



## Bericht zur Dokumentation von Forschungsansätzen zur Erzeugung einer hochaufgelösten Böenklimatologie

Stanley, K. | Primo, C. | Brune, S. | Dalelane, C. | Mannig, B.

## **Zitiervorschlag**

Stanley K, Primo C, Brune S, Dalelane C, Mannig B (2024) Bericht zur Dokumentation von Forschungsansätzen zur Erzeugung einer hochaufgelösten Böenmeteorologie. Produktbericht des Schwerpunktthemas 101 (SPT-101) im Themenfeld 1 des BMDV-Expertennetzwerks.

## **Impressum**

Digitale Publikationsreihe des Themenfeldes 1 Klimawandelfolgen und Anpassung im BMDV-Expertennetzwerk „Wissen – Können – Handeln“ zu den Forschungsergebnissen (der Schwerpunktthemen) aus der 2. Forschungsphase (2020–2025).

Die Arbeiten des BMDV-Expertennetzwerk „Wissen – Können – Handeln“ wurden durch das Bundesministerium für Digitales und Verkehr (BMDV) gefördert. Internet: [BMDV-Expertennetzwerk](#)

## **Herausgeber**

Deutscher Wetterdienst (DWD)  
Frankfurter Straße 135  
63067 Offenbach am Main  
Tel: +49 (0) 69 / 8062 - 0  
E-Mail: [info@dwd.de](mailto:info@dwd.de)

## **Autorinnen und Autoren:**

Deutscher Wetterdienst:  
Kelly Stanley, Cristina Primo, Sebastian Brune, Clementine Dalelane, Birgit Mannig

**Redaktionsschluss:** 11 / 2024

**Titelbild:** Das Titelbild besteht aus Waben in verschiedenen Farben, sowie den Logos des BMDV-Expertennetzwerks und den beteiligten Behörden. Vier Fotos sind abgebildet: Eine Böenfront über Bettenhausen (Quelle: Reiner Müller); ein Regenschirm, der im Sturm weht (Quelle: Rainer Fuhrmann – stock.adobe.com); eine Gewitterfront in Schönbrunn bei Bamberg (Quelle: Anne Köhler); einem Sturm exponierte Bäume in Steinach/Thüringen (Quelle: Rüdiger Manig/ Deutscher Wetterdienst).



## Inhaltsverzeichnis

1	Kurzfassung .....	1
2	Hintergrund und Ziele .....	2
3	Daten .....	3
4	Methoden.....	5
5	Ergebnisse.....	7
6	Schlussfolgerungen und Ausblick .....	11
7	Literaturverzeichnis.....	iv
8	Anhang.....	v



# 1 Kurzfassung

Ein flächendeckender, hochaufgelöster Datensatz zur stündlichen Maximalwindgeschwindigkeit ist notwendig, um die Risiken durch extreme Windereignisse besser abschätzen zu können und zur Entwicklung von Anpassungsstrategien beizutragen. In Deutschland existiert nur eine begrenzte Anzahl von Stationen, die Windböen messen. Qualitativ hochwertige, gerasterte Beobachtungsdaten existieren bislang nicht für das Gebiet der Bundesrepublik Deutschland. Diese Datenlücke soll durch die Erstellung eines qualitativ hochwertigen Rasterdatensatz der maximalen stündlichen Windgeschwindigkeit in 10 m Höhe über der Landoberfläche geschlossen werden. Ziel ist ein Datensatz mit einer räumlichen Auflösung von ca. 1 km und einer zeitlichen Abdeckung ab 1995 bis zur Gegenwart.

Dafür wird ein auf Künstlicher Intelligenz (KI) basierender Ansatz nach der modifizierten Methode von Primo et al. (2024) verwendet. Mittels eines sogenannten Distributional Regression Network (DRN) werden verschiedene Variablen aus mehreren DWD-Datensätzen (COSMO-REA6, HOSTRADA, RADKLIM), topographische Daten sowie Landbedeckungsdaten verknüpft. DRN können komplexe, nicht-lineare Zusammenhänge zwischen verschiedenen Prädiktoren abbilden. Die Kombination dieser Datenquellen ermöglicht es dem DRN, optimale Prädiktoren für die Vorhersage maximaler stündlicher Windgeschwindigkeiten und deren Unsicherheit zu erlernen und auf bisher ungesehene Daten anzuwenden.

Die in diesem Bericht vorgestellten Ergebnisse stellen die besten derzeitigen Zwischenergebnisse dar. Diese Studie hat gezeigt, dass mittels eines DRN eine verbesserte Vorhersage der stündlichen maximalen Windgeschwindigkeiten möglich ist, insbesondere durch die Einbeziehung einer Vielzahl unterschiedlicher Prädiktoren. Die Korrelation der Vorhersagen mit den Stationsdaten wurde gegenüber den ursprünglichen COSMO-REA6-Daten erheblich verbessert, und die Fehlermaße, wie der Mittlere Absolute Fehler (MAE) und die Wurzel des mittleren quadratischen Fehlers (RMSE), konnten nahezu halbiert werden.

Um die Methode weiter zu evaluieren und zu optimieren, werden weitere Tests zur Neuronale Netze (NN)-Architektur und zu den Hyperparametern durchgeführt. Es wird geprüft, welche Prädiktorvariablen in der finalen Version des Produkts verwendet werden. Das Ziel ist, die Prädiktorvariablen auf eine Kernauswahl relevanter Variablen zu reduzieren, die die Vorhersage der Stärke von Windböen am besten repräsentieren. Dies ist notwendig, um die Verarbeitungszeit und die Rechenleistung zur Ausführung des Programms zu optimieren. Ein bedeutender Schritt in der Weiterentwicklung der Methode wird die geplante Integration des neuen COSMO-R6G2-Datensatzes sein, der ab Anfang 2025 verfügbar sein soll. Derzeit werden noch Daten aus COSMO-REA6 zur Entwicklung und Testung des DRN-Ansatzes verwendet, doch wird COSMO-R6G2 in Zukunft COSMO-REA6 ersetzen.

## 2 Hintergrund und Ziele

Durch ihre unvorhersehbaren und plötzlichen Kräfte stellen Windböen eine erhebliche Gefahr für Fahrzeuge, Bauwerke und die Verkehrsinfrastruktur dar. Solche Windereignisse können Schäden an Bäumen, Gebäuden und der Infrastruktur verursachen und den Straßen- sowie Schienenverkehr erheblich beeinträchtigen. Ein flächendeckender, hochaufgelöster Datensatz zur stündlichen Maximalwindgeschwindigkeit ist notwendig, um die Risiken durch extreme Windereignisse besser abschätzen zu können und zur Entwicklung von Anpassungsstrategien beizutragen.

Eine Windböe ist als das kurzzeitige Maximum der horizontalen Windgeschwindigkeit während einer längeren Messperiode definiert. Die Messungen erfolgen hauptsächlich mit Schalen- und Ultraschall-Anemometern, die idealerweise in einer Referenzhöhe von 10 m angebracht werden. In Deutschland existiert nur eine begrenzte Anzahl von Stationen, die Windböen messen. Bis zur Digitalisierung der Windmessungen in den 1990er Jahren war die Aufzeichnung solcher Daten stark limitiert, sodass kontinuierliche Böendatensätze meist bis maximal 1990 zurückreichen (Suomi und Vihma, 2018). Bei diesen Stationsdaten handelt es sich um punktuelle Messungen. Qualitativ hochwertige, gerasterte Beobachtungsdaten existieren bislang nicht für das Gebiet der Bundesrepublik Deutschland.

In Wetter- oder Klimamodellen wiederum sind die Prozesse zur Entstehung von Böen, wie Grenzschichtturbulenz und Konvektion in der Regel nicht aufgelöst und werden stattdessen durch Parametrisierungen abgebildet. Daher kann auch der bisher genutzte Reanalysedatensatz COSMO-REA6 mit einer räumlichen Auflösung von etwa 6 km das kleinskalige Phänomen der Windböen nicht adäquat abbilden.

Diese Datenlücke soll durch die Erstellung eines qualitativ hochwertigen Rasterdatensatz der maximalen stündlichen Windgeschwindigkeit in 10 m Höhe über der Landoberfläche geschlossen werden. Dafür wird ein auf künstlicher Intelligenz (KI) basierender Ansatz entwickelt. Mittels einer Verteilungsregression, also eines sogenannten Distributional Regression Networks (DRN), werden verschiedene Variablen aus mehreren Datensätzen des Deutschen Wetterdiensts (DWD) verknüpft. Ziel ist ein Datensatz stündlicher Maximalwindgeschwindigkeiten mit einer räumlichen Auflösung von ca. 1 km und einer zeitlichen Abdeckung ab 1995 bis zur Gegenwart.

### 3 Daten

Für die Modellierung der stündlichen Maximalwindgeschwindigkeiten in Deutschland wurden verschiedene meteorologische und topographische Daten sowie Landbedeckungsdaten integriert. Diese Datensätze umfassen sowohl punktuelle Beobachtungen als auch hochaufgelöste, flächendeckende Rasterinformationen und bieten eine solide Grundlage für die Erstellung einer präzisen Windböenklimatologie. Für eine vollständige Auflistung der in dieser Studie verwendeten Datensätze und Variablen, die zur Erzielung der in Abschnitt 5 dargestellten Ergebnisse geführt haben, siehe Tabelle A1 im Anhang. Wir werden noch weitere Tests bezüglich der im DRN verwendeten Variablen durchführen. Folgende Datensätze wurden für die hier vorgestellten Ergebnisse verwendet:

- **[Stationsmessungen der Windgeschwindigkeit](#)**: Diese enthalten stündliche Maximumwerte der Windgeschwindigkeit, welche von 275 Messstationen des DWD stammen (DWD, 2024a). Die standardisierte Messhöhe beträgt 10 m, variiert jedoch abhängig von den lokalen Bedingungen an einzelnen Stationen. Es ist noch nicht festgelegt, wie die Stationsmessungen für das Endprodukt ausgewählt werden, um im DRN als „Zielvariable“ zu dienen. Es gibt die Optionen entweder alle maximalen Windmessungen in Deutschland aus allen Messhöhen zu nutzen, oder alle Windböen-Messungen von 10 m +/- einer bestimmten Höhe. Weitere Tests sind erforderlich. Abschnitt 5 zeigt die Ergebnisse eines DRN-Modells, welches auf allen beobachteten maximalen Windgeschwindigkeiten in allen gemessenen Höhen als „Zielvariable“ trainiert wurde.
- **[COSMO-REA6](#)**: Der COSMO-REA6-Datensatz ist ein regionaler Reanalysedatensatz mit einer räumlichen Auflösung von etwa 6 km über Europa (EURO-CORDEX-Region) und deckt den Zeitraum von 1995 bis 2018 ab (DWD, 2019). Der Datensatz integriert Bodenbeobachtungen, Satellitendaten und Modellvorhersagen, um eine dreidimensionale Beschreibung des atmosphärischen Zustands zu liefern. Diese umfasst u.a. Temperatur-, Druck-, Wind- und Feuchtigkeitsfelder. Für diese Studie wurden die genutzten COSMO-REA6 Variablen gezielt ausgewählt, um Variablen auszuschließen, die sich auf die Bedingungen an der Erdoberfläche beziehen oder die mehrere Versionen gleicher oder ähnlicher meteorologischer Parameter darstellen.

Da die Produktion von COSMO-REA6 aufgrund der Einstellung des zugrunde liegenden globalen Datensatzes ERA-Interim nicht weitergeführt wird, wird COSMO-R6G2, die nächste Generation des Reanalysedatensatzes, für das Endprodukt eingesetzt. COSMO-R6G2 wird voraussichtlich den Zeitraum von 1995 bis zur Gegenwart abdecken. Es ist vorgesehen, die Anzahl der COSMO-R6G2-Variablen, die als Prädiktoren im DRN verwendet werden, auf eine Kernliste relevanter Windböen-Prädiktoren zu reduzieren. Diese Auswahl basiert auf Feature-Importance-Tests und der Eliminierung redundanter meteorologischer Variablen durch Feature-Selection-Tests.

- **[HOSTRADA](#)**: Der „Hochaufgelöste Stündliche Rasterdatensatz für Deutschland“ (HOSTRADA) Version 1.0 (1 km x 1 km) stellt seit 1995 bis heute eine Vielzahl meteorologischer Parameter zur Verfügung, darunter Temperatur, Globalstrahlung und mittlere Windgeschwindigkeit (DWD, 2024b). HOSTRADA basiert auf der Interpolation von Stationsdaten und integriert zusätzliche Informationen aus Satelliten- und Modelldaten.
- **[RADKLIM](#)**: RADKLIM Version 2017.002 bietet eine radargestützte Niederschlagsklimatologie für Deutschland mit einer räumlichen Gitterweite von 1 km x 1 km und Daten ab dem Jahr

2001 bis heute (DWD, 2018). Die Daten wurden anhand von Bodenmessungen kalibriert, um die Genauigkeit der Niederschlagsmengen zu verbessern.

- **GlobCover**: GlobCover ist ein Dienst der Europäischen Weltraumorganisation (ESA), der globale Komposite und Landbedeckungskarten erstellt (ESA, 2009). Die Landbedeckungsklassen aus GlobCover wurden in dieser Studie mit äquivalenten Rauigkeitslängen (m) abgeglichen, die aus Floors et al. (2021) übernommen wurden. Die Verwendung der GlobCover-Daten ermöglicht eine genaue Bestimmung der Rauigkeit der Landoberfläche, die eine entscheidende Rolle bei der Modellierung der Windgeschwindigkeiten in bodennahen Schichten spielt.
- **Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM)**: ASTER GDEM Version 2 wurde gemeinsam vom Ministerium für Wirtschaft und Industrie (METI) Japans und der US-amerikanischen Raumfahrtbehörde NASA entwickelt (NASA und METI, 2011). Diese topographischen Daten sind für die Anpassung der Windmodelle an die Geländehöhen in Deutschland von Bedeutung, um die lokalen Einflüsse der Topographie auf das Windfeld berücksichtigen zu können.
- **Tageszeit und Tag des Jahres**: Zur Analyse zyklischer Muster wurden „Tageszeit“ und „Tag des Jahres“ in kontinuierliche Werte umgewandelt. Die Stundenangabe der Tageszeit wurde mithilfe einer Sinus-Transformation in einen periodischen Wert umgewandelt. Der „Tag des Jahres“ wurde ebenfalls in einen Sinuswert überführt, wobei zwischen Schaltjahren (366 Tage) und normalen Jahren (365 Tage) unterschieden wurde. Diese Umwandlungen unterstützen die Untersuchung von täglichen und jährlichen Mustern in den Daten.

Die Kombination dieser Datenquellen ermöglicht es dem DRN, optimale Prädiktoren für die Vorhersage maximaler stündlicher Windgeschwindigkeiten zu erlernen und auf bisher ungesehene Daten anzuwenden. Diese Datenvielfalt verbessert die Fähigkeit des Modells, komplexe Wechselwirkungen zwischen atmosphärischen Bedingungen und der Landoberfläche zu berücksichtigen, und trägt damit zur Genauigkeit und Zuverlässigkeit der resultierenden Windböenmeteorologie bei.

## 4 Methoden

Für die Postprozessierung der Maximalen Windgeschwindigkeit aus Modellen existieren mehrere Ansätze. Der DWD verwendet seit vielen Jahren Model Output Statistics (MOS), basierend auf linearen Regressionen, um Unsicherheiten in Modellen zu korrigieren. Jüngste Studien (Schulz und Lerch, 2022; Primo et al., 2024) empfehlen die Anwendung von Methoden, die auf Neuronalen Netzen (NN) basieren, insbesondere eine Verteilungsregression (DRN). Diese Modelle können komplexe, nicht-lineare Zusammenhänge zwischen verschiedenen Prädiktoren abbilden.

In dieser Studie wurde ein DRN nach der modifizierten Methode von Primo et al. (2024) verwendet, um die maximale Windgeschwindigkeit vorherzusagen. Die stündlichen Vorhersagevariablen von Primo et al. (2024) wurden durch meteorologische Variablen aus den beschriebenen DWD-Datensätzen ersetzt. Das DRN wurde mit einer Kombination aus Stationsdaten und meteorologischen Informationen von benachbarten Gitterpunkten trainiert. Der Ansatz wurde angepasst, indem die Stations-Embedding-Funktion entfernt wurde und ein gemeinsames Netzwerk für alle Stationen über alle Zeitpunkte hinweg aufgebaut wurde. So kann für jeden Gitterpunkt die Zielvariable (maximale Windgeschwindigkeit innerhalb einer Stunde in 10 m Höhe) vorhergesagt werden. Die Prädiktorvariablen bestehen aus meteorologischen Datensätzen des DWD (COSMO-REA6, HOSTRADA, RADKLIM) sowie hochaufgelösten Geländemodellen, Landnutzungsdaten, die Tageszeit und den Tag des Jahres (siehe Datenabschnitt 3).

In diesem Abschnitt setzen wir grundlegende Kenntnisse über NN und die zugrunde liegende Terminologie voraus. Für ausführlichere Informationen zu Deep Learning verweisen wir auf Goodfellow et al. (2016). Das DRN verwendet eine Standard-NN-Architektur mit zwei Hidden Layers (ca. 50 und 25 Knoten) zwischen dem Input-Layer und dem Output-Layer. Die Trainingsprozedur basiert auf dem Adam-Algorithmus und optimiert die Continuous Ranked Probability Score (CRPS) Verlustfunktion. Die zugrunde liegende Verteilung im Output-Layer ist eine trunkierte logistische Verteilung, die davon ausgeht, dass die Windgeschwindigkeit streng positiv ist und über Verteilungen modelliert wird, die links bei Null trunkiert sind. Die Parameter dieser Verteilung werden so bestimmt, dass Aussagen über die wahrscheinlichste Windböe (Location Parameter -  $\mu$ ) sowie über die damit verbundene Unsicherheit (Scale Parameter -  $\sigma$ ) möglich sind. Eine Softplus-Aktivierungsfunktion wird auf die Output-Layer angewendet, um das Training zu stabilisieren und den Modus (Location Parameter) der trunkierten logistischen Verteilung auf positive Werte zu beschränken.

Die weiteren Hyperparametereinstellungen des verwendeten DRN-Modells umfassen eine Lernrate von 0,001, eine Batchgröße von 256, eine maximale Anzahl von 1000 Epochen für das Training sowie eine Early-Stopping-Strategie, die den Trainingsprozess beendet, wenn der Validierungsverlust über eine festgelegte Anzahl von Epochen (10) konstant bleibt, um ein Overfitting des Modells zu verhindern. Während des DRN-Modelltrainings wurden die beobachteten stündlichen maximalen Windgeschwindigkeiten der Stationen als „Sample Weight“ verwendet, um den stärksten Windböen im trainierten DRN-Modell ein höheres Gewicht zu geben. Geplant sind weitere Tests mit dieser „Sample Weight“-Funktion, bei denen beispielsweise die quadrierten beobachteten stündlichen maximalen Windgeschwindigkeiten als Input genutzt werden. Insgesamt berichteten Schulz und Lerch (2022), dass die Ergebnisse des DRN relativ robust gegenüber einer breiten Palette an Einstellungen der Hyperparameter sind und dass eine Erhöhung der Anzahl der Layer oder der Knotenanzahl innerhalb eines Layers die Vorhersagegenauigkeit nicht verbesserte. Das DRN-Modell ist mit den Python-Bibliotheken

Keras (Version 2.11.0; Chollet et al. 2015) und TensorFlow (Version 2.11.1; Abadi et al. 2016) implementiert und wurde auf dem Hochleistungsrechner (HPC) des DWD gerechnet.

Die Daten wurden in Trainings-, Validierungs- und Testdatensätze unterteilt:

- Trainingsdaten (2014–2016): Zur Anpassung der Modellparameter und zur ersten Kalibrierung.
- Validierungsdaten (2017): Zur Optimierung der Hyperparameter und zur Überwachung des Modelltrainings mittels Early-Stopping.
- Testdaten (2018): Zur abschließenden Überprüfung der Modellleistung und zur Validierung der generalisierten Vorhersagefähigkeit.

Die Datenvorverarbeitung umfasste die Normalisierung der Prädiktoren, um die Konvergenz des neuronalen Netzes zu verbessern, sowie die Identifikation und Behandlung von Ausreißern in den Stationsdaten. NN-basierte Vorhersagemodelle werden häufig mehrfach mit zufällig initialisierten Gewichten und Batches ausgeführt, um ein Ensemble von Vorhersagen zu erstellen und so der Zufälligkeit des Trainingsprozesses, der auf Stochastic Gradient Descent Methoden basiert, Rechnung zu tragen. Um diese inhärenten Unsicherheiten im Training des NN-Modells zu berücksichtigen und die Vorhersagegenauigkeit zu verbessern, folgen wir Schulz und Lerch (2022) und erstellen ein Ensemble von 10 DRN-Modellen zur Vorhersage von Windböen. Dieses Ensemble wird zu einer finalen Vorhersage zusammengefasst, indem die Verteilungsparameter gemittelt werden.

Zur Bewertung der Gesamtvorhersageleistung des DRN auf den Testdaten wurden proper scoring rules sowie andere Evaluationsmetriken herangezogen, wie der CRPS, der mittlere absolute Fehler (MAE), die Wurzel der mittleren quadratischen Abweichung (RMSE) und die Korrelation.

Um die Methode weiter zu evaluieren und zu optimieren, werden weitere Tests zur NN-Architektur und zu den Hyperparametern durchgeführt. Wir werden zukünftig prüfen, welche Prädiktorvariablen in der finalen Version des Produkts verwendet werden. Das Ziel ist, die Prädiktorvariablen auf eine Kernauswahl relevanter Variablen zu reduzieren, die die Vorhersage der Stärke von Windböen am besten repräsentieren. Dies ist notwendig, um die Verarbeitungszeit und die Rechenleistung zur Ausführung des Programms zu optimieren.

## 5 Ergebnisse

Wir haben zahlreiche Tests zum Training des DRN unter verschiedenen Einstellungen und Kombinationen von Prädiktorvariablen und Datensätzen durchgeführt und setzen diese fort. Hier werden die Ergebnisse des trainierten DRN-Modells gezeigt, das derzeit die besten Ergebnisse in den Testdaten zeigt und alle in Tabelle A1 (im Anhang) aufgeführten Prädiktorvariablen und Datensätze verwendet. Tests haben gezeigt, dass DRN-Modelle, welche mit allen Prädiktoren trainiert wurden, bessere Werte für den MAE und den RMSE erzielen, also solche, welche zum Beispiel nur mit HOSTRADA- oder COSMO-REA6-Daten trainiert wurden.

Die Zielvariablen des hier gezeigten DRN-Modells sind alle maximalen stündlichen Windgeschwindigkeiten, die an DWD-Wetterstationen gemessen wurden. Die verwendeten Windgeschwindigkeiten sind weder höhenkorrigiert noch wurden Messungen herausgefiltert, die auf anderen Messhöhen der Stationssensoren als 10 m gemessen wurden. Das DRN wurde auf Daten aus den Jahren 2014–2018 trainiert und getestet. Der Zeitraum 2014–16 wurde für das Training verwendet, das Jahr 2017 für die Validierung und das Jahr 2018 wurde als unabhängiger Testdatensatz genutzt. Den Messstationen wurden die Rasterdaten an den nächstgelegenen Gitterpunkten der Prädiktordatensätze zugeordnet.

Abb. 1 zeigt, dass das DRN die Darstellung der maximalen stündlichen Windgeschwindigkeit in 10 m Höhe (VMAX\_10M) im Vergleich zu COSMO-REA6 verbessert. Die obere Reihe in Abb. 1 zeigt die Histogramme aller stündlichen VMAX\_10M-Werte aus den Testdaten des Jahres 2018 und ihre statistischen Werte. Abb. 1(a) zeigt das Histogramm aller beobachteten maximale stündliche Windgeschwindigkeit im Jahr 2018 der Wetterstationen. Abb. 1(b) zeigt das Histogramm aller VMAX\_10M-Werte aus COSMO-REA6, die dem nächstgelegenen Gitterpunkt und Zeitschritt für die beobachteten maximale Windgeschwindigkeit-Werte im Jahr 2018 entnommen wurden. Abb. 1(c) zeigt das Histogramm aller vorhergesagten VMAX\_10M-Werte des trainierten DRN-Modells (Location Parameter). Insgesamt wird das Histogramm der beobachteten VMAX\_10M-Werte sowohl von COSMO-REA6 als auch vom DRN gut repräsentiert. Allerdings stellt das DRN die VMAX\_10M-Werte im höchsten Perzentil (>99 %), also extreme Windgeschwindigkeiten >35 m/s, sehr gut dar, wie an den Maximalwerten der Datensätze im Streudiagramm in Abb. 1(e) zu sehen ist. Abb. 1(d) zeigt, dass diese Werte im COSMO-REA6-Datensatz hingegen gar nicht repräsentiert sind.

Die Streudiagramme zwischen Stationsbeobachtungen und COSMO-REA6 (Abb. 1(d)) sowie den DRN-Vorhersagen (Abb. 1(e)) zeigen, dass die Genauigkeit der VMAX\_10M-Werte von COSMO-REA6 im Vergleich zu den DRN-Vorhersagen insgesamt geringer ist, wie die Fehlermaße MAE und RMSE belegen. Es wird deutlich, dass VMAX\_10M über Deutschland vom trainierten DRN sehr gut dargestellt werden kann, auch für nicht im Trainingsdatensatz enthaltene Zeiträume. Dies gilt insbesondere für VMAX\_10M >35 m/s, die von COSMO-REA6 nicht simuliert werden können. Für die Testdaten des Jahres 2018 lieferte das DRN deutlich bessere Vorhersagen der VMAX\_10M als COSMO-REA6. Im Vergleich zu den Fehlermaßen der COSMO-REA6-Testdaten (MAE: 1,67, RMSE: 2,29, Korrelation: 0,81) zeigt das DRN deutlich verbesserte Werte (MAE: 0,87, RMSE: 1,21, Korrelation: 0,95). Die DRN-Fehlermaße (MAE, RMSE, Korrelation) basieren ausschließlich auf den Location Parameter.

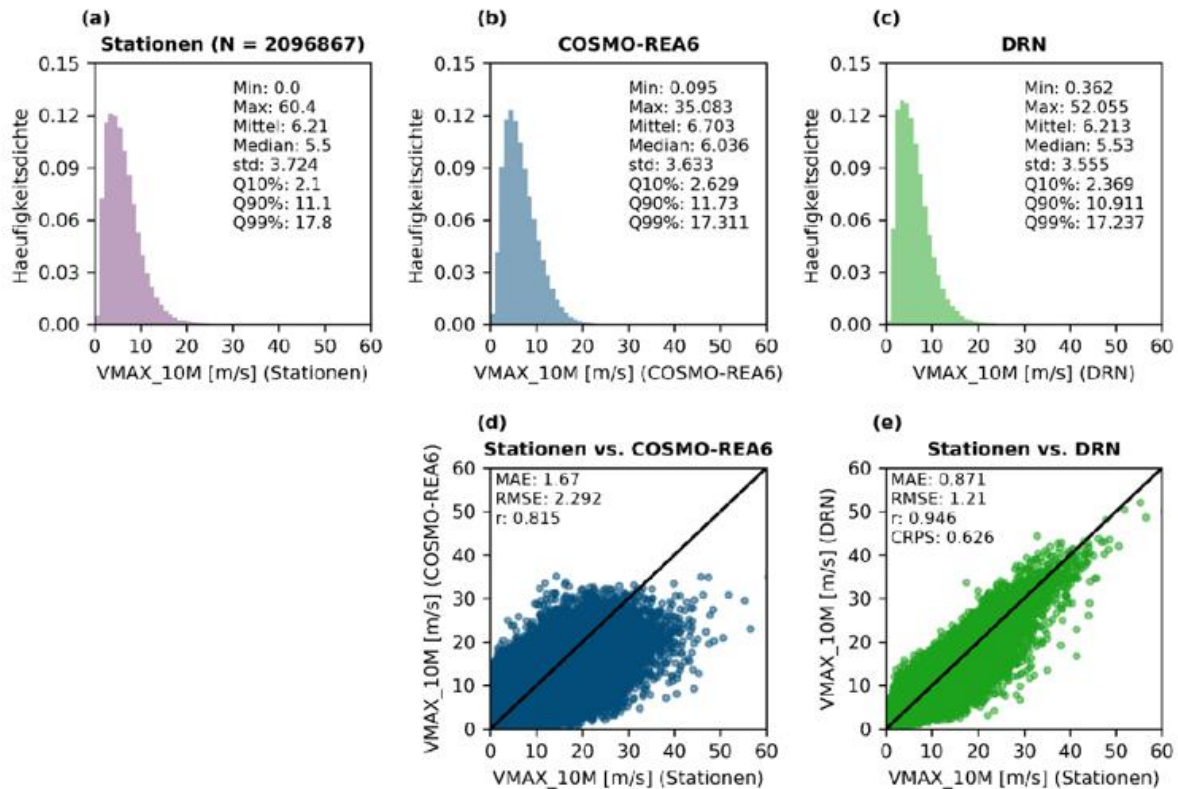


Abb. 1: DRN-Ergebnisse für die Testdaten (2018). Die obere Reihe zeigt die Histogramme und ihre statistischen Werte für die maximale stündliche Windgeschwindigkeit in 10 m Höhe (VMAX\_10M) im Jahr 2018 von (a) Stationen, (b) COSMO-REA6 und (c) DRN (Location Parameter). Min=Minimumwert, Max=Maximumwert, Mittel=Mittelwert, Median=Median, std=Standardabweichung, Q10%=10. Perzentilwert, Q90%=90. Perzentilwert, Q99%=99. Perzentilwert. Die untere Reihe zeigt die Streudiagramme für VMAX\_10M aus Stationsdaten im Jahr 2018 im Vergleich zu: (d) VMAX\_10M von COSMO-REA6 und (e) dem vom DRN vorhergesagten VMAX\_10M (Location Parameter). MAE=Mittlerer absoluter Fehler, RMSE=Wurzel des mittleren quadratischen Fehlers, r=Korrelation, CRPS=Continuous Ranked Probability Score.

Zusätzlich wurden die DRN-Ergebnisse während des Sturms Friederike im Januar 2018, der nicht im Trainingsdatensatz enthalten ist, genauer untersucht, um die Vorhersagegenauigkeit bei extremen Wetterereignissen zu überprüfen. Abb. 2(a) zeigt die Zugbahn von Sturm Friederike über Europa vom 17. bis 19. Januar 2018. Vom Sturmfeld an der Südflanke des Tiefs war insbesondere die Mitte Deutschlands betroffen, d.h. ein Streifen von Nordrhein-Westfalen bis Sachsen. Die dort gemessenen Spitzenböen zählen zu den höchsten, die im Vergleichszeitraum 1981 bis 2010 aufgetreten sind (Haeseler et al., 2018). In Norddeutschland war der Sturm dagegen kaum spürbar.

Abb. 2(b) zeigt die VMAX\_10M-Werte (Location Parameter) eines Zeitschrittes, die vom DRN-Modell während des Sturms Friederike am 18. Januar 2018 um 14 UTC über Deutschland mit einer Auflösung von  $0,1^\circ$  ( $\sim 10$  km) produziert wurden. Für die Analyse wurden zwei Wetterstationen entlang der Zugbahn der höchsten Windgeschwindigkeiten während des Sturms Friederike ausgewählt. In Abb. 2(b) wird die Lage der Station „Werl“ durch den gelben Kreis und die Lage der Station „Erfurt-Weimar“ durch das gelbe Dreieck dargestellt.

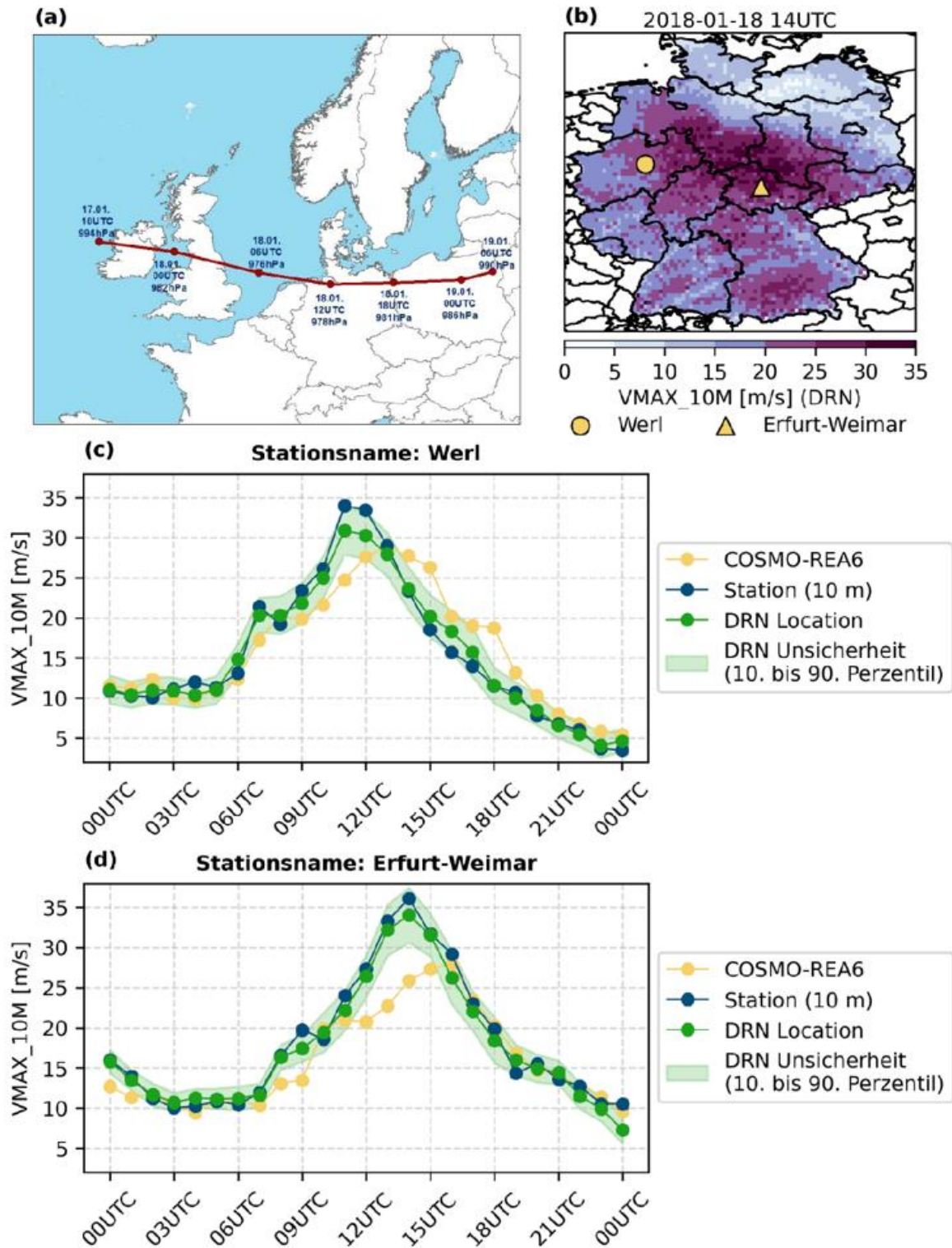


Abb. 2: (a) Zugbahn des Sturms Friederike über Europa vom 17. bis 19. Januar 2018 (Haeseler et al., 2018). (b) VMAX\_10M über Deutschland aus dem DRN (Location Parameter) während des Sturms Friederike am 18. Januar 2018, 14 UTC. 24-stündige Zeitreihenplots von VMAX\_10M für den 18. Januar 2018 zeigen COSMO-REA6 (gelbe Punkte), Stationsmessungen (blaue Punkte), den DRN Location Parameter (grüne Punkte) sowie das 80%-Konfidenzintervall um den DRN Location Parameter (grün schattiertes Band) für Werl (c) und Erfurt-Weimar (d).

Zur Quantifizierung der Unsicherheit in der DRN VMAX\_10M-Vorhersage haben wir das 10. und 90. Perzentil einer trunkierten logistischen Verteilung berechnet, parametrisiert durch den vorhergesagten Location ( $\mu$ ) und den Scale Parameter ( $\sigma$ ), um ein 80%-Konfidenzintervall der prognostizierten VMAX\_10M-Werte darzustellen. Um diese Unsicherheit zu visualisieren, wurde der Bereich zwischen dem 10. und 90. Perzentil als grün schattiertes Band um den Location Parameter in Abb. 2(c) und Abb. 2(d) abgebildet. Diese Darstellung veranschaulicht den wahrscheinlichen Wertebereich der VMAX\_10M und hebt das Konfidenzintervall des Modells um den zentralen Vorhersagewert ( $\mu$ ) hervor. Dieser Ansatz zur Unsicherheitsquantifizierung erwies sich als zuverlässig und die Werte stimmen gut mit den beobachteten VMAX\_10M-Werten überein, wodurch die inhärente Variabilität in den VMAX\_10M Vorhersagen erfasst wird.

Die Zeitreihen der stündlichen VMAX\_10M-Werte für den 18. Januar 2018, gemessen an den Wetterstationen und aus COSMO-REA6, sowie die DRN Location Parameter und das 80% Konfidenzintervall sind für Werl in Abb. 2(c) und für Erfurt-Weimar in Abb. 2(d) dargestellt. Die vorhergesagten VMAX\_10M-Werte (Location Parameter) des DRN stimmen gut mit den gemessenen Spitzenböen überein (Stationsmessungen sind meist innerhalb des 80%-Konfidenzintervalls des DRN) und zeigen eine verbesserte zeitliche Darstellung im Vergleich zu den COSMO-REA6-Ergebnissen. Insbesondere die Fähigkeit des DRN, lokal sehr hohe Böen zu erfassen, erwies sich als deutlicher Vorteil gegenüber dem COSMO-REA6-Ansatz.

## 6 Schlussfolgerungen und Ausblick

Diese Studie hat gezeigt, dass mittels eines DRN eine verbesserte Vorhersage der stündlichen maximalen Windgeschwindigkeiten möglich ist, insbesondere durch die Einbeziehung einer Vielzahl unterschiedlicher Prädiktoren. Das DRN liefert als Output die wahrscheinlichste Windböe sowie deren Vorhersage-Unsicherheit. Die Korrelation der DRN-Vorhersagen mit den Stationsdaten konnte im Vergleich zu den ursprünglichen COSMO-REA6-Daten erheblich verbessert und die Fehlermaße (MAE und RMSE) nahezu halbiert werden.

Die in diesem Bericht vorgestellten Ergebnisse stellen die besten derzeitigen Zwischenergebnisse dar. Um die Methode weiter zu evaluieren und optimieren, werden zusätzliche Tests durchgeführt. Voraussichtlich kann das DRN Modell durch weitere Anpassungen der Hyperparameter verbessert werden. Auch die Berücksichtigung eines längeren Zeitraums für das Training des DRN sollte zu weiteren Verbesserungen in der Vorhersage der stündlichen extremen Windgeschwindigkeiten führen. Zukünftige Arbeiten umfassen die Anpassung der Modellarchitektur und die Reduktion der Eingangsvariablen durch Feature-Importance-Analysen. Diese Analysen sollen helfen, die wichtigsten Prädiktoren für die Modellleistung zu identifizieren und das Modell dadurch effizienter zu gestalten.

Jeder der verwendeten Prädiktor-Datensätze hat dabei seine eigenen Stärken und Schwächen. Beispielsweise bietet COSMO-REA6 eine räumliche Auflösung von etwa 6 km sowie Daten in verschiedenen atmosphärischen Höhen. HOSTRADA hingegen liefert eine höhere räumliche Auflösung von 1 km, ist jedoch auf Daten an der Erdoberfläche beschränkt. Durch die Kombination dieser Prädiktor-Datensätze, die meteorologische und topographische Bedingungen über der Landoberfläche Deutschlands abbilden, kann das DRN die optimale Mischung von Prädiktoren erlernen und identifizieren. So wird es möglich, die Intensität der stündlichen maximalen Windgeschwindigkeiten auch in bisher ungesehenen Daten akkurat zu erfassen und vorherzusagen.

Ein bedeutender Schritt in der Weiterentwicklung der Methode wird die geplante Integration des neuen COSMO-R6G2-Datensatzes sein, der ab Anfang 2025 verfügbar sein soll. COSMO-R6G2 stellt eine zweite Generation der COSMO-REA6-Reanalyse dar und basiert auf dem aktuellen globalen Reanalysedatensatz ERA5. Die neue COSMO-Reanalyse soll den Zeitraum ab 1995 abdecken und jeweils bis zur Gegenwart aktualisiert werden. Derzeit werden Daten aus COSMO-REA6 zur Entwicklung und Testung des DRN-Ansatzes verwendet, doch wird COSMO-R6G2 in Zukunft COSMO-REA6 ersetzen. Die Verwendung von COSMO-R6G2 gewährleistet, dass der erstellte Windböen-Datensatz auch über das Jahr 2025 hinaus kontinuierlich aktualisiert und weitergeführt werden kann.

Ziel des Projekts ist es, einen finalen Datensatz zu entwickeln, der die beste Repräsentation der stündlichen maximalen Windgeschwindigkeiten in Deutschland liefert, mit einer räumlichen Auflösung von ~1 km und einer zeitlichen Abdeckung von 1995 bis 2025.

## 7 Literaturverzeichnis

Abadi M, Barham P, Chen J, et al. (2016) TensorFlow: A System for Large-Scale Machine Learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16).

Chollet F, et al. (2015) Keras. <https://keras.io>

DWD (2019) COSMO-REA6 Regionale Reanalyse [Datensatz]. Climate Data Center (CDC). [https://opendata.dwd.de/climate\\_environment/REA/](https://opendata.dwd.de/climate_environment/REA/)

DWD (2018) RADKLIM – Radarbasierte Niederschlagsanalyse für Deutschland Version 2017.002 [Datensatz]. Climate Data Center (CDC). [https://opendata.dwd.de/climate\\_environment/CDC/help/landing\\_pages/doi\\_landingpage\\_RADKLIM\\_RW\\_V2017.002-de.html](https://opendata.dwd.de/climate_environment/CDC/help/landing_pages/doi_landingpage_RADKLIM_RW_V2017.002-de.html)

DWD (2024a) Stundenmaximum aus Stationsmessungen der Windgeschwindigkeit für Deutschland, v24.03 [Datensatz]. Climate Data Center (CDC). [https://opendata.dwd.de/climate\\_environment/CDC/observations\\_germany/climate/hourly/extreme\\_wind/](https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/hourly/extreme_wind/)

DWD (2024b) HOSTRADA - Hochaufgelöster Stündlicher Rasterdatensatz für Deutschland, Version 1.0 [Datensatz]. Climate Data Center (CDC). [https://opendata.dwd.de/climate\\_environment/CDC/grids\\_germany/hourly/hostrada/](https://opendata.dwd.de/climate_environment/CDC/grids_germany/hourly/hostrada/)

ESA (2009) GlobCover 2009: Global Land Cover Map. [https://due.esrin.esa.int/page\\_globcover.php](https://due.esrin.esa.int/page_globcover.php)

Floors R, Badger M, Troen I, Grogan K und Permien FH (2021) Satellite-based estimation of roughness lengths and displacement heights for wind resource modelling. Wind Energy Science 6(6), 1379–1400, <https://doi.org/10.5194/wes-6-1379-2021>

Goodfellow I, Bengio Y und Courville A (2016) Deep Learning. MIT Press, Cambridge.

Haeseler S, Lefebvre C, Bissolli P, Daßler J und Mamtimin B (2018) Sturmtief Friederike vom 18. Januar 2018 in Europa. Deutscher Wetterdienst (DWD), Offenbach. [https://www.dwd.de/DE/leistungen/besondereereignisse/stuerme/20180123\\_friederike\\_europa.pdf](https://www.dwd.de/DE/leistungen/besondereereignisse/stuerme/20180123_friederike_europa.pdf)

NASA und METI (2011) ASTER Global Digital Elevation Model Version 2. <https://asterweb.jpl.nasa.gov/gdem.asp>

Primo C, Schulz B, Lerch S und Hess R (2024) Comparison of model output statistics and neural networks to postprocess wind gusts. arXiv preprint, arXiv:2401.11896.

Schulz B und Lerch S (2022) Machine learning methods for postprocessing ensemble forecasts of wind gusts: A systematic comparison. Monthly Weather Review 150(1), 235–257, <https://doi.org/10.1175/MWR-D-21-0150.1>

Suomi I und Vihma T (2018) Wind gust measurement techniques—From traditional anemometry to new possibilities. Sensors 18(4), 1300, <https://doi.org/10.3390/s18041300>

## 8 Anhang

Tabelle A1: Datensätze und Variablen, die zum Training des DRN-Modells verwendet wurden, um die maximale stündliche Windgeschwindigkeit in 10 m Höhe vorherzusagen, deren Ergebnisse in Kapitel 5 dargestellt sind.

Parameter	Beschreibung	Einheit
<b>Stationsmessungen der Windgeschwindigkeit</b>		
FX_911	Stundenmaximum der Windgeschwindigkeit	m/s
<b>COSMO-REA6*</b>		
*Hinweis: COSMO-REA6 wird im Endprodukt durch Parameter von COSMO-R6G2 ersetzt		
CLCT	Gesamtbedeckungsgrad mit Wolken	%
DURSUN	Sonnenscheindauer	s
H_PBL	Höhe der planetaren Grenzschicht	m
HSURF	Höhe der Oberflächentopographie	m
LHFL_S	Latenter Wärmefluss am Boden	W/m <sup>2</sup>
LWU_S	Aufwärts gerichtete langwellige Strahlung an der Oberfläche	W/m <sup>2</sup>
PMSL	Auf Meeresspiegel reduzierter Bodendruck	Pa
PS	Unreduzierter Bodendruck	Pa
QV_2M	2 m spezifische Feuchte	kg/kg
QV_S	Spezifische Feuchte an der Oberfläche	kg/kg
RAIN_CON	Konvektiver Regen	kg/m <sup>2</sup>
RAIN_GSP	Skaliger Regen	kg/m <sup>2</sup>
RELHUM_2M	2 m relative Feuchte	%
SHFL_S	Sensibler Wärmefluss am Boden	W/m <sup>2</sup>
SNOW_CON	Konvektiver Schneefall	kg/m <sup>2</sup>
SNOW_GSP	Skaliger Schneefall	kg/m <sup>2</sup>
SOBS_RAD	Kurzwellige Strahlungsbilanz an der Oberfläche	W/m <sup>2</sup>
SWDIFDS_RAD	Diffuse abwärts gerichtete kurzwellige Strahlung an der Oberfläche	W/m <sup>2</sup>
SWDIRS_RAD	Direkte kurzwellige Strahlung an der Oberfläche	W/m <sup>2</sup>
T_2M	2 m Temperatur	K
T (L35/36/37/38/39/40)	Temperatur auf unterschiedlichen Höhenleveln	K
TKE (L35/36/37/38/39/40)	Turbulente kinetische Energie auf unterschiedlichen Höhenleveln	m <sup>2</sup> /s <sup>2</sup>
TOT_PRECIP	Gesamtniederschlag	kg/m <sup>2</sup>
TQV	Vertikal integrierter Wasserdampf	kg/m <sup>2</sup>
U_10M	Zonale 10 m Windgeschwindigkeit	m/s

## Anhang

Parameter	Beschreibung	Einheit
V_10M	Meridionale 10 m Windgeschwindigkeit	m/s
VMAX_10M	Maximum der 10 m Windgeschwindigkeit	m/s
WS_10M	10 m Windgeschwindigkeit: $WS_{10M} = \sqrt{U_{10M}^2 + V_{10M}^2}$	m/s
WS (L35/36/37/38/39/40)	Windgeschwindigkeit: $WS = \sqrt{U^2 + V^2}$ auf unterschiedlichen Höhenleveln	m/s
Z0	Oberflächenrauigkeit	m
Q (L35/36/37/38/39/40)	Spezifische Feuchte auf unterschiedlichen Höhenleveln	kg/kg
<b>HOSTRADA</b>		
clt	Bedeckungsgrad	octa
dem	Digitale Geländerhöhe	m
hurs	Relative Feuchte	%
ps	Luftdruck auf Stationshöhe	Pa
psl	Luftdruck reduziert auf Meereshöhe	Pa
rsds	Abwärts gerichtete kurzwellige Strahlung an der Oberfläche	W/m <sup>2</sup>
sfcWind	Oberflächennahe Mittlere Windgeschwindigkeit	m/s
tas	Temperatur in 2 m	C
tdew	Taupunkttemperatur	C
<b>RADKLIM</b>		
RR	Niederschlagsmenge	kg/m <sup>2</sup>
<b>ASTER GDEM</b>		
Z	Höhe über dem Meeresspiegel	m
<b>GlobCover</b>		
Z0	Oberflächenrauigkeit	m
<b>Tageszeit und Tag des Jahres</b>		
sin_time	Sinuswert der Tageszeit	-
sin_day	Sinuswert des Tages im Jahr	-

Tabelle A2: Abkürzungsverzeichnis.

ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
BMDV	Bundesministerium für Digitales und Verkehr
COSMO-REA6	Consortium for Small-Scale Modeling ReAnalysis, 6-km-Auflösung
COSMO-R6G2	Consortium for Small-Scale Modeling ReAnalysis, 6-km-Auflösung, Generation 2
CRPS	Continuous Ranked Probability Score
DRN	Distributional Regression Network
DWD	Deutscher Wetterdienst
ECMWF	Europäisches Zentrum für Mittelfristige Wettervorhersage
ERA-Interim	Globale atmosphärische Reanalyse vom ECMWF (Vorgänger von ERA-5)
ERA-5	Fünfte Generation der globalen atmosphärischen Reanalyse vom ECMWF
ESA	Europäisches Weltraumorganisation
EURO-CORDEX	Coordinated Downscaling Experiment - Europäische Domäne
GDEM	Global Digital Elevation Model
HOSTRADA	Hochauflöser Stündlicher Rasterdatensatz für Deutschland
HPC	Hochleistungsrechner
KI	Künstliche Intelligenz
MAE	Mittlere Absolute Fehler
METI	Ministerium für Wirtschaft und Industrie (Japan)
MOS	Model Output Statistics
NASA	US-Bundesbehörde für Raumfahrt und Flugwissenschaft
RADKLIM	Radarklimatologie für Deutschland
RMSE	Wurzel des mittleren quadratischen Fehlers
SPT	Schwerpunktthema
VMAX_10M	Maximalen stündliche Windgeschwindigkeit in 10m Höhe