Automatisiertes Clustering der Schallemissionen von Windenergieanlagen

Susanne Martens, Lukas Findeisen, Tobias Bohne und Raimund Rolfes

Leibniz Universität Hannover, Institut für Statik und Dynamik, ForWind, 30167 Hannover, Deutschland,

Email: s.martens@isd.uni-hannover.de

Einleitung

Der abgestrahlte Schall von Windenergieanlagen (WEA) und dessen Ausbreitung wird von den vorherrschenden atmosphärischen Bedingungen sowie den Betriebszuständen der Anlagen stark beeinflusst. Die drei Faktoren – Akustik, Atmosphäre, Windenergieanlage – stehen hierbei in Wechselwirkung zu einander. Die bisherigen Untersuchungen in Bezug auf diese Wechselwirkungen basieren auf einer manuellen Einordnung von Messdatensätzen, welche akustische, meteorologische und anlagenspezifische Größen enthalten. Mit diesem Ansatz können grundlegende Zusammenhänge in der WEA-Schallemission und -immission identifiziert werden. Für eine detaillierte Untersuchung sind jedoch automatisierte Methoden des Clusterings notwendig.

Der erste Schritt hierzu wird in diesem Beitrag präsentiert. Es wird eine Methode zum automatisierten Clustering von WEA-Schallemissionen vorgestellt und auf umfangreiche Datensätze angewandt. Anhand des k-Means Algorithmus werden akustische Daten in Abhängigkeit von Betriebsparametern in Gruppen mit ähnlichen Eigenschaften zusammengefasst. Hierbei werden zwei Anwendungsfälle, die sich auf die gemessenen Gesamtschalldruckpegel und Terzspektren beziehen, betrachtet.

Messungen und Datenselektion

Für die Untersuchungen werden Datensätze, welche im Rahmen einer Messkampagne in einem Bürgerwindpark in Schleswig-Holstein aufgenommen wurden, verwendet. Im Frühjahr/Sommer 2020 wurden für 14 Wochen akustische und meteorologische Daten in der Nähe einer WEA der 3 MW-Klasse aufgenommen sowie die SCADA-Daten der Anlage erfasst.

In diesem Beitrag wird der Clustering-Algorithmus lediglich auf die Anlagendaten der WEA sowie auf die gemessenen akustischen Daten in 178 m Entfernung zur Anlage angewandt. Diese Entfernung entspricht der Nabenhöhe zuzüglich des halben Rotordurchmessers und kommt dementsprechend der empfohlenen Messposition zur Bestimmung der Schallemission gleich [1]. Die Messposition sowie die Messumgebung sind in Abbildung 1 (links) gezeigt. Die Landschaft wird als flach und homogen eingestuft und ist überwiegend mit Gras bedeckt. Für die Aufnahme der akustischen Größen (hochauflösende Zeitsignale, Terzspektren und Gesamtschalldruckpegel) wurde eine autarke akustische Dauermessstation entwickelt, welche in Abbildung 1 (rechts) dargestellt ist. Der Schalldruckpegelmesser ist in einer Höhe von 1.70 m positioniert und ist zur Reduktion von windinduzierten Geräuschen am Mikrofon mit zwei Windschirmen und einem Nasenkonus ausgestattet. Die Stromversorgung erfolgt über ein Solarpanel sowie einer externen Batterieeinheit. Die Entwicklung der Windschirme wird in [2] beschrieben.

Für gute Ergebnisse im Clustering ist die Selektion der Datensätze essentiell. Es werden lediglich 'gesunde' und vollständige Datensätze mit dominanten WEA-Geräuschen ausgewählt. Letzteres wird über statistische und frequenzbezogene Kriterien sowie über einen ausreichenden Signal-zu-Rausch Abstand sichergestellt. Die gemessenen WEA-Geräusche liegen demnach über der linearen Regressionskurve der gemessenen Hintergrundgeräusche. Folglich wird davon ausgegangen, dass externe Quellen - wie windbedingte Vegetationsgeräusche und Geräusche von Tieren - die ausgewählten akustischen Datensätze nicht beeinflussen. Im Rahmen der Datenselektion wurde zudem darauf geachtet, dass die akustischen Aufnahmen nicht von benachbarten Anlagen gestört werden. Wie in Richtlinien [1] empfohlen, wurden für die Repräsentation der WEA-Schallemission lediglich Daten in Mitwindrichtung der Anlage ausgewählt. Weitere Informationen zu den Messungen und der Datenselektion sind in [3] zu finden.

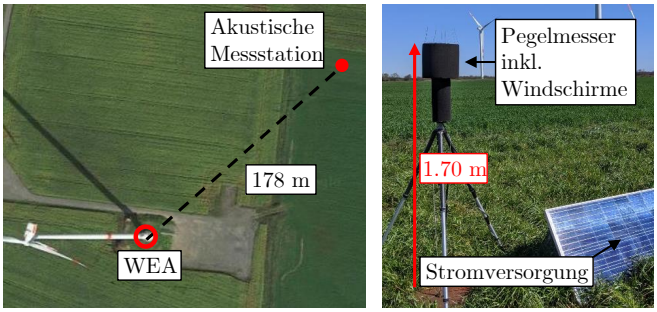

Abbildung 1: Messumgebung inkl. Messposition (links) und Beispiel einer autarken akustischen Dauermessstation (rechts)
Kartendaten©2021 GeoBasis-DE/BKG(©2019)
Bilder©2021 Aerodata International Surveys, GeoBasis-DE/BKG, GeoContent, Maxar Technologies

Methode des Clusterings

Um Datensätze mit ähnlichen Eigenschaften in Gruppen einzuteilen, wird in dieser Arbeit die k-means Clusteringmethode angewandt. Diese ist den partitionierenden Verfahren zugeordnet und gehört zu den am häufigsten angewandten Methoden. Mit dem Algorithmus werden Datensätze in K nicht überlappende Cluster mit ähnlichen Eigenschaften eingeteilt. Die Ähnlichkeiten werden hierbei anhand von Distanzmaßen bestimmt. Die einzelnen Cluster werden durch ihre Zentren C, auch Zentroide genannt, definiert.

Im Folgenden wird zunächst der allgemeine Clustering-Prozess kurz beschrieben. Anschließend wird auf das Verfahren zur Bestimmung der optimalen Clusteranzahl sowie zur Beurteilung der Clusterqualität eingegangen.

Der Prozess des k-means Clustering wird nach [4] folgend beschrieben:

1 Auswahl von K Anfangszentren (K wird in der Regel von dem Benutzer festgelegt, hier: optimale Anzahl über den prozentualen Anteil der erklärten Varianz)

2 Initialisierung der Startzentren

3 Jeder Datenpunkt D wird den nächstgelegenen Zentrum C zugeordnet, Datenpunkte mit gleichem Zentrum bilden ein Cluster

4 Schwerpunkt des Clusters (d.h. das Zentrum) wird auf Basis der zugehörigen Datenpunkte aktualisiert

5 Vorgang 3 - 4 wird wiederholt bis kein Datenpunkt mehr den Cluster wechselt oder Konvergenzkriterien (z.B. vorgegebene Anzahl an Iterationen) erreicht sind

Aus mathematischer Sicht wird der k-means Algorithmus über die Zielfunktion

$$\min_{\{C_k\},1\leq k\leq K} \sum_{k=1}^{K} \sum_{x\in S_k} \pi_x \mathrm{dist}(x, C_k) \tag{1}$$

mit

den Datenpunkten $D = \{x_1, ..., x_n\}$,
der Gewichtung der Punkte π_x,
dem Cluster S_k und
dem Zentrum des Clusters C_k

beschrieben. Die Funktion 'dist' berechnet den Abstand zwischen dem Datenpunkt x und dem Zentrum C_k. In dieser Arbeit wird als Entfernungsmaß der quadratische euklidische Abstand $||x - C||^2$ verwendet.

Zur Auswahl der optimalen Clusteranzahl wird der kumulative prozentuale Anteil der erklärten Varianz in Abhängigkeit der Clusteranzahl betrachtet. Dieser Anteil ergibt sich aus dem Verhältnis der erklärten Varianz eines Clusters zur Gesamtvarianz. Die Gesamtvarianz entspricht der kumulativen Summe der erklärten Varianzen aller einzelnen Cluster. [5]

In Abbildung 2 ist diese Abhängigkeit für das Anwendungsbeispiel I, welches sich auf das automatisierte Clustering der gemessenen Gesamtschalldruckpegel und Anlagenleistung bezieht, gezeigt. Für die Auswahl der Clusteranzahl wird das Kriterium festgelegt, dass der prozentuale Anteil der Varianz über 99% liegen muss. Die Clusteranzahl für das gegebene Beispiel ist 11. Mit einer höheren Anzahl werden keine besseren Ergebnisse erzielt. Mit einer sehr geringen Clusteranzahl sinkt der prozentuale Anteil deutlich.

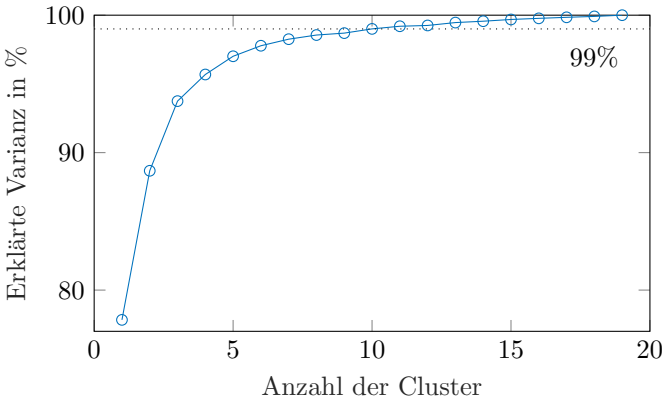

Abbildung 2: Kumulative prozentuale Anteil der erklärten Varianz in Abhängigkeit der Clusteranzahl

Die Qualität des Clusterings wird über die Silhouttenkoeffizienten bewertet. Die Silhouette wird für jeden Punkt x mit

$$S_x = \frac{(b_x - a_x)}{max(a_x, b_x)} \tag{2}$$

berechnet. Hierbei ist a_x der durchschnittliche Abstand des x-ten Punktes zu den anderen Punkten im selben Cluster und b_x repräsentiert den minimalen durchschnittlichen Abstand des x-ten Punktes zu Punkten in einem anderen Cluster. Wie auch im Clustering Prozess, wird für die Bestimmung von S_x der euklidische Abstand verwendet.

Der Silhouttenkoeffizient wird über

$$S_C = \frac{1}{n_c} \sum_{x\in C} S_x \tag{3}$$

bestimmt und ist das arithmetische Mittel der Silhouetten. Mit Gleichung 3 kann der Koeffizient sowohl für einzelne Cluster als auch für einen gesamten Datensatz ermittelt werden. Die Koeffizienten weisen Werte zwischen -1 und 1 auf. Je höher der Wert, desto besser passt der Punkt x zu dem eigenen und desto schlechter zu den benachbarten Clustern. Der Silhouettenkoeffizient gibt also Auskunft über die Strukturierung und ist somit ein Bewertungskriterium für Ergebnisse im Clustering. In Tabelle 1 sind die nach [6] empfohlenen Interpretationen zur Clusterstruktur in Abhängigkeit der Koeffizienten zusammengefasst. Koeffizienten über 0.71 weisen demnach auf eine starke Clusterstruktur und folglich auf gute Clusterergebnisse hin.

Tabelle 1: Interpretation des Silhouettenkoeffizienten [6]

Silhouettenkoeffizienten	Interpretationen
0.71 - 1.00	starke Struktur
0.51 - 0.70	sinnvolle Struktur
0.26 - 0.50	schwache Struktur, ggf. artifiziell
$\leq$ 0.25	keine Struktur

Ergebnisse

Die vorgestellte Clusteringmethode wird auf zwei Fälle angewandt. Im ersten Anwendungsfall wird der Gesamtschalldruckpegel in Abhängigkeit der Anlagenleistung geclustert. Im zweiten Anwendungsfall werden die gemessenen Terzspektren eines Clusters des ersten Falls gruppiert.

Für ein erfolgreiches Clustering ist eine Normierung der Datensätze notwendig. Die Anlagenleistung wird auf die Nennleistung der WEA normiert. Zur Normierung des gemessenen Gesamtschalldruckpegels werden die Emissionwerte aus einem Schallgutachten der WEA herangezogen. Der gemessene Gesamtschalldruckpegel wird auf den Immissionswert, der über die geometrische Ausbreitung auf Basis des Emissionswertes berechnet wurde, bezogen. Die gemessenen Terzspektren werden auf den jeweiligen maximalen Terzpegel normiert. Das Datenclustering inkl. der Normierung wird mittels der Software *Matlab* durchgeführt. In Tabelle 2 sind die gewählten Einstellungen zum Clustering zusammengefasst. Für weitere Informationen bzgl. der Einstellungen wird auf [7] verwiesen.

Tabelle 2: Einstellungen zum Clustering

Clustering Algorithmus	k-means
Anzahl der Cluster	Bestimmung durch %-Satz der Varianz
Abstandsmaß	quadratischer euklidischer Abstand
Max. Anzahl an Iterationen	100
Methode zur Auswahl der anfänglichen Clusterzentren	k-means++
Anzahl an Wdh. mit neuen anfänglichen Clusterzentren	10

Anwendungsfall I: Gesamtschalldruckpegel

Anhand des k-mean Algorithmus wird der normierte Gesamtschalldruckpegel in Abhängigkeit der normierten Anlagenleistung in 11 Cluster eingeteilt. Diese Einteilung ist in Abbildung 3 dargestellt. In Abbildung 4 sind die gemittelten Silhouettenkoeffizienten der einzelnen Cluster gezeigt. Der Übersicht halber wird auf die Darstellung der Silhouettenkoeffizienten je Datenpunkt verzichtet.

Der gemittelte Silhouettenkoeffizient aller Cluster beträgt im Anwendungsfall I 0.66. Dies entspricht nach Tabelle 1 einer sinnvollen, mittel ausgeprägten Clusterstruktur. Der höchste Koeffizient wird für Cluster 1 (S=0.86) und der niedrigste für Cluster 2 (S=0.55) berechnet. Die beiden Cluster liegen im selben Bereich der normierten Anlagenleistung. Im Bereich der Nenndrehzahl, d.h. bei normierten Leistungen höher als 0.5, liegen die Koeffizienten über 0.6. Folglich lassen sich die Datenpunkte in diesem Bereich gut den Clustern zu ordnen. Cluster 10 hat beispielsweise einen Koeffizienten von 0.65 und beinhaltet 484 Datenpunkte. Letzteres entspricht 7% des gesamten Datensatzes.

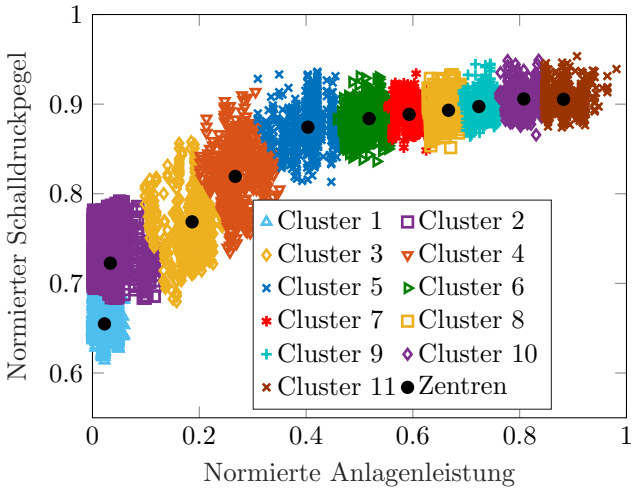

Abbildung 3: Anwendungsfall I: Einteilung der Gesamtschalldruckpegel in 11 Cluster

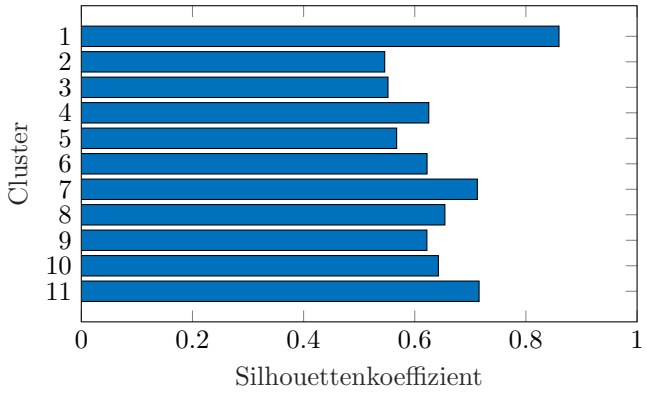

Abbildung 4: Anwendungsfall I: Berechneter Silhouettenkoeffizient je Cluster

Anwendungsfall II: Terzspektrum

Im zweiten Anwendungsfall werden die Terzspektren des 10. Clusters aus dem ersten Anwendungsfall betrachtet. Mittels des Clusterings werden zwei Gruppen identifiziert, dessen Terzspektren in Abbildung 5 aufgetragen sind. In Abbildung 6 sind die Silhouettenkoeffizienten der einzelnen Datenpunkt dargestellt.

Mit einem durchschnittlichen Silhouettenkoeffizienten von 0.77 wird die Clusterstruktur als stark ausgeprägt beurteilt. Der durchschnittliche Koeffizient des zweiten Clusters ist dabei mit 0.83 höher als der des ersten Clusters (0.71). Der zweite Cluster beinhaltet zudem eine höhere Anzahl an Datenpunkten. Etwa 65% der Daten werden dem zweiten Cluster zugeordnet.

Wie in Abbildung 5 zu erkennen, unterscheiden sich die Terzspektren der Cluster 1 und 2 leicht. Die Zentren der Cluster sind im Frequenzband verschoben. Der maximale Pegel des ersten Clusters liegt bei 315 Hz, während der des zweiten Clusters bei 500 Hz identifiziert wird. Im niedrigen Frequenzbereich sind die Pegel im Cluster 1 höher als im Cluster 2. Im hohen Frequenzbereich sind entgegen die Pegel des zweiten Clusters höher. Cluster 1 und 2 weisen demnach unterschiedliche Eigenschaften

auf, die auf betriebstechnische oder auch umgebungsbedingte Gründe zurückzuführen sein könnten. Beispielsweise könnte die WEA im Pitch betrieben werden oder die Turbulenzintensität in der Umgebung ist erhöht, sodass sich die Schallentstehung und -abstrahlung am Rotorblatt verändert.

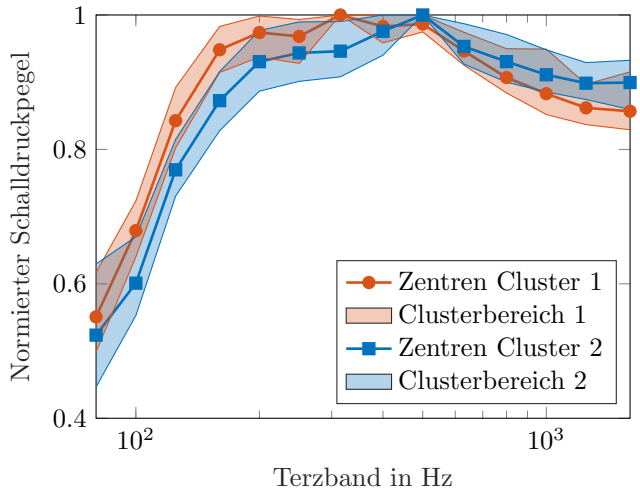

Abbildung 5: Anwendungsfall II: Einteilung der Terzspektren in zwei Cluster

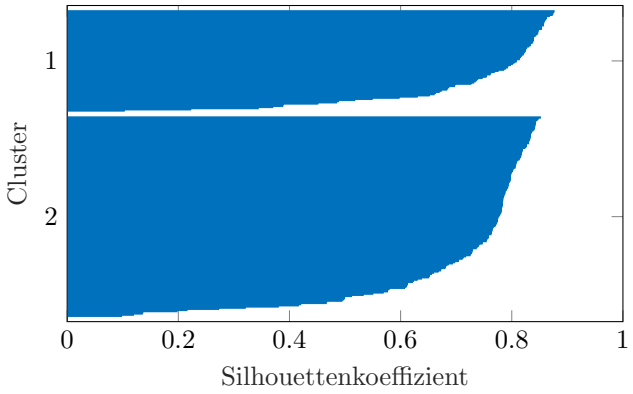

Abbildung 6: Anwendungsfall II: Berechneter Silhouettenkoeffizient je Datenpunkt

Schlussfolgerung und Ausblick

Automatisierte Clusteringmethoden sind für die Anwendung im Bereich des WEA-Schalls geeignet und bieten die Möglichkeit unbekannte Zusammenhänge in der WEA-Schallemission und –immission zu identifizieren. Der erste Schritt hierzu wird in diesem Beitrag vorgestellt. Es wird eine Methodik zum automatisierten Clustering präsentiert sowie auf zwei verschiedene Fälle angewandt. Anhand des k-means Clusterings wird der Gesamtschalldruckpegel in Abhängigkeit der Anlagenleistung in 11 Cluster eingeteilt. Mit einem gemittelten Silhoulettenkoeffizienten von 0.66 wird die Clusterstruktur als sinnvoll und mittel ausgeprägt eingestuft. Darüber hinaus werden die Terzspektren eines dieser 11 Cluster ebenfalls mittels einer Clusteranalyse betrachtet. Hierbei werden zwei Cluster mit leicht unterschiedlichen Verläufen der Terzspektren identifiziert, welche mit einem gemittelten Silhoulettenkoeffizienten von 0.77 eine starke Clusterstruktur aufweisen. Ohne das angewandte Clustering, d.h. mit manueller Zuordnung, wäre diese Identifikation nicht möglich gewesen.

In naher Zukunft werden weitere Methoden des automatisierten Clustering betrachtet und mit dem k-means Algorithmus verglichen. Vor- und Nachteile der einzelnen Methoden werden analysiert. Langfristig werden anhand eines ausgewählten Algorithmus die Zusammenhänge in der WEA-Schallemission und -immission untersucht.

Danksagung

Dieser Beitrag ist im Rahmen des Projektes „WEA-Akzeptanz“ entstanden, welches vom *Bundesministerium für Wirtschaft und Energie* (BMWi) finanziert wird. Darüber hinaus ist das Institut für Statik und Dynamik der Leibniz Universität Hannover Teil des Zentrums für Windenergie ForWind. Die Autoren möchten sich für die Unterstützung beim Zuwendungsgeber sowie beim Betreiber des Windparks *BÜRGERWINDPARK JANNEBY eG* bedanken. Weitere Informationen über das Projekt „WEA-Akzeptanz“ sind auf der Projekthomepage unter www.wea-akzeptanz.uni-hannover.de zu finden.

Literatur

[1] IEC 61400-11:2012/AMD1:2018 International Electrotechnical Commission: Wind turbines - Part 11: Acoustic noise measurement techniques.

[2] Martens, S., Boas, M., Bohne, T. und Rolfes, R. (2019): Towards the use of secondary windscreens to improve wind turbine sound measurements, Proceedings of 15th EAWE PhD Seminar on Wind Energy, Nantes, Frankreich

[3] Martens, S., Bohne, T. und Rolfes, R. (2020): An evaluation method for extensive wind turbine sound measurement data and its application. Proceedings of Meetings on Acoustics, Acoustical Society of America (41)

[4] Wu, J. (2012): Advances in K-means Clustering: A Data Mining Thinking, Springer-Verlag, Berlin Heidelberg.

[5] SPSS Statistics, Version 2020 URL: `https://www.ibm.com/docs/en/spss-statistics/24.0.0?topic=reduction-total-variance-explained`

[6] Kaufman, L. und Rousseeuw, P. J. (2005): Finding Groups in Data: An Introduction to Cluster Analysis, New York: Wiley.

[7] Matlab - Help Center, kmeans, URL: `https://www.mathworks.com/help/stats/kmeans.html`