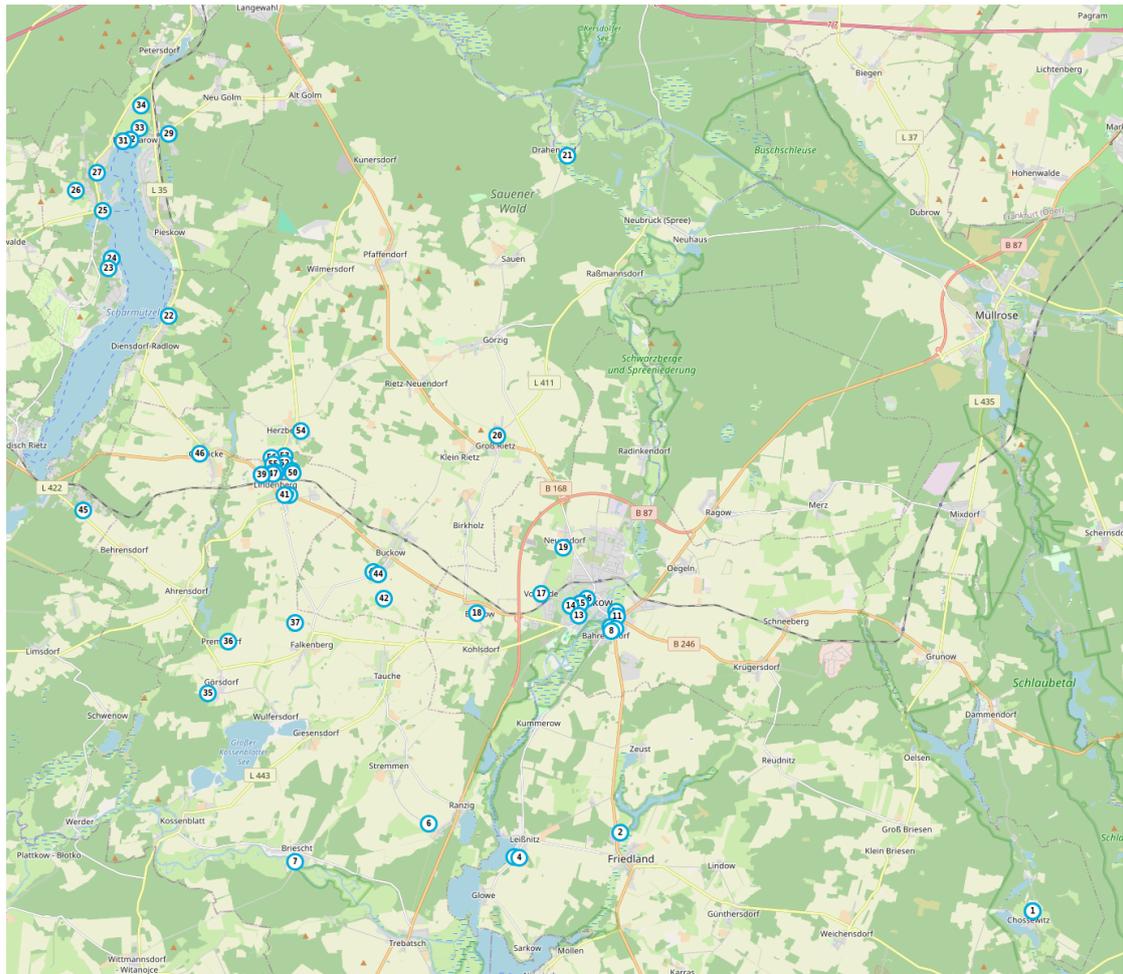


Abbildung 2: Standorte der MESSIs

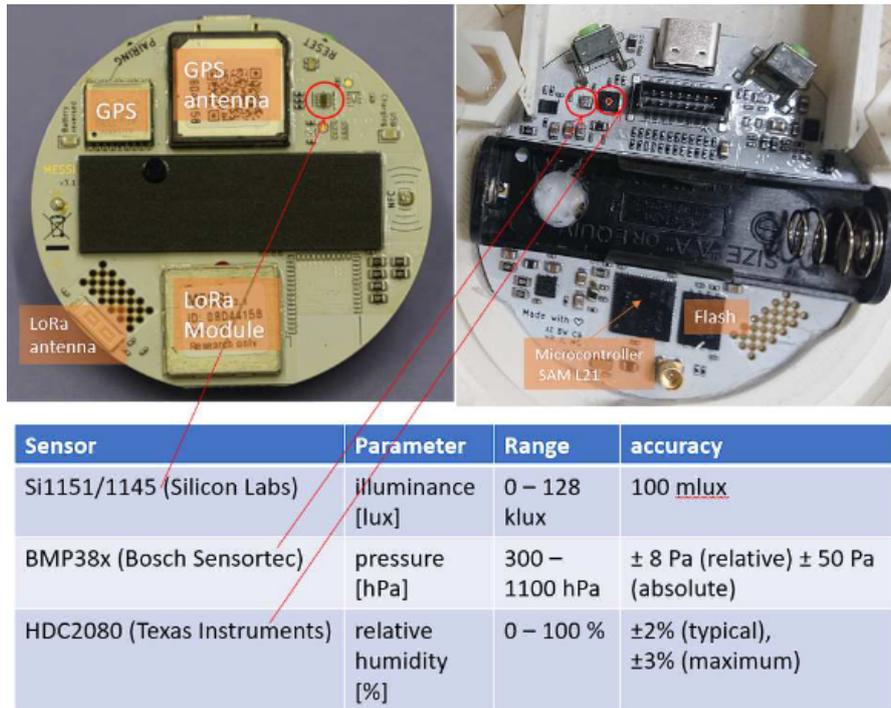


Abbildung 3: Hauptplatine des MESSIs mit Sensorik

Somit kann mit statistischen Methoden die Trägheit des Gehäuses unter Nutzung der Information von  $T_{out}$  und der Beleuchtungsstärke minimiert werden.

## 4 Daten

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 Tabelle 1 ist eine Übersicht der gemessenen und abgeleiteten Messgrößen des MESSIs abgebildet.

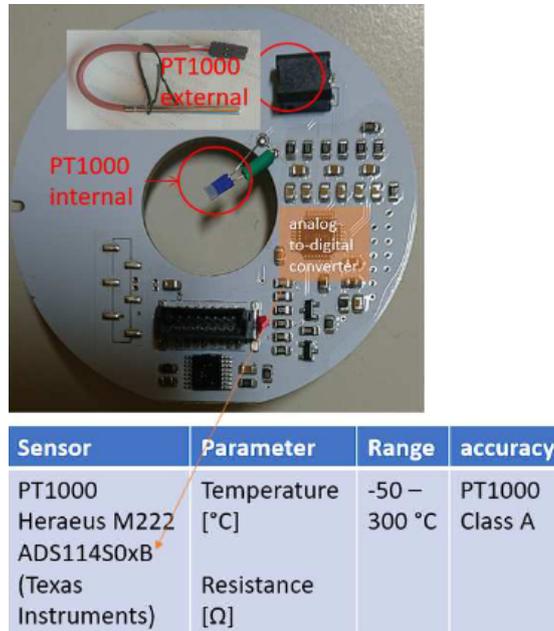


Abbildung 4: Erweiterung zur Temperaturmessung mit 2 PT1000 Thermometern

Variablenname	Beschreibung	Einheit
$T_{in}$	Lufttemperatur innerhalb des Gehäuses	°C
$T_{out}$	Lufttemperatur außerhalb des Gehäuses	°C
$T_{mod}$	Lufttemperatur modelliert	°C
$P$	Luftdruck in Höhe des MESSIs	hPa
$P_0$	Luftdruck auf Meeresniveau reduziert	hPa
$RF$	relative Feuchte	%
$E$	Beleuchtungsstärke	Lux

Tabelle 1: Übersicht über die gemessenen Variablen

## 4.1 Qualitätsmanagement

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
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 „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.

Bei der Messstation MESSI konnten für die Messgrößen Lufttemperatur, Luftdruck und Beleuchtungsstärke Qualitätsflags ermittelt werden. Die Luftfeuchtigkeit konnte wegen der bereits angedeuteten Probleme in Kap. 3.2 keiner Qualitätsprüfung unterzogen werden.

Bevor eine automatische Prüfung stattfinden konnte, wurden zunächst die eingesetzten Geräte den Standorten der Messkampagne zugeordnet. Es ist möglich, dass an einem Standort verschiedene Messgeräte eingesetzt wurden, weil durch Ausfälle ein Auswechseln

nötig wurde. Außerdem wurde der Beginn und das Ende der Messreihe visuell begutachtet, um einen plausiblen Messzeitraum zu bestimmen, auf den dann die automatische Qualitätsprüfung angewendet werden kann.

#### 4.1.1 Lufttemperatur

Zunächst wurde die Differenz  $\Delta T = T_{in} - T_{out}$  beider Sensoren ermittelt. Da kurzzeitig größere Differenzen durch unterschiedliche Trägheit der Sensorik möglich sind, wurde ein gleitendes Mittel der  $\Delta T$ -Werte über eine Stunde gebildet (zentriert). Nur wenn der Betrag dieses gleitenden Mittels größer 2.5 K ist, 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.

Neben der automatischen Detektion solcher Probleme wurden noch Temperaturdifferenzen  $\Delta T_{in} = T_{ref} - T_{in}$  bzw.  $\Delta T_{out} = T_{ref} - T_{out}$  zu professionellen Stationen ( $T_{ref}$ ) des Deutschen Wetterdienst berechnet. Bei Abweichungen mit  $|\Delta T_{in/out}| > 7 K$  wurden Ausgaben zur manuellen Prüfung erzeugt. 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 „m“ (manuell) gekennzeichnet. Zwei verschiedene Listen werden bei diesem manuellen Verfahren gepflegt. Eine sogenannte *Blacklist*, die Zeiträume angibt, die herausgefiltert werden und eine sogenannte *Whitelist*, die Zeiträume kennzeichnet die fälschlicherweise als dubios eingestuft wurden.

#### 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 ( $E < 200 lux$ ) bzw. nachts unplausibel hohe Werte ( $E > 1000 lux$ ) 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 wurden<sup>4</sup>.

---

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

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 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).

### 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*<sup>5</sup> der auf den Meeresspiegel reduzierte Luftdruck  $P_0$  bestimmt. 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 – in unserem Fall haben wir allerdings den Wert  $k = 2.5$  verwendet, weil dadurch weniger, durchaus noch plausible Messwerte, als fehlerhaft markiert wurden. Dieses Vorgehen hatte allerdings noch eine Schwäche: Etwa bei Coldpools, die zum raschen Anstieg des Luftdrucks führen, 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 5 sieht man,

---

<sup>5</sup>Die Höhe  $h$  der Messstation über dem Meeresspiegel wurde wie folgt geschätzt: Zunächst wurden Digital Terrain Model Daten 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](https://www.opendem.info/downloads/srtm_germany_dtm.zip); Informationen bzgl. Auflösung: [https://www.opendem.info/srtm\\_processing.html](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  $h = h_E + 2m$ .

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.

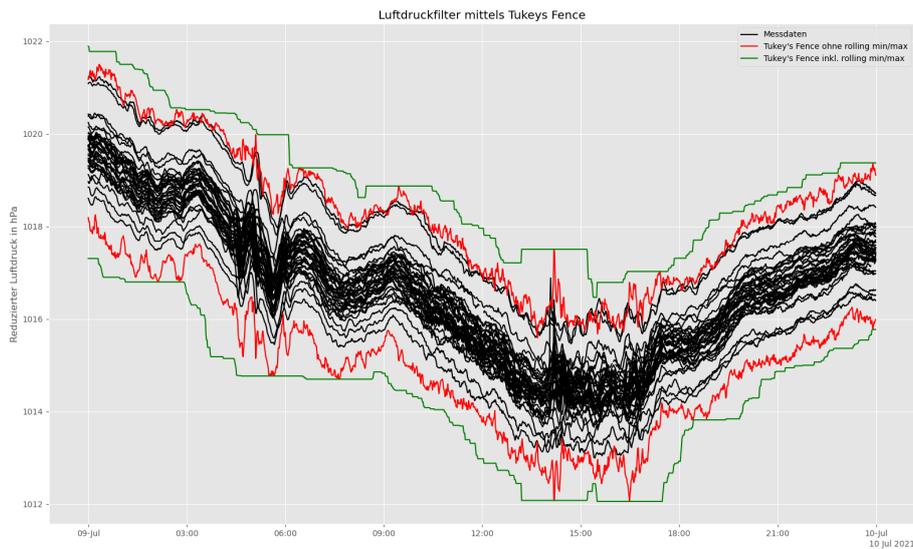


Abbildung 5: Luftdruckfilter

## 4.2 Modellierung

Die für unsere Untersuchungen relevanteste Messgröße ist die Lufttemperatur. Aus diesem Grund veröffentlichen wir neben den zwei direkt gemessenen Temperaturen  $T_{in}$  und  $T_{out}$  noch eine mittels eines statistischen Modells adjustierte Temperatur  $T_{mod}$ . Diese adjustierte Temperatur wurde wie folgt bestimmt.

Neben einem möglicherweise 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 DWD 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 DWD-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 Fehlmessungen 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 (0.24 K gegenüber 0.64 K bei den Rohdaten). An einer weiteren Position des Messnetzes (Messwiese Falkenberg) hatten wir ebenfalls die Situation, dass ein MESSI fast direkt neben einer belüfteten Messstation des DWD positioniert war. Hier liegt der RMSE des MESSIs gegenüber der DWD-Station bei 0.92 K und der RMSE bei Anwendung der oben beschriebenen Justierung liegt bei 0.53 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 0.39 K, gegenüber einem RMSE von 0.64 K der Rohdaten. Beim Vergleichsfall auf der Messwiese Falkenberg hatte dieses Modell einen RMSE von 0.51 K gegenüber einem RMSE von 0.92 K der Rohdaten.

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

## Literatur

- Cathy Hohenegger, Felix Ament, Frank Beyrich, Ulrich Löhnert, Henning Rust, Jens Bange, Tobias Böck, Christopher Böttcher, Jakob Boventer, Finn Burgemeister, et al. Fesstval: The field experiment on submesoscale spatio-temporal variability in lindenbergl. *Bulletin of the American Meteorological Society*, 104(10):E1875–E1892, 2023.
- 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/>.
- Clemens Simmer, Gerhard Adrian, Sarah Jones, Volkmar Wirth, Martin Göber, Cathy Hohenegger, Tijana Janjic, Jan Keller, Christian Ohlwein, Axel Seifert, et al. Herz: The german hans-ertel centre for weather research. *Bulletin of the American Meteorological Society*, 97(6):1057–1068, 2016.
- 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.