FREIE UNIVERSITÄT BERLIN

DOCTORAL THESIS

An approach for the classification of European windstorms

Inaugural-Dissertation to obtain the academic degree Doctor rerum naturalium (Dr. rer. nat.)

submitted to the

Department of Earth Sciences

of Freie Universität Berlin

by

Christian Passow

May 16, 2022

1st reviewer:	Univ.-Prof. Dr. Uwe Ulbrich Institut für Meteorologie Freie Universität Berlin
2st reviewer:	Univ.-Prof. Dr. Henning Rust Institut für Meteorologie Freie Universität Berlin
Date of defense:	Berlin, 24. January 2023

Danksagung

Eine Arbeit von solch einem Ausmaß entsteht nicht über Nacht und auch nicht ohne die Unterstützung von vielen verschiedenen Menschen, seien es nun die Betreuer, Familie, Freunde oder Kollegen. Es liegt mir daher sehr am Herzen, meinen Dank und Anerkennung an all die Menschen auszudrücken, ohne die die Fertigstellung dieses Projekts nicht möglich gewesen wäre.

Mein erster Dank geht an meine beiden Betreuer Uwe Ulbrich und Henning Rust, die mir mit Rat zur Seite standen und mir die Freiheit ließen, meine Visionen umzusetzen. Ihr habt mir die Chance gegeben in diesem Projekt über mich hinaus zu wachsen, nicht nur als Wissenschaftler, sondern auch als Mensch. Im gleichen Atemzug danke ich auch all den anderen Kollegen in *CliDia* und *StatMet*. Vom wissenschaftlichen Austausch an der Kaffeemaschine, Vorträgen und Ratschlägen, bis hin zu gemeinsamen Dienstreisen, Mittagessen und Feierlichkeiten wart es immer ihr die dafür sorgten, dass sich dieses Institut in den letzten 10 Jahren wie ein zweites Zuhause angefühlt hat. Auch danke ich *NatRiskChange* für die Finanzierung und lehrreichen Kursen. Ich wünschte ich hätte die Gelegenheit gehabt jeden von euch besser kennen zu lernen.

Besonderer Dank geht an Oscar und Jana, die beide mit mir zusammen die beschwerliche Reise eines Doktoranden angetreten haben. Ohne euch wären die letzten Jahre triste und eintönig gewesen. Bei euch fand ich stets ein offenes Ohr, sei es nun wegen der Arbeit oder den endlosen Irrungen und Wirrungen meines Lebens. Ich hoffe ihr habt meine Gesellschaft so sehr genossen, wie ich die eure.

Auch möchte ich mich bei Reik Donner bedanken. Dies ist mein erstes großes Projekt bei dem du mir nicht als Co-Autor und Mentor zur Seite gestanden hast. Dein Einsatz und Interesse für die Wissenschaft hat mich nachhaltig zum besseren geprägt und ich bin stets bemüht meine Standards an den deinen zu messen.

Als letztes danke ich meiner Familie. Euer Rückhalt hat mir die notwendige Zuversicht gegeben, dass ich mich stets an euch wenden kann, was immer da auch kommen solle. Es ist unmöglich jahrelang einer Leidenschaft zu folgen, ohne solch ein Vertrauen und Unterstützung zu fühlen.

Selbstständigkeitserklärung

Hiermit erkläre ich an Eides Statt, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe angefertigt, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Berlin, May 16, 2022

Abstract

Winter windstorms are among the most severe natural hazards in Europe and are frequently the cause of enormous economic losses. Therefore, it is not surprising that studies on windstorms often revolve around their potential insured losses, return values or overall severity. Each of these measures is an important indicator for decision makers to understand the potential impacts of current and future windstorm events. Especially now, as evidence of an advancing climate change becomes more apparent. A much less researched but equally informative topic in this context are the characteristics of windstorms. Duration, size, and intensity are only a few examples of windstorm characteristics that are not only statistical properties of the event but also measures of severity themselves. In this dissertation, we study multiple European winter windstorms characteristics with the goal of identifying and understanding key parameters determining these characteristics, quantify their impact and investigate potential trends. Using the new ERA5 reanalysis product of the European Centre for Medium-Range Weather Forecasts, windstorms are tracked and matched to a parent cyclone. A set of windstorm characteristics is designed based on commonly evaluated characteristics from the windstorm and natural hazard community. Based on these characteristics, a set of windstorms is partitioned for the purpose of constructing objective windstorm classes, not only for further evaluation, but also in an effort to create universally applicable windstorm classes for a wide range of end-users. In the process, we introduce our newly developed quasi-supervised k-means (QSKM), a semi-supervised clustering technique for grouping windstorm events with respect to a catalog of historically severe windstorm events. QSKM constructs three different windstorm classes, one of which closely resembling the windstorm catalog in its inherent characteristics. In a comprehensive evaluation of the constructed classes, we show that large events with high wind speeds and an exceptionally long lifetime often origin in the West Atlantic near the US east coast, intensify over the mid and eastern parts of the Atlantic and usually hit Central or North Europe. Their occurrence can be associated with a strong jet stream and a deep parent cyclone. Affected areas experience strong wind gust between 15-20 hours, with peak wind speeds in the early 10 hours of occurrence. Comparable weaker and smaller events are found in the Mediterranean region. Those events often develop in the East Atlantic or Mediterranean Sea in which they also dissolve due to their short lifetime. However, due to their slow moving character, they often affect local areas up to 20 hours and more regardless of their short lifetime. Similar small and short events, but with high wind speeds are typical for northern Europe. These type of events usually origin in the western parts of the North Atlantic and further intensify as they travel across the open water. Although only exhibiting a third of the size the first class of events, the core pressure of their parent cyclone can also drop below 970 hPa.

Zusammenfassung

Stürme gehören zu den gefährlichsten Naturgefahren in Europa und sind häufig die Ursache enormer wirtschaftlicher Verluste. Daher ist es nicht verwunderlich, dass sich Studien über Stürme häufig um potenzielle versicherte Schäden, Wiederkehrwerte oder die Gesamtschwere der Ereignisse drehen. Jede dieser Größen ist ein wichtiger Indikator für Entscheidungsträger, um die potenziellen Auswirkungen aktueller und zukünftiger Stürme anzuschätzen. Vor allem jetzt, da die Anzeichen für einen fortschreitenden Klimawandel immer deutlicher werden. Ein weit weniger erforschtes, aber ebenso aufschlussreiches Thema in diesem Zusammenhang sind die Eigenschaften von Stürmen. Dauer, Größe und Intensität sind nur einige Beispiele für die Merkmale von Stürmen, die nicht nur statistische Eigenschaften des Ereignisses sind, sondern auch ein Maß für die Schwere des Ereignisses selbst. In dieser Dissertation untersuchen wir mehrere europäische Sturmcharakteristika mit dem Ziel, die Schlüsselparameter, die diese Charakteristika bestimmen, zu identifizieren und zu verstehen, ihre Auswirkungen zu quantifizieren und mögliche Trends zu untersuchen. Unter Verwendung des neuen ERA5 Reanalyze-Produkts des Europäischen Zentrums für mittelfristige Wettervorhersage werden die Stürme verfolgt und einem übergeordneten Zyklon zugeordnet. Eine Reihe von Sturmmerkmalen wird auf der Grundlage von häufig untersuchten Eigenschaften aus der Sturm- und Naturgefahrengemeinschaft entwickelt. Basierend auf dieser Grundlage werden die Stürme unterteilt, um objektive Sturmklassen zu erstellen, nicht nur für die weitere Auswertungen, sondern auch in dem Bestreben, universell einsetzbare Sturmklassen für ein breites Spektrum von Endnutzern zu schaffen. Dabei stellen wir unser neu entwickeltes quasi-supervised k-means (QSKM) vor, ein halb-überwachtes Clustering-Verfahren zur Partitionierung von Sturmereignissen in Bezug auf einen Katalog historisch schwerer Sturmereignisse. QSKM konstruiert drei verschiedene Sturmklassen, von denen eine dem Sturmkatalog in ihren Eigenschaften sehr ähnlich ist. In einer umfassenden Auswertung der konstruierten Klassen zeigen wir, dass große Ereignisse mit hohen Windgeschwindigkeiten und einer außergewöhnlich langen Lebensdauer häufig im Westatlantik nahe der US-Ostküste entstehen, sich über dem mittleren und östlichen Teil des Atlantiks verstärken und in der Regel auf Mittel- oder Nordeuropa treffen. Ihr Auftreten kann mit einem starken Jetstream und instensiven Tiefdruckgebiet in Verbindung gebracht werden. In den betroffenen Gebieten treten starke Böen über einen Zeitraum von 15-20 Stunden auf, wobei die höchsten Windgeschwindigkeiten meist in den ersten 10 Stunden des auftreten. Vergleichbare schwächere und kleinere Ereignisse finden sich im Mittelmeerraum. Diese Ereignisse entwickeln sich oft im Ostatlantik oder im Mittelmeer, wo sie sich aufgrund ihrer kurzen Lebensdauer auch oftmals wieder auflösen. Aufgrund ihrer langsamen Zuggeschwindigkeit sind betroffene Gebiete dennoch oft bis zu 20 Stunden und länger lokalen Windböen ausgesetzt. Ähnlich kleine und kurze Ereignisse, aber mit detulich höheren Windgeschwindigkeiten, sind typisch für Nordeuropa. Diese Art von Ereignissen hat ihren Ursprung in der Regel in den westlichen Teilen

des Nordatlantiks und verstärkt sich weiter, wenn sie über das offene Wasser ziehen. Obwohl sie nur ein Drittel der Größe der ersten Klasse aufweisen, kann der Kerndruck des übergeordneten Zyklons auch unter 970 hPa fallen.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

On January 15th in 2007, the low pressure system that soon would give rise to the windstorm Kyrill formed over Newfoundland, marking the starting date of one of the most severe European windstorm in recent history. Kyrill made first landfall in Ireland on the 17 January and continued its journey across Europe, leaving a path of death and destruction, until it finally disappeared on 19 January. Within those few days, Kyrill claimed at least 46 lives, uprooted up to 62 million trees, significantly disrupted the European infrastructure, and caused approximately 7 billion Euro of insured losses in Europe (Fink et al., 2009; Roberts et al., 2014).

While events of such an enormous magnitude like Kyrill are rare, windstorms are among the most common and severe natural hazard in Europe. According to the German reinsurance company Munich RE, 42% of the natural hazards that cause damage in Europe between 1980-2012 are of meteorological nature (includes all types of storm and hail) and they alone are already responsible for 62% of insured losses (Munich RE, 2013). Among those, winter windstorms associated with extra-tropical cyclones are by far the events with the largest economic impact (Handmer et al., 2012). In Germany, four of the ten most severe natural hazards (measured by insured losses) between 2002-2017 are windstorm events (GDV, 2018). In summary, windstorms are one of, if not the most relevant natural hazard in Europe, when measured in terms of insured losses.

Given the tremendous economic impact that such a windstorm event can have and the frequency in which they occur, it is not surprising that studies of windstorms often revolve around their occurrence (e.g. Nissen et al., 2010; Donat et al., 2010), potential insured loss (e.g. Donat et al., 2011a; Pardowitz et al., 2016a; Pardowitz et al., 2016b), return values (e.g. Della-Marta et al., 2009; Donat et al., 2011b) or their overall severity (e.g. Klawa and Ulbrich, 2003; Leckebusch et al., 2008), since all those measures are direct or indirect indicators for property damage or obstruction of daily life. Especially in the context of a changing climate, each of these topics gains even more relevance, as it is important for decision makers and risk assessment to understand and being able to predict the potential impact of future windstorms.

However, an aspect of windstorms that is considerably less discussed are their inherent characteristics such as duration, size or travel speed. This is surprising, considering that these characteristics are each a measure of windstorm severity themselves. For example, larger and longer windstorm events affect wider areas for comparable longer periods of time than smaller and

shorter events. Analogously, events that travel faster leave people with less time to response and/or prepare for the incoming hazard. The list of examples can be continued for numerous other windstorm characteristics. Yet, the literature primarily focuses on the previously mentioned measures that outline the economic impact of the event. In cases where characteristics of windstorms are discussed, the focus is usually on intensity, which is often measured in terms of wind speed (e.g. Nissen et al., 2010).

We believe that the characteristics of windstorms are an important aspect that can not only be directly linked to the severity of the event, but also distinguishes them from each other. Not every windstorm is as devastating as Kyrill, but each of them is extreme in its own way. Understanding the differences in characteristics between a windstorm event that causes billions of Euros worth of damage and a windstorm that causes only minor inconveniences in the daily life of European citizens, as well the source behind those differences, could be a major contribution to our general understanding of windstorm development and impact.

Therefore, in this dissertation we focus on the characteristics of European windstorms, understanding the factors governing their variability, and their changes in recent years. For this purpose, we aim to develop a classification scheme of European windstorm types. The goal is to identify windstorm classes based on their characteristics, to understand their differences and identify atmospheric large-scale conditions that drives their development. Finally, we explore trends in each of the identified European windstorm classes and show the potential impact different types of windstorms can have.

The structure of this thesis is a follows. In the remainder of this chapter, we provide an overview of the current state of the art for European windstorms and discuss challenges of windstorm classification, as well as the necessary tools for the task and the overall research questions we intend to answer. In Chapter 2, we describe the data used in this dissertation, how windstorms and cyclones are tracked, and, more importantly, we motivate and define the windstorm characteristics that are part of our classification. The 3rd chapter covers the centerpiece of this work, the introduction, application and evaluation of quasi-supervised k-means, a novel semi-supervised clustering technique specifically developed by us to identify windstorm classes with respect to an existing windstorm catalog. Chapter 4 builds upon the results from Chapter 3 and shows how the identified windstorm classes differ in impact, occurrence and driving mechanism. Chapter 4 ends on a trend analysis of the different windstorm classes. In the last chapter, a synopsis of the thesis is presented.

1.1 Extratropical cyclones and windstorms

Extratropical cyclones (sometimes called midlatitude cyclones) are low-pressure systems formed in the middle or high latitudes and are one of the driving forces behind the day-to-day variability of weather in Europe and the midlatitudes in general. Their lifetime spans multiple days and their size

ranges from several hundred to several thousand kilometers. Particularly strong extratropical cyclones can have a core pressure far below 970 hPa and are often accompanied by extreme weather events such as strong wind gusts, hail and severe amounts of precipitation.

In the midlatitudes, extratropical cyclones usually develop along the polar front. Here, the warm air masses of the equator meet the cold air masses of the poles, creating a small region of particularly strong baroclinicity. The sharp meridional temperature gradient in this region creates a vertical shear of strong zonal wind and flow that is unstable to small perturbations, which in turn can grow into cyclones. Due to its key role, baroclinic instability has been widely excepted as the primary mechanism by which extratropical cyclones develop.

Windstorms, on the other hand, are winds that are strong enough to potentially cause damage (Pielke, 2007) and which may or may not be accompanied by precipitation. Dependent on what causes them, they may last for only a few minutes (e.g. downbursts) or hours and even days (e.g. wind gusts from extratropical cyclones). The key part of the definition of windstorm is the damage aspect associated with the occurrence of strong wind gusts, as it implies that windstorms are defined through surface wind speeds and their impact. Therefore, unlike extratropical cyclones, windstorms are meteorological extreme events (or natural hazards) by definition.

From these two definitions it becomes clear that there is a connection between extratropical cyclones and windstorms. As a matter of fact, strong winter cyclones are frequently the cause of intense and long-lasting windstorms. As already mentioned above, extratropical cyclones typically form along the polar front, a region of strong baroclinicty created between the warm subtropical and cold polar air masses. In winter, the meridional temperature gradient increases as less sunlight reaches the Pole. This strengthens the polar front, which in turn favors the development of cyclones (and windstorms) in the midlatitudes. Which is why, historically, most severe windstorm events occurred in winter (see Roberts et al., 2014).

Due to the close connection between windstorm and extratropical cyclones and the fact that both occur more frequently and stronger during winter, we will focus on European winter windstorms that can be connected to extratropical cyclone. Hence, we assume that each windstorm is part of a unique parent cyclone. By doing so we establish a direct link between upper-level atmospheric processes that favor the development and intensification of extratropical cyclones and extreme surface wind speeds that potentially cause damage or even threaten the wellbeing of people. An introduction to how windstorms and extratropical cyclones are identified and tracked is given in Sect. 2.2.1 and Sect. 2.2.2, respectively. The method that we use to match both types of events is described in Sect. 2.2.3.

1.2 Trends in cyclone and windstorm activity

In 2014, the fifth assessment report of the Intergovernmental Panel on Climate Change (IPCC) summarized the accumulated studies on trends in cyclone and windstorm activity (see IPCC, 2014). They state that studies using reanalysis data indicate a northward and eastward shift in the Atlantic cyclone activity during the last 60 years, consequently leading to more frequent and more intense wintertime cyclones in the high-latitude Atlantic and fewer in the mid-latitude. However, they also point towards a variance in fidelity of findings dependent on the used reanalysis product and the often contradicting results from long-term studies using *in situ* measurement. Furthermore, they mention the problem of incomparable results due to the sensitivities of trend estimations to the used cyclone identification and tracking schemes and/or different definitions for extreme cyclones. For the Northern Hemisphere, the report concludes that confidence in large-scale changes in the intensity of extreme extratropical cyclones, or storminess in general, within the last century is low.

Nevertheless, we want to provide a brief overview of trends in cyclone and windstorm activity from the literature to clarify the current state of the art, especially with respect to Europe. In the framework of the Intercomparison of Mid Latitude Storm Diagnostics (IMILAST) project, Neu et al. (2013) investigated trends in cyclone numbers per winter season for the Northern Hemisphere and their potential dependence on the tracking algorithm. They found an overall increase in the total number of cyclones for the period 1989-2009 for most methods, especially in Europe, but a general decrease in the number of deep cyclones (minimum core pressure below 980 hPa). Donat et al. (2011c) studied long-term trends in European storminess using the 20th Century reanalysis (20CR) product (Compo et al., 2011) of the National Oceanic and Atmospheric Administration and found significant upward trends in the number of stormy days and local wind speeds for central, northern and western Europe since 1871. In response to the work of Donat et al. (2011c), Befort et al. (2016) compared long-term trends in the 20CR and the ERA-20C reanalysis dataset (Poli et al., 2016), the twentieth century reanalysis of the ECMWF. Their results show substantial difference between both datasets for Northern Europe, where the number of cyclone tracks starts to slowly increase in ERA-20C starting in 1950, while they are consistently decreasing in 20CR. For deep cyclones (here, minimum core pressure below 970 hPa) and windstorms, they found a significant increase in activity for the period 1961-1990 in ERA-20CR. With a focus on the Mediterranean region, Nissen et al. (2010) found a significant decrease in cyclone numbers over the western Mediterranean basin and over the Black Sea, and an upward trend in cyclone numbers over the Levant region.

In summary, the literature points towards an increase in cyclone and windstorm activity over Europe. However, there is a strong difference in the magnitude, and sometimes even the sign, of these trends dependent on the data, method, region or period of interest. These differences become even more substantial for deep or extreme cyclones.

1.3 Windstorm classification and machine learning

As already mentioned about, we aim for a classification scheme of European winter windstorms based on their characteristics. Therefore, it is beneficial for the readers to first familiarize themselves with the purpose and applications of classification. In this section, we provide a brief overview of the role of classification in machine learning (Sect. 1.3.1) and show examples of extratropical cyclone and windstorm classification from the literature (Sect. 1.3.2). For a detailed discussion of classification and its purpose and goal in machine learning, data mining or knowledge discovery, we suggest Kloesgen and Zytkow (2002). For a comprehensive overview and description of classification techniques, we suggest James et al. (2013)

1.3.1 The role of classification in machine learning

The goal of a classification is to derive knowledge or information from data by ordering or dissecting it into small and comprehensive classes (i.e. units, subgroups, subsets, partitions), often for the purpose of eventually predicting classes based on new information. Given the goal, there are at least three common interpretations for classification in the scope of data mining and knowledge discovery: classification as ...

1. ... ordering system for objects
2. ... discrimination or class assignment
3. ... class construction or clustering

The first interpretation covers the systematic classification of objects by ordering them into groups with a common ground. This can include classification based on very simple properties such as ordering people into groups of different heights (i.e. small, medium, tall), but also more complex relationships such as the ordering of chemical elements in the periodic table. Of the three interpretations, it is the one most commonly associated with categorization.

For us, the second and third interpretation are more relevant as they cover the more statistical spectrum of classification. Discrimination and class assignment refer to a practice of supervised learning. In these cases, it is known that each object originates from one of N already existing classes (with $N \geq 2$) and can be describe by a set of features. However, it is unknown to which class each object belongs and has to be reconstructed from the features. The solution being a single class is what distinguishes discrimination or class assignment from ordering, where objects can belong to different classes dependent on the ordering system. Class construction or clustering, on the other hand, refers to a practice of unsupervised learning, where there are no classes to begin with and have to be constructed from the features.

1.3.2 Examples of extratropical cyclones and windstorms classification in climate science

Classifications have always been a popular tool in meteorology and climate science because they allow atmospheric situations and the resulting weather impacts to be summarized into comprehensive and predictable packages. A good example for classification in meteorology are European weather regimes (Rex, 1951), characteristic large-scale atmospheric circulation patterns, which have long been proficient tools in weather prediction because of their reliable connection to surface weather and weather extremes such as heat waves or widespread heavy precipitation in Europe. Given the popularity and success, it is not surprising that their have been classifications of extratropical cyclone and windstorms in the past.

A comprehensive and good review of extratropical cyclone classification and its use in climate studies has been published by Catto (2016). They have distinguished the current state of the art in cyclone classification into eight different categories, ranging from simple conceptual models to synoptic-dynamic classifications, classifications based on impacts or satellite images, and more. In their work, Catto (2016) points towards the past and potential future gain of insight into extratropical cyclone dynamics that has been and can be achieved through classification, and the aid these techniques provide in the forecasting of cyclone systems.

In the context of this dissertation, extratropical cyclone classification based on impacts is closest to our intentions of classifying windstorms. Catto (2016) identified three major impacts associated with extratropical cyclones that could be used to classify the severity of an event: wind, precipitation and serial clustering. With respect to wind, there have been numerous studies discussing windstorm-generating extratropical cyclones in terms of wind damage (e.g. Donat et al., 2011a; Munich RE, 2013; Handmer et al., 2012), their structure (Hewson and Neu, 2015) or trends in occurrence (e.g. Paciorek et al., 2002; Neu et al., 2013; Befort et al., 2016). Similar, large proportions of precipitation in the midlatitudes, as well as the most extreme precipitation events, can be associated with extratropical cyclones and their fronts (Pfahl and Wernli, 2012; Catto et al., 2012; Zhang and Colle, 2017). Serial clustering, on the other hand, describes the case in which extratropical cyclones occur in close succession within the same region. In cases of serial clustering, the socioeconomic damage can greatly enhance due to long and constant exposure of extreme winds and precipitation (e.g. Mailier et al., 2006; Dacre and Pinto, 2020). A prominent example are the extratropical cyclones Anathol, Lothar and Martin (names given by the German weather service) from December 1999 which hit Europe in close succession and caused billions of insured losses (Munich RE, 2002).

A classification explicitly focusing on windstorms and their local impacts was published by Dreveton et al. (1998). Their goal was a semi-automated classification of windstorms over France primarily based on expert knowledge. For this purpose, they first designed, in consultation with an experienced weather forecaster, a number of meteorological features from a list

of storm events provided by insurers. The result was a classification table with seven different possible outcomes based on characteristics such as origin, atmospheric flow and affected areas (see Appendix B.1). In the second stage, the expert knowledge was automated with numerical criteria, calculated from meteorological surface observation data.

Using their semi-automated classification model, they identified four meteorological key parameters to evaluate storm risk:

1. Maximum value of the wind gust during the event,
2. Duration of strong winds,
3. Whether the direction of the wind is typical (damage is more significant if not),
4. Precipitation amounts above 35 mm (higher damage potential).

Based on their results, Dreveton et al. (1998) argue for the necessity of windstorm classification because the different storm types showed obvious differences even at the same station. However, they also conclude that their methodology must be carefully applied and extensively tested to be consistent with the expert knowledge.

1.3.3 Challenges in windstorm classification

The key problem in windstorm classification is the lack of clear and obvious inherent classes, and therefore classes usually must be constructed in advance. This distinguishes windstorms from other extreme events that can be easily classified, for example, based on their physical state of matter such as precipitation (e.g., solid or liquid) or based on their type of occurrence like floods (e.g., flash flood or river flooding). However, classes are often constructed to meet the needs of the user, which always makes them subjective to some degree and usually limits their use to the extent that they can only be used to answer the scientific question for which they were originally created. A good example is the previously mentioned work of Dreveton et al. (1998), which heavily builds on expert knowledge. In their work it is neither discussed or shown, nor can it be predicted, how their classification and results would change if a different expert would have provided the foundation for their methodology. Therefore, the challenge of windstorm classification is to create objective and general applicable windstorm classes.

1.4 Research questions and contribution to science

The main goal of this dissertation is to classify European winter windstorms based on characteristics such as duration, size and intensity. The classification is designed to meet the following criteria:

- the classes are as objectively designed as possible, and thus applicable even outside the scope of this thesis,

- the classes are relevant for end users such as scientists, insurers, or decision makers.

At first glance, the two criteria seem to contradict each other, as it is difficult to maintain an unbiased methodology while producing results for a target audience. It should be noted, however, that not all windstorms have the socioeconomic impact of an event of the magnitude of Kyrill. A windstorm classification based on characteristics should naturally distinguish between these type of events. Therefore, our criteria are compatible in a sense that at least one class should consist of "Kyrill-like" events that pose a thread to nature, buildings and/or human lives.

For the purpose of class construction, we propose a novel clustering technique called quasi-supervised k-means. The method is specifically developed for creating classes with respect to an already classified subset of events, and we use it to ensure that extreme windstorm events such as Kyrill are well represented within our constructed classes. A comprehensive description and derivation of this new clustering algorithm is given in Chapter 3. The identification and tracking of the necessary extratropical cyclones and windstorms are described in Chapter 2. In the same chapter, we also motivate and calculate our windstorm characteristics. These characteristics are designed with the intent to represent the most basic attributes of a natural hazard.

The second goal is an extensive evaluation of the resulting windstorm classes. First, we aim to identify and understand key characteristics determent for these classes (Sect. 3.4). This will help us understand what characteristics define each windstorm class and to identify potential threads that can be expected from them. Second, we related the classes to characteristics large-scale situation such as teleconnection modes, baroclinicity fields and attributes of the parent cyclone (Sect. 4.2). The goal is to identify and quantify atmospheric situations that or growth factors that favor the occurrence of an event of a certain windstorm class, and to provide a model for predictions. Third, we aim to quantify potential regional impacts (Sect. 4.1) in the event of a windstorm of a certain class. Finally, we also aim for a trend analysis in windstorm occurrence for each of the classes (Sect. 4.3).

From our goals, we derive four research question that we intend to answer in the course of this dissertation:

1. Can windstorm characteristics be used to construct objective and generally applicable windstorm classes?

2. What are the potential impacts in the event of a windstorm of a certain class?

3. Can we quantify the influence of large-scale atmospheric features, such as teleconnections or the parent cyclone, on windstorms using modern machine learning approaches?

4. Are there significant trends in windstorm characteristics within the observed period?

In summary, in this dissertation we aim to provide objective and universal applicable classes for European winter windstorms, to identify key windstorm characteristics that govern their class, to provide an understanding of their driving atmospheric mechanisms and their potential impacts, as well as trend analysis.

1.5 Important notations and remarks

In science, there are often multiple terms for the same topics, subjects, or methods, as terminology changes depending on the scientific field. For the reader, this can easily lead to confusion and hamper the process of understanding or learning of the topic. A prominent example for this are empirical orthogonal functions (EOFs) and principal component analysis (PCA), which are essentially the same statistical method with the same output, but with completely different names and terminology. Therefore, it is not only in the interest of our readers but also our own, to give a brief overview of our used terminology.

In general, we use the terminology currently popularized by the increased use of machine learning in science. Meaning, we use terms such as target (i.e. target variable) and features in favor of response variable and covariates. Within this context, a single data point consist of a target value and feature value(s). In addition, for simplicity, we use generalizations for certain terms. Hereafter, we refer to extratropical cyclones as cyclones unless otherwise noted. Analogously, models will always refer to statistical models rather than climate models. A list of our used statistical terminology and generalizations, as well as a few alternative notations, are given in Tab. 1.1.

TABLE 1.1: List of used terms and generalizations, plus alternative notations.

Used term	Alternative term(s)
Target	Regressand, response variable, dependent variable
Feature	Explanatory variables, covariates, independent variables
Model	Statistical model
Cyclone	Extratropical cyclone
Windstorm	Cyclone footprint, wind footprint
Jet stream	Polar jet stream
Accuracy	Fraction correct, percent correct score

Chapter 2

European winter windstorms - Identification, tracking and characteristics

In this chapter, we give a general overview of the data and methods that are used to identify, track and match extratropical cyclones and windstorms [1] (Sect. 2.1-2.2). By matching the tracked windstorms to parent cyclones, we will later have additional information for the classification. A short summary of the identified and tracked windstorms will be given in Sect. 2.3. Afterwards, we will introduce and motivate the windstorm characteristics that we will focus on in the remainder of this study (Sect. 2.4). The chapter will close on a first look at the tracked windstorms (Sect. 2.5), followed by a summary of the chapters content and how the shown results further motivate our research project (Sect. 2.6).

2.1 Data

We use two sets of reanalyses: ERA5 (Hersbach et al., 2020) and ERA-Interim (ERAINT; Berrisford et al., 2011). The ERA5 reanalysis is our primary data source, which we use to identify and track windstorms (Sect. 2.3), estimate features from multiple meteorological variables at different atmospheric levels (Sect. 4.2), or quantify trends (Sect. 4.3). ERAINT, on the other hand, is exclusively used to identify and track cyclones (Sect. 2.3).

ERAINT is a widely popular climate reanalysis of the European Center for Medium-Range Weather Forecasts (ECMWF) and is operational since 2006. Its 6-hourly analysis are available for the period 1979 to August 2019, has a spatial resolution of approximately 80 km (0.75°) and resolves the atmosphere on 60 pressure levels from the surface (i.e., the 1000 hPa level) up to 0.1 hPa.

ERA5 is the latest climate reanalysis of the ECMWF and successor of ERAINT. It was put into operation in 2016 and has fully replaced ERAINT since September 2019. This new reanalysis product provides hourly estimates for more than 240 atmospheric, land and oceanic climate variables for the period 1950 to present. The data is resolved horizontally on a 31 km

[1] A definition of cyclones and windstorms in the context of this thesis is given above in Sect. 1.1.

(0.25°) grid across the globe and vertically on 137 pressure levels from the surface up to a height of 0.01 hPa.

Although ERA5 outmatches ERAINT in every technical aspect, we use ERAINT for the identification and tracking of extratropical cyclones. The reason for this is of purely technical nature. Here, we use a numerical scheme based on the works of Murray and Simmonds (1991) to identify and track cyclones (see Sect. 2.2.2). This scheme includes a set of parameters that are dependent on the spatial and temporal resolution of the input data. At the time of the writing of this thesis, these were not adapted for the new higher resolution of ERA5. As the adaptation of the parameters is not trivial and outside of the scope of this thesis, we decided to use ERAINT for the cyclone tracking instead of ERA5. This decision should only have a low impact on the quality of the tracking as large cyclones are reasonably well resolved by ERAINT (Hewson and Neu, 2015).

As we intent to match the cyclone tracks from ERAINT with the windstorm tracks from ERA5, two necessary adjustments have to be made. First, both reanalysis sets require a common observation period and are therefore reduced to only cover the years 1981-2017. Second, ERA5's temporal resolution is reduced to the 6-hourly interval of ERAINT.

2.2 Methods

2.2.1 Windstorm tracking and the storm severity index

For the identification and tracking of windstorms, we use the WiTRACK scheme. It was first introduced by Leckebusch et al. (2008) with the aim to design an objective tool for identifying and tracking windstorms based on nothing more than fields of extreme wind speeds. Since then, the algorithm behind the scheme has been further developed in great detail by Kruschke (2015) and used in several studies of European windstorms (e.g. Nissen et al., 2010; Pardowitz et al., 2016a; Walz et al., 2018). In the following, we give a short outline on how windstorm objects are identified and tracked. For a more comprehensive description of the algorithm we refer to Kruschke (2015).

In the identification process, the algorithm detects contiguous areas of surface wind speeds exceeding a certain threshold at each time step (dark gray area in Fig. 2.1). The threshold at each grid box is a parameter that must be set by the user. Usually it is set to the local climatological 98^{th} percentile, meaning the detected wind speed at a grid box belongs to the upper 2% of the local wind speed distribution. This relationship comes from the literature, where the strongest 2% of surface wind speeds are often associated with damage on housing or nature (e.g. Klawa and Ulbrich, 2003; Leckebusch et al., 2007; Donat et al., 2011b).

In the tracking process, the previously identified contiguous areas are assigned to each other according to a nearest neighbor scheme. For a better illustration of this process imagine the tracking of a single windstorm as in Fig 2.1. First, the center of the contiguous area is calculated using the

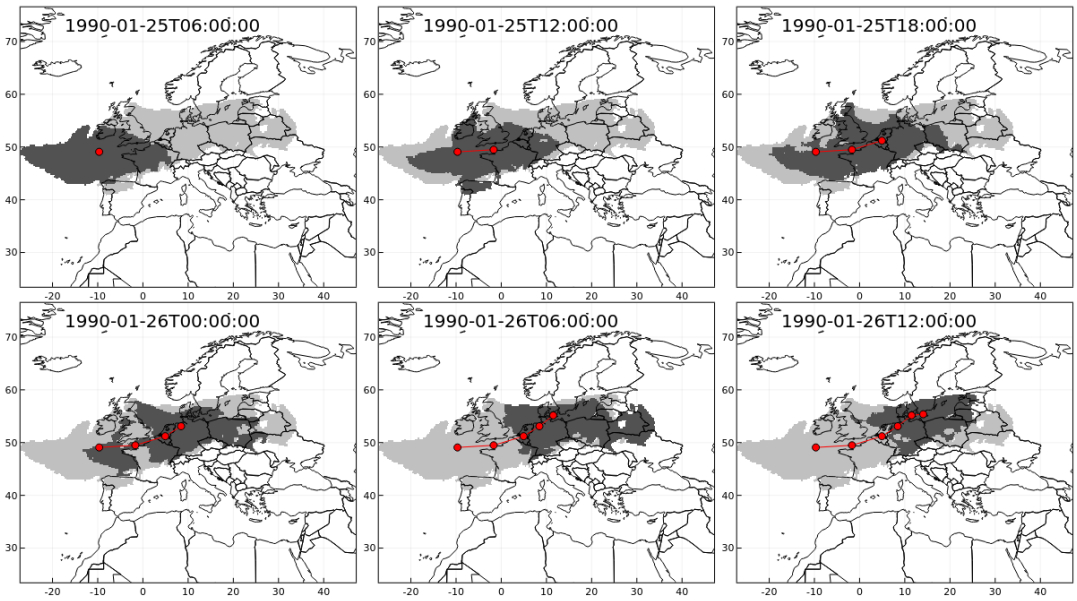

FIGURE 2.1: Tracking of 6 time steps of the windstorm "Daria". Dark grey grid boxes exceed the local climatological 98th percentile of 10 m wind speed at the current time step. Grid boxes that where hit at least once during the tracking are shown in light gray (i.e. the footprint). The red dots mark the cluster centers. Linked together, the centers build the windstorm track (red solid line).

weighted average of the longitudes and latitudes of all grid boxes within the identified areas (Fig 2.1, red dot). A distance is calculated between the currently tracked center and the center of each object in the following time step. The center with the shortest distance is linked to the current center (Fig 2.1, red solid line). Centers that exceed a maximum distance of 600 km are excluded from this process. If no center falls below the maximum distance, the tracking is stopped.

Finally, a filter is applied in which the tracks are checked for a minimum length, minimum total physical area, and maximum translational velocity. Should a track not exceed a minimum of 4 time steps (here, equivalent to 18 hours), it gets discarded from the tracking. The same applies to windstorm events that do not have at least a minimum total size of 150000 km^2 or exceed a translational velocity of 600 km/h (Kruschke, 2015).

Storm severity index

In addition to introducing an identification and tracking scheme for windstorms, Leckebusch et al. (2008) proposed an objective index for the estimation of their severity, namely the storm severity index (SSI). The SSI is based on the same exceedance of local wind speed thresholds that was previously used to track the windstorm and is defined as

$$\mathrm{SSI} = \frac{1}{A_0} \cdot \sum_t \sum_x A(x) \left(\max \left\{ \frac{v(t,x)}{v_{98}(x)} - 1, 0 \right\} \right)^3 , \tag{2.1}$$

where $A(x)$ and $v_{98}(x)$ are the area and 98$^{\text{th}}$ wind speed percentile, respectively, of a grid box x, and $v(t,x)$ the local measured wind speed at x at time t. The parameter A_0 corresponds to a reference area chosen for normalization and is set to 12363.7 km^2, approximately the area of a $1^\circ \times 1^\circ$-grid box at the equator (Kruschke, 2015), and primarily serves the purpose of keeping the SSI dimensionless. Hence, the SSI is the cubed scaled wind speed summed over the size and lifetime of the windstorm.

From Eq. 2.1 we can see that the SSI depends on three windstorm characteristics: the duration of the event, the size of the event, and the intensity of the surface wind speed. The SSI is proportional to each of these characteristics, i.e. the larger the characteristics, the higher the magnitude of the SSI. Therefore, the index is a useful tool to rank windstorms, identify particular extreme events or investigate climate trends in windstorm severity (e.g. Leckebusch et al., 2008; Roberts et al., 2014; Moran, 2019). In this study, we primarily use the SSI as a measure for the potential severity of our identified windstorm classes.

2.2.2 Cyclone tracking

For the identification and tracking of cyclones we use an objective numerical scheme that was developed by Murray and Simmonds (1991) for the Southern Hemisphere and later adapted for the Northern Hemisphere by Pinto et al. (2005). The scheme is fully automatized and only requires fields of mean sea-level pressure (MSLP) as input. Its general process is comparable to the WiTRACK scheme described in Sect. 2.2.1 and consists of three major components: (*i*) identification of the cyclone center, (*ii*) tracking of the centers and (*iii*) selection by user defined criteria.

In the identification step, cyclone centers are located by detecting local maximum in the Laplacian of pressure ($\nabla^2 p$) for every time step. In case such a maximum exist, the algorithm searches for a pressure minimum within a 1200 km radius. If such a minimum is found, the cyclone is classified as closed depression and its center is the longitude and latitude of the aforementioned pressure minimum. However, if no such pressure minimum is found, the algorithm searches for a minimum in the pressure gradient instead. These type of cyclones are classified as open depressions and their center is the location of the minimum pressure gradient. Finally, the cyclone centers are filtered by strength. Closed depressions are discarded for $\nabla^2 p < 0.3$ hPa deg.lat.$^{-2}$ and open depression for $\nabla^2 p < 0.2$ hPa deg.lat.$^{-2}$. Additionally, cyclones which are localized in areas with an elevation of over 1500 m are removed as well.

The tracking of the identified cyclone centers consists of multiple steps. To illustrate, imagine a single cyclone center for a given time step. First, the most likely position of this cyclone center is predicted for the next time step based on the so-called steering-flow, a measure calculated from past motion and pressure tendencies (e.g. Simmonds et al., 1999). Afterwards, the newly predicted and previously identified cyclone centers of the next time

step are evaluated regarding their distance. If the distance falls below a critical threshold of 12.5 deg. lat., the identified center gets assigned to the current track. Should multiple cyclone centers fall below the threshold, the currently tracked cyclone center is matched with the most probable candidate based on distance and core pressure differences. The cyclone center is denoted as decayed or newly generated, if it cannot be further matched.

Finally, the cyclone tracks are filtered based on user defined criteria. Traditionally, these criteria revolve around the cyclones minimum lifetime or whether the system is closed or not. Here, we remove every track that has not at least 5 time steps (equivalent to 24 hours).

2.2.3 Windstorm and cyclone matching

For the matching of the windstorm and cyclone tracks we use the method proposed by Nissen et al. (2010). Their matching process is divided into three iterations. At each iteration, slightly different criteria are used to match a windstorm event to a single cyclone. Only windstorm events that cannot be matched at the current iteration enter the next one. If they pass through all three iterations without being matched to a cyclone, they are discarded. The individual iterations are shortly described below.

In the first iteration, all cyclones that cover the full lifetime of the windstorm event are singled out, and the average distance between the centers of the windstorm and filtered cyclones is calculated. Afterwards, if there is more than one pair with an average distance of below 800 km, the strongest cyclone in terms of averaged Laplacian of pressure is selected. If no cyclone meets this criterion, the distance threshold is increased to 1200 km and the cyclone with the shortest distance is selected for pairing instead.

After the first iteration, the vast majority of windstorms are each matched to a cyclone, but in cases where the event is the result of two consecutive cyclones, a second iteration is usually required. Here, cyclones and windstorms are matched based on a period of coexistence. This period of coexistence must cover at least 4 time steps and must not exceed an average distance of 1200 km between the centers. If both criteria are met, the pair with longest period of coexistence is selected.

The third and last iteration covers a special case. So far, the matching process in the first two iteration was conducted amongst cyclones that have been closed and stronger than 0.7 hPa deg.lat.$^{-2}$ at least once during their lifetime. In the Mediterranean region, which is part of our study area, cyclones that never have been closed during their lifetime can be generated due to local effects. To account for these, the process from second iteration is repeated again, but this time for open depressions

2.3 Identification and tracking

Windstorms and cyclones are both tracked on the Northern Hemisphere between 20-90°N for the extended winter period October to March in all years

(1981-2017) following the methods described in Sect. 2.2.1 and Sect. 2.2.2. As mentioned in Sect. 2.1, we use the ERA5 data for tracking windstorms and the ERAINT data for tracking cyclones.

Since we are only interested in winter windstorms with a potential impact on Europe, we filter all tracked windstorm events that passed through the region 25°E-45°W and 25°-75°N (see region in Fig. 2.1). As an criteria for this filter, the centers of the events must have spend at least 24 hours of their lifetime within the above defined European region. This way we remove all events that only scratched the boarders of our region of interest. After the filtering, we are left with 2234 windstorm events.

Windstorm and cyclone tracks are matched following the procedure of Nissen et al. (2010) described in Sect. 2.2.3. After the first iteration, 1911 of 2234 (∼85.5%) windstorms were successfully assigned to a closed and intense cyclone. Another 277 of the 323 remaining windstorms were successfully assigned to a closed and intense cyclone within the second iteration. In the last iteration, 15 of 46 remaining windstorm events were successfully assigned to an open cyclone. Overall, 2203 of 2234 (∼98.6%) windstorm events could be assigned to a parent cyclone.

2.4 Windstorm characteristics and severity

One can define numerous windstorm characteristics, ranging from obvious measures such as the intensity or size of the event to more complex measures that describe the development of the cyclone or footprint, and each one of them can be defined in multiple different ways. For example, the intensity of a windstorm can be the average, maximum or a specific quantile of the wind speed. It could also be defined through a severity index such as the SSI (Leckebusch et al., 2008) or Beaufort number (see McIlveen, 1991). Dependent on the task, it might also be relevant whether the characteristics are calculated over the full lifetime of the event or for each time step individually. Therefore, it is important to clarify in advance which characteristics we are interested in and how they are defined.

As already discussed in the introduction (Chapter 1), our intention is to focus on windstorm characteristics that are relevant for end-user, such as the fire brigade or insurance companies, or risk communication in general. For this purpose, we look at windstorm characteristics from the perspective of natural hazards. According to the CapHaz-Net report on "Risk Communication and Natural Hazards" (Höppner et al., 2010), characteristics of natural hazards that are commonly regarded to impact on the needs for risk communication can be divided into two groups: characteristics that relate to the occurrence of hazard events and those that relate to their impact. The occurrence group includes characteristics such as the speed of onset, which is the time between the start of the event and its peak, the frequency or return period, and the typical size of the event. In other words, the group covers characteristics that generally describe the occurring event. The impact group, on the other hand, includes characteristics that describe the level of threat an

TABLE 2.1: Characteristics of natural hazards and the respective challenges for risk communication from Höppner et al. (2010).

Characteristics of natural hazards	**Main challenge for risk communication**
Occurrence of hazard event:	
Speed of onset: for example, sudden or slow/creeping	Perception of hazard or risks; spatial and temporal specificity of forecasts/communication; warning
Frequency or return period	Perception of risks, awareness; memory
Size	Perception of hazards and risks; spatial and temporal specificity of forecasts/communication; warning
Novelty/Change	Perception of hazards and risks; awareness; experiences
Impact of hazard event:	
Area: focused or diffuse	Perception of hazards, impacts and risks; awareness, warning
Intensity	Perception of hazards, impacts, consequences and vulnerability
Immediacy and duration of impacts	Perception of impacts, consequences and vulnerability; awareness

event is for a region or population. Hence characteristics that answer questions such as: How intense is the event? Is only one area affected or several? Are the effects immediate or delayed? How long are areas affected by the event? Table 2.1 shows a list of characteristics Höppner et al. (2010) found to be important, the group they belong too, and what type of challenges they pose for the communication of risk.

From the characteristics listed in Tab. 2.1, we decided to adapted the speed of onset, size, intensity and duration for study. However, a key difference between the work of Höppner et al. (2010) and our study is, that we are only interested in windstorm characteristics that describe the occurring event and not its impact on certain regions. In other words, characteristics of the occurrence group in Tab. 2.1. To further elaborate this difference, consider the definition of the intensity of a windstorm. Following Höppner et al. (2010), the intensity of a windstorm could be defined in terms of the damage it causes, be it insured loss, human casualties or fallen trees. For our purposes, however, it is more appropriate to define intensity in terms of physical properties of the events, such as average or maximum wind speed. This applies for the duration as well. Note that we did not included any types of characteristics representing the frequency, return period or novelty/change from the definitions in Tab. 2.1 as we believe that those are better investigate in the framework of a trend analysis (Sect. 4.3) or climate change study. Area

and immediacy, on the other hand, were not taken into account, as they cannot be directly associated with the windstorm without reference to an impact area.

Alongside the characteristics of natural hazard in Tab. 2.1, we decide to also look into windstorm specific characteristics that might not be covered by the general definitions of Höppner et al. (2010). Two factors, that we could not directly or indirectly related to any of the listed characteristics, are the origin and region of impact. Here, origin does not necessarily refer to the region of cyclogenesis, as the associated windstorm (i.e. footprint) often occurs hours or even days after the cyclone has formed. However, various studies (e.g. Gray and Dacre, 2006; Pinto et al., 2009; Allen et al., 2010) showed that the origin of European cyclones is a good indicator for their intensity. Therefore, we assume that origin of the windstorm might also be influential itself. The region of impact usually refers to certain subregions of the European continent. In the literature, windstroms are studied with respect to one or more subregions to either investigate the influence or importance of different atmospheric process for different subregions (e.g. Walz et al., 2018) or, more commonly, to highlight windstorm activities in a specific subregion (e.g. Donat et al., 2010; Nissen et al., 2010; Pardowitz et al., 2016b). As our own addition to the commonly used characteristics from the literature, we also investigate the travel distance and speed of windstorm events. Both could be important features for understanding extreme windstorms from a more complex perspective than just high wind speeds or large size. For example, from a risk assessment perspective, there is much less time to respond to a nearby and fast-moving extreme event than to an event that was predicted days ago.

In summary, we look at seven windstorm characteristics, namely the intensity, size, duration, origin, region of impact, travel speed and distance. In the following subsections, we give a detailed description of how we define and calculate each of the named characteristics. A short summary of those descriptions, plus the respective abbreviations of the characteristics, is given in Tab. 2.2.

Intensity and size

As mentioned in Sect. 1.5, a windstorm footprint is defined as the sum of all contiguous area of grid boxes x where the 10 m wind speed $v(x,t)$ exceeded the local climatological 98$^{\text{th}}$ wind speed percentile $v_{98}(x)$ at the time step t (see light gray area in Fig. 2.1). For such a footprint, we defined the intensity as the average wind speed (MEANV) of the footprint, averaged over the duration of the event, i.e.

$$\text{MEANV} = \frac{1}{|t|} \sum_{t_i \in t} \sum_{x_i \in x} v(x_i, t_i) / |x_{t_i}|, \tag{2.2}$$

where $|x_{t_i}|$ is the number of grid boxes that exceed the 98$^{\text{th}}$ wind speed percentile at time step t_i and $|t|$ the number of time steps. Therefore, MEANV

TABLE 2.2: List of windstorm characteristics, plus their abbreviation (Abbr.), unit and a short description. For more details, see the respective subsections.

Characteristic	Abbr.	Unit	Description
Intensity	MEANV	m/s	Average wind speed
Duration	DUR	h	Lifetime of the event
Size	AREA	km^2	Size of the event
Speed of onset	ONSET	h	Hours between start and peak of the event (max. SSI)
Origin	OLON,OLAT	degree	Longitude/Latitude of first detection
Destination	DLON,DLAT	degree	Longitude/Latitude of last detection
Distance	DIST	km	Full travel distance of the event
Travel speed	TSPEED	km/h	Moving speed of the event

can be interpreted as average wind speed across all contiguous areas of the event.

Analogously, we define the size of the event as the area (A) of all grid boxes in the footprint, averaged over the duration of the event:

$$\text{AREA} = \frac{1}{|t|} \sum_{x_i \in x} A(x_i). \tag{2.3}$$

Note, that we use AREA as abbreviation for the size of the event, while Höppner et al. (2010) use area to characterize the spatial distribution of an impact caused by a hazard. We do this because it is consistent with the naming in the scheme proposed by Kruschke (2015) and more accurate given that the unit is km^2.

Duration and speed of onset

The duration of an event is defined as the hours between the first and last time step of the event track. Hence, lets $t = t_1, \ldots, t_N$ be a series of N time steps at which a windstorm event was tracked, then the duration is calculated as

$$\text{DUR} = t_N - t_1. \tag{2.4}$$

Similar, the speed of onset are the number of hours between the first time step and the time step where the event reaches its largest SSI (t_{max}):

$$\text{ONSET} = t_{max} - t_1. \tag{2.5}$$

Origin, destination, travel speed and distance

Origin and destination (as proxy for the region of impact) are defined as the respective longitude (LON) and latitude (LAT) of the first and last tracked storm center of an event (see red dots in Fig. 2.1). Again, if $t = t_1, \ldots, t_N$ is a series of N time steps at which a windstorm event was tracked, the origin (with prefix O) is calculated as

$$\text{OLON} = \text{LON}(t_1) \quad \text{and} \quad \text{OLAT} = \text{LAT}(t_1). \tag{2.6}$$

Analogously, the destination (with the prefix D) is calculated as

$$\text{DLON} = \text{LON}(t_N) \quad \text{and} \quad \text{DLAT} = \text{LAT}(t_N). \tag{2.7}$$

To calculate the travel speed, we first have to calculate the distance. The distance is calculate using the haversine formula as proposed in Gade (2010), which is a formula for calculating the spherical distance d between two points on the surface of a sphere. The formula requires three inputs: the coordinates of the first and second point, i.e. $p_i = [LON_i, LAT_i]$ and $p_j = [LON_j, LAT_j]$, and the radius of the sphere:

$$d_{i,j} = \text{haversine}(p_i, p_j, r), \tag{2.8}$$

with $r = 6371$ km, the averaged radius of the earth. For a windstorm footprint, we calculate d between consecutive storm centers and sum the resulting d's for the overall travel distances of the event:

$$\text{DIST} = \sum_{i=2}^{N} d_{i-1,i}. \tag{2.9}$$

Afterwards, we divide the previously calculated d's by 6 to get hourly velocities (remember, we track 6-hourly data), which we average to the overall travel speed of the event:

$$\text{TSPEED} = \frac{1}{N-1} \sum_{i=2}^{N} d_{i-1,i}/6. \tag{2.10}$$

We have to average by dividing through $N - 1$, since there is one less d than time steps t.

2.5 Basic characteristics of windstorm events

In this section, we provide an overview of our previously tracked windstorm events, focusing on spatial and temporal resolution and representation of historically severe events, as well as a first look at our previously defined windstorm characteristics. We do this for several reasons. For one, at the time of writing, there is not a lot of published work about windstorms tracked from the newly introduced ERA5 reanalysis data set. Therefore, this overview

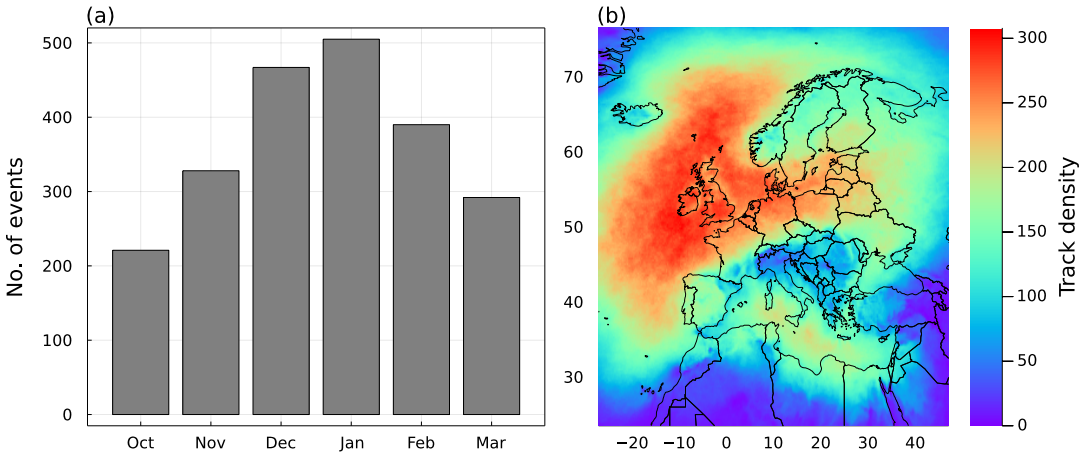

FIGURE 2.2: Temporal and spacial distribution of windstorm events from 1981-2017 over Europe with (a) showing the number of events per month and (b) the number of times the 98^{th} wind percentile was exceeded within a grid box.

serves as a type of quality control in which we evaluate whether or not windstorm tracks show what already has been established in the windstorm literature based older reanalysis data. Second, a general overview will help us determine the contribution of each individual windstorms class to the overall picture of European windstorms.

The majority of the events occur in the winter months December to February (Fig. 2.2a). Windstorm activity is generally higher over the water body of the North Atlantic than over the European landmass (Fig. 2.2b). The maximum track density, here defined as number of measured exceedences of the 98^{th} wind percentile, lies southwest to the British Isles. From there, the high-density field (values above 200) splits into two sidearms, one extending northward and the other toward the northern parts of central Europe. Additionally, Fig. 2.2b shows smaller areas of high activity in the Mediterranean. In summary, both attributes resemble the results from the literature. An increased number of windstorms from December to Feburary is a common phenomenon in the midlatitudes (see Sect. 1.1 or Pinto et al., 2009) linked to an increased meridional temperature gradient in this season. Similar, the shape of the high track density can be connected to the large-scale atmospheric flow, which is primarily west to southwest in the midlatitudes of the Northern Hemisphere.

In terms of overall severity, Fig. 2.3 shows that the large majority ($\sim$ 75%) of windstorms have a SSI value below five. This information by itself does not mean much, since per definition of the SSI (Eq. 2.1) each of the tracked events is extreme in its own way. However, for comparison, we marked the SSI values (gray dashed lines in Fig. 2.3) of three of the most severe windstorm events in Europe based on insurance loss (Roberts et al., 2014): "Kyrill" ($\sim$ 6.7 bn USD), "Lothar" ($\sim$ 8.0 bn USD) and "Daria" ($\sim$ 8.2 bn USD). All three events show significantly higher SSI values than a large majority of our tracked windstorms. In fact, 19 out of the 23 windstorms highlighted by the XWS open access catalog of extreme European windstorms from 1979 to 2012 (Roberts et al., 2014) have a SSI larger than five. Again, this does not justify

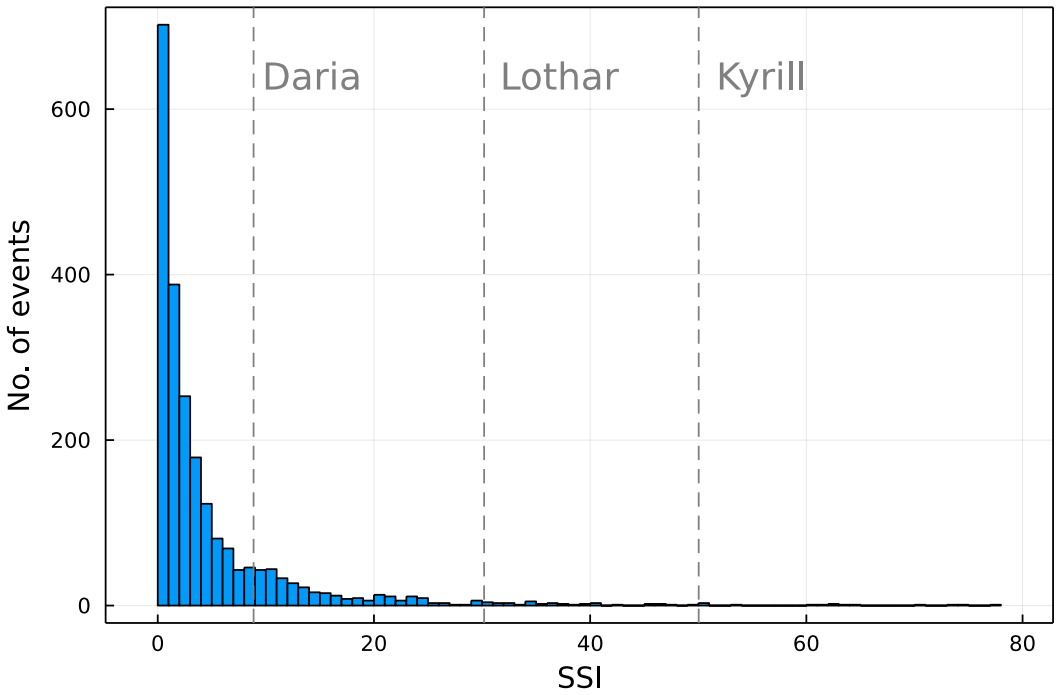

FIGURE 2.3: Frequency distribution of the SSIs of the tracked windstorm events from 1981-2017. The dashed gray lines mark the SSIs of the cyclones "Daria", "Lothar" and "Kyrill" (names given by Freie Universität Berlin and used by the German Meteorological Service, DWD).

to remove a large amount of tracks from the data, especially since our SSI threshold of five was set rather arbitrary. However, it shows that even within the context of extreme events, there is a subset of windstorms at the very top-end of the distribution that are not only extreme but also severe in terms of social-economical relevance.

To get an overview of how likely an event with a certain characteristic is to occur, Fig. 2.4 shows the fraction of events per year in relation to the characteristics. Each box shows the variation of the respective event characteristic over the period from 1981-2017. The duration (Fig. 2.4a) follows the expected behavior, with shorter events occurring more frequently than longer ones. Approximately 35-60% of the events have a duration between one or two days, while events with a duration above four days hardly account for 10% of the annual windstorm events. The average wind speed (Fig. 2.4b), on the other hand, displays a completely different curve. There is a monotonous increase of windstorm frequency for our wind speed bins up to 15 m/s, followed by a drop between 15-17.5 m/s and recovery between 17.5-20 m/s. Events with an averaged wind speed of above 20 m/s usually account for less than 5% of the observed windstorms per year. Windstorm size and SSI show a very similar pattern (Fig. 2.4c and d), with the frequency of events decreasing as the magnitude increases. In both cases, the higher fraction in the last bin is an artifact of counting all events above a certain threshold (1200 and 12 respectively) into one and the same bin.

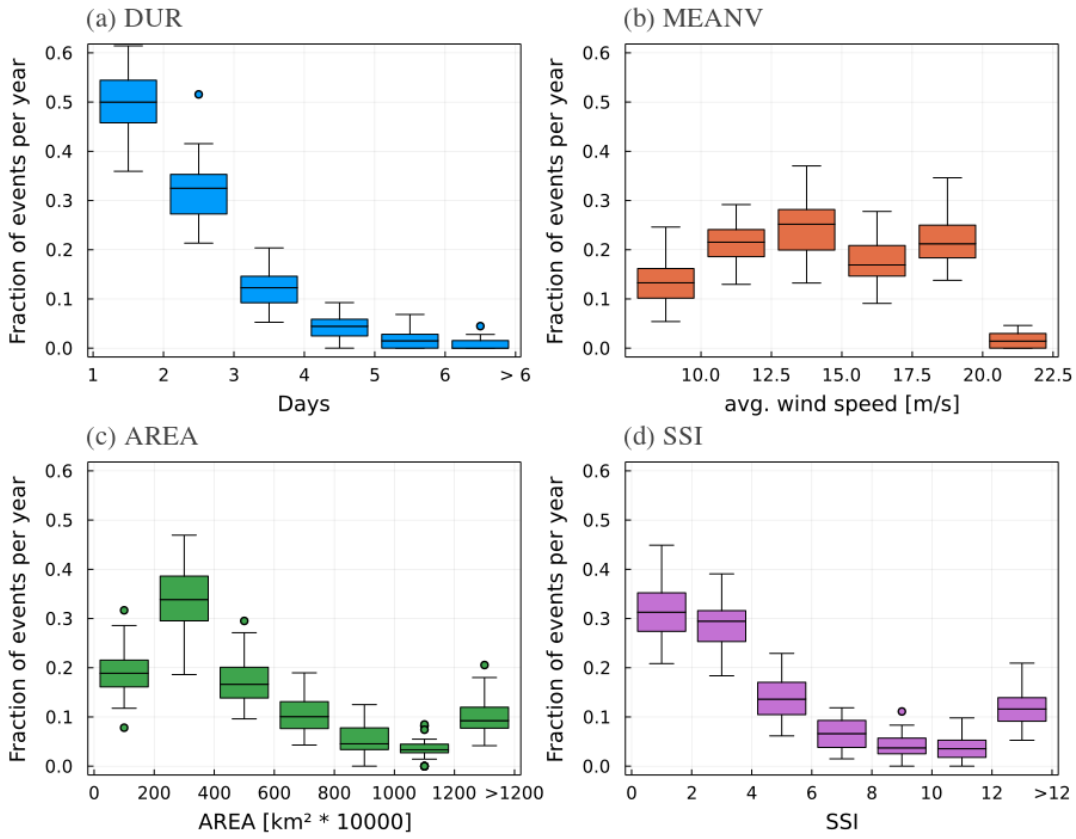

FIGURE 2.4: Box-whisker plots displaying the fraction of events per year for (a) duration, (b) averaged wind speed, (c) size and (d) SSI. Each box indicates the fraction of windstorms in the respective parameter range for the period 1981-2017. For example, according to (a), every year approximately 0.2-0.4 (i.e., 20-40%) of the events had a lifetime between two to three days.

2.6 Summary

In this chapter, we identified, tracked and matched cyclones and windstorm events following the approaches described in Sect. 2.2 and using the climate reanalysis data of the ECMWF. Windstorm tracks were generated by applying the WiTRACK scheme developed by Leckebusch et al. (2008) to the ERA5 reanlysis. The cyclone tracks, on the other hand, were generated following the approach of Murray and Simmonds (1991) using the ERAINT reanalysis. After the identification and tracking, both products where first matched following Nissen et al. (2010) and then filtered for the European region displayed in Fig. 2.1.

Overall, 2203 windstorm events were identified and tracked within the European region for the extended winter period October to March. Events have a lifetime above 18 hours and a minimum total size of at least 150000 km^2. Approximately 98% of the windstorm tracks could be matched to a parent cyclone that was closed and reached a Laplacian of pressure of more than 0.7 hPa deg.lat.$^{-2}$ at least once during their lifetime.

Following the identification and tracking, we introduced a set of 10 windstorm characteristics that will be further investigated throughout the remainder of this thesis. The set is a composite of the characteristics of natural hazards listed by Höppner et al. (2010), a group of windstorm specific characteristics derived from the literature and two additional characteristics that we found to be interesting. A short description of the characteristics, the functions used to estimated them and their abbreviations are given in Tab. 2.2.

Finally, for the purpose of providing a general overview, we investigated the tracked windstorms with respect to their spatial and temporal distribution, their overall severity and how common windstorm events with certain characteristics are. We found no surprising results regarding the spatial and temporal distribution (Fig. 2.2). The highest windstorm activity can be observed during the winter months from December to February and the primarily affected areas are the northern parts of central Europe and large parts of the North Atlantic. The density distribution of the SSI (Fig. 2.3) showed that the large majority of events do not compare in terms of severity to a small subset of very extreme events highlighted by the XWS catalog of extreme windstorms by Roberts et al. (2014). The distribution of characteristics shown in Fig. 2.4 complements the distribution of the SSI. Smaller, shorter and less severe events a more frequent than their counterparts over the span of an extended winter season. The frequency pattern of the wind speed is less clear as it shows a comparably flat curve. Since the frequency distribution of the wind speed is significantly different from the distributions of the size and duration, this could indicate that wind speed is independent of both variables. However, a more comprehensive study of this hypothesis is required before a clear answer can be formulated.

Chapter 3

Quasi-supervised k-means (QSKM)

As discussed in Sect. 1.4, we intend to cluster our windstorm tracks (see Chap. 2) with respect to their characteristics to obtain labels for our classification. The main challenge of this task is to design clusters that are not only representative of all the subsets within the data, but also relevant from a risk and hazard perspective. For this purpose, we developed quasi-supervised k-means (QSKM), a semi-supervised clustering technique that takes advantage of an already know subset of labeled data to improve the quality and relevance of the clustering result.

This chapter will cover an introduction of QSKM clustering. Since k-means is an integral part of QSKM, we will start by giving a comprehensive overview of the traditional k-means clustering in Section 3.1. We will discuss its purpose and classical algorithms as well as methodological flaws, advantages, and disadvantages of the clustering method. In Section 3.3, we motivate the transition from k-means to QSKM, introduce the necessary additional processes, and how these are integrated into the clustering algorithm. Finally, this chapter will close on Section 3.4, where we will validate the performance of QSKM compared to traditional k-means clustering.

3.1 Traditional k-means clustering

Let $X \in \mathbb{R}^{n \times p}$ be a set of n observations, each with p features. In k-means clustering X is partitioned into k subgroups (the so-called clusters) $C_1, \ldots, C_K$ in such a way that three important criteria are met:

1. $C_1 \cup C_2 \cup \ldots \cup C_K = \{x_1, \ldots, x_n\}$,
2. $C_k \cap C_{k'} = \emptyset$ for $k \neq k'$,
3. $\underset{C_1,\ldots,C_K}{\text{minimize}} \left\{ \sum_{k=1}^{K} W(C_k) \right\}$, where $W(C_k)$ is the within-cluster variation (see Eq. 3.1).

The first criteria states that each observation has to belong to at least one of the K clusters, while the second one says that no observation belongs to more than one cluster. Together, both criteria give $C_1, \ldots, C_K$ the properties of being non-empty and non-overlapping.

The third criteria, on the other hand, says that X has to be partitioned in such a way that the sum of the within-cluster variation $W(C_k)$ is minimized, where $W(C_k)$ is commonly defined as

$$W(C_k) = \frac{1}{N_k} \sum_{x_i, x_{i'} \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2, \tag{3.1}$$

with N_k being the total number of observations in the kth cluster. If we insert Eq. 3.1 into the third criteria we get the optimization problem that defines k-means clustering:

$$\underset{C_1,\ldots,C_K}{\text{minimize}} \left\{ \sum_{k=1}^{K} \frac{1}{N_k} \sum_{x_i, x_{i'} \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \right\}. \tag{3.2}$$

Given Eq. 3.2, all that is required is an algorithm that partitions the observations into K clusters such that the criteria are met. This, however, is not a trivial task as there are up to K^n possible ways to partition n observations into K clusters. Under those conditions, clustering a moderate number of observations already becomes a computational challenge. For large amounts of observations, the optimization problem becomes virtually impossible to solve. Therefore, k-means algorithms follow a heuristic approach of solving Eq. 3.2. This is usually done by providing some initial starting conditions. Given the starting conditions, the algorithm converges towards a local minimum for Eq. 3.2 in favor of the initially wanted global minimum.

The by far most commonly used heuristic approach of k-means clustering is the so-called Lloyd's algorithm (Lloyd, 1982), which is shown in Algorithm 3.1. To understand why Alg. 3.1 decreases the value of the objective in

Algorithm 3.1 Lloyd's algorithm for k-means clustering

1. **Initialization step:** Randomly select k observations from $X \in \mathbb{R}^{n \times p}$. These serve as initial means $\{m_1, \ldots, m_k\} \in \mathbb{R}^p$ (the so-called cluster centroids)

2. Iterate until cluster assignments no longer change:

   (a) **Assignment step:** Assign each observation to the the cluster with the closest centroid (usually calculated using Euclidean distance)

   (b) **Update step:** Recalculate centroids for observations assigned to each cluster

$$m_k = \frac{1}{N_k} \sum_{i \in C_k} x_{ij} \tag{3.3}$$

Eq. 3.2, remember the following identity:

$$W(C_k) = \frac{1}{N_k} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 = 2 \sum_{i \in C_k} \sum_{j=1}^{p} (x_{ij} - m_k)^2. \tag{3.4}$$

Step 2 of Alg. 3.1 is designed to further update m_k until the observations are no longer reassigned to other clusters. Therefore, per definition of Eq. 3.4, the within-cluster variation of a subgroup C_k is minimized. As result, the sum of the within-cluster variation is minimized as well.

3.2 Challenges of k-means clustering and consequent extensions of the algorithm

Traditional k-means clustering is one of the most popular algorithms for finding inherent subgroups in data (Wu et al., 2008) and the method comes with a lot of advantages. The algorithm is mathematically intuitive, easy to implement and fast. Yet, despite all its simplicity, k-means clustering can still compete with far more complex and computationally expensive clustering algorithms (Rodriguez et al., 2019).

However, there are two major problems with using k-means, namely imperfect initial conditions for the algorithm and determining the optimal number of clusters (k). The first problem refers to the fact that the initial conditions in Alg. 3.1 are chosen randomly. As discussed above, Alg. 3.1 only finds a local optimum for the k-means optimization problem (Eq. 3.2). If the initial conditions are chosen randomly, there is no way to guarantee that the local minimum is even close to the global minimum. In other words, the clustering solution found with k-means is arbitrarily far from the best available solution. The second problem refers to the fact that the algorithm does not specify k, which therefore must be chosen by the user. However, the correct number of clusters within the data is usually unknown. Both of these problems have a strong impact on the clustering result and therefore have to be accounted for in any k-means application.

It is possible to account for imperfect initial conditions either by (*i*) repeatedly using k-means or by (*ii*) directly improving the initialization step. As long as the initialization includes some degree of randomness, the repeated use of k-means will produce several different cluster solution. The user then chooses the 'best-fit' from among the candidates, where best is usually defined as the clustering result with the overall smallest sum of within-cluster variation (Eq. 3.1). Improving the initialization step, on the other hand, is less straight forward as finding the location of the initial centroids is not significantly easier than the clustering problem itself. Numerous initialization techniques have been proposed in the literature such as kmeans++ (Arthur and Vassilvitskii, 2007) or Maxmin (Gonzalez, 1985). As a comprehensive discussion about different initialization techniques is outside the scope of this thesis, we refer to Peña et al. (1999), Steinley and Brusco (2007), or Fränti and Sieranoja (2019) for additional information about this topic

Fränti and Sieranoja (2019) investigated the impact of using a initialization techniques compared to simple repeated runs of k-means with random starting conditions and found that both processes generally improve the clustering result, especially if used together. However, they also found that cluster overlap is big factor in the initialization process. If there is high overlap, k-means works well regardless of the initialization technique. If there is no overlap, then k-means success depends completely on the initialization technique. Therefore, it may depend on the data whether a more complex initialization is useful or not. However, if the additional computation time caused by the initialization technique is not relevant, it should always be taken into account when using k-means, since it has no negative impact on the clustering result.

There are several methods for determining the optimal number of clusters k, where the most popular are the elbow method (Thorndike, 1953), average silhouettes (Rousseeuw, 1987) and the gap statistic (Tibshirani et al., 2001). All three use a method-specific measure to quantify the quality of the clustering result. These measures are generally independent of k and therefore, if applied to a range of $k = \{2, \ldots, k_{max}\}$, can be used to estimate and compare the quality of a range of clustering results. Again, a comprehensive study of all relevant methods for determining k would exceed the scope of this thesis and we therefore refer to the cited literature for the respective methods. That said, we intend to use average silhouettes at a later stage of this thesis. Compared to the other two, silhouettes are easier to calculate and interpret than gap statistics, and the elbow method provides only a visual interpretation tool for determining k, making it unsuitable for inclusion in an algorithm. Since all three methods are capable of determining the correct k (Yuan and Yang, 2019), our choice fell on silhouettes. A description of the method is given in Appendix A.1.

From the discussed challenges of k-means and the demonstrated solutions to the resulting problems, we can derive a generic algorithm (Alg. 3.2) for how k-means should be applied to any clustering problem. The algorithm should include an initialization technique such as kmeans++ for a better choice of starting conditions. To even further improve the clustering result, k-means should be repeated multiple times to get the best fit from a range of clustering solutions. Finally, the whole process should be tested for different k to find the optimal number of clusters.

3.3 From k-means to quasi-supervised k-means

K-means clustering is an unsupervised statistical learning method that aims to cluster data based on a user-specified set of features. In Sect. 3.1 we introduced the traditional k-means algorithm (Alg. 3.1), discussed methodological challenges that come with its application, and showed necessary extensions to deal with them (Sect. 3.2). Finally, we derive a generic algorithm for the application of k-means to real data (Alg. 3.2).

Quasi-supervised k-means is also a clustering method, yet unlike k-means the clustering process is guided not only by a set of features, but also by a

Algorithm 3.2 Generic k-means algorithm for the application to real data

1. For $k = 2 \rightarrow k_{max}$ do:
   - Iterate I times:
     - Estimation of initial centroids using an improved initialization technique
     - Iterate until cluster assignments no longer change:
       (a) Assignment step
       (b) Update step (Eq. 3.3)
   - Pick 'best-fit' from I clustering result (e.g., min. sum of within-cluster variation)
2. Determine clustering result with optimal k (e.g., Silhouettes, Gap statistic or Elbow method)

previously labeled subset of data called the reference. Hence, the idea behind QSKM is to handle clustering in a semi-supervised manner, where we retain the ability of unsupervised learning methods like k-means to discover hidden subsets within the data, while benefiting from the response-oriented problem solving of supervised methods.

The advantage of such a semi-supervised approach to clustering is that the added labels will inevitable improve the relevance of the resulting clusters, provided they are representative of a clusters within the data. If this is the case, the clustering process will find observations that are close to the labeled data and assign them the respective labels. Therefore, under the condition that at least two clusters are found (which is the default for all clustering methods), in the worst case scenario semi-supervised clustering will at least extend the sample size of a known subset of data, increasing its representation within the data. This is already of great value, because the performance of many statistical methods (e.g., trend analysis or statistical moments) scale with sample size. In the best case scenario, semi-supervised clustering not only increases the representativeness of a subsample, but also further stabilize the clustering process and eventually improves the clustering result by using additional information.

The algorithm of QSKM consist of three parts: the generic k-means algorithm shown in Alg. 3.2, a feature sampling, and a cluster verification against a reference. The extended k-means algorithm is the core of QSKM, where the actual clustering is performed. The feature sampling and cluster verification are further described below. Both processes are contributions by the author and are the primary factors that change k-means from an unsupervised into a semi-supervised approach.

For the purpose of better illustration, the full algorithm is shown in Alg. 3.3. Let $X \in \mathbb{R}^{n \times p}$ be a set of n observation of p features and REF the reference, a subset of observation from $X^{n \times p}$ handpicked by the user. The first step is the

Algorithm 3.3 Quasi-supervised k-means

1. Draw J subsets $S_1, \ldots, S_j \in \mathbb{R}^{n \times q}$ from $X \in \mathbb{R}^{n \times p}$ such that $q \leq p$ and $S_1 \neq S_2 \neq \ldots \neq S_j$
2. For $j = 1 \rightarrow J$ do:
   (a) Apply Alg. 3.2 to S_j to get K clusters $C_{j,1}, \ldots, C_{j,K}$
   (b) Count the no. of elements from the reference (REF) inside each of the K cluster and find the maximum:
   $$\gamma_j = \max \left[\text{count} \left(C_{j,1} \bigcap \text{REF} \right), \ldots, \text{count} \left(C_{j,K} \bigcap \text{REF} \right) \right] \quad (3.5)$$
   (c) Calculate the distance between the centroid of the reference and the centroid of the cluster(s) with most events from the reference using the full feature space:
   $$\delta_j = \min \left[\text{distance} \left(C_{j,max}, \text{REF} \right) \right] \quad (3.6)$$
   with $C_{j,max}$ is a placeholder for the clusters in γ_j
3. Select clustering result j with smallest δ from the clustering results with the largest γ

feature sampling. Here, the algorithm draws J subsets $S_1^{n \times q}, \ldots, S_J^{n \times q}$ from X such that $q \leq p$ and $S_j \neq S_{j'}$ for $j \neq j'$. Both parameters, J and q, have to be specified by the user. The idea behind the feature sampling is that not all features are necessary or suited to represent REF. In some cases, they may even prevent a representative cluster from being found at all. Thus, the goal of feature selection is to find the best set of features to describe the REF. The second step consist of a loop where at each iteration j k-means (Alg. 3.2) is applied to the respective subset S_j. Each of the resulting clusters $C_{j,1}, \ldots, C_{j,K}$ is checked for the number of observations from REF to find the clusters with the highest number of elements (γ_j) from REF. Afterwards, the distance is calculated between the centroids of the clusters in γ_j and the centroid of REF itself. Note, the distance is calculated using all p features. This way, multiple different settings of q can be compared without having to respect a potential bias in δ_j due to a reduced/increased feature space. The case of multiple cluster in γ_j is rare, since the elements of REF would need to be evenly distributed across more than one cluster, but can happen in theory and is account for by δ_j, which is used to identify the cluster from γ_j closes to REF. Therefore, after step two there is only one γ_j and δ_j for each of the J subsets. In the third and final step, all found clusters are evaluated in terms of number of events and distance from REF. The 'best-fit' is the cluster with the largest number of events from REF and the smallest distance.

As a function for the distance in Eq. 3.6, we decided to use the root mean square deviation (RMSD):

$$\text{RMSD} = \sqrt{\frac{1}{p}\sum_{i=1}^{p}(\text{REF}_i - c_i)^2}, \tag{3.7}$$

where REF_i and c_i are the ith element of the reference and cluster centroid respectively. Larger errors have a disproportionately large effect on the RMSD, making it sensitive to outliers. Therefore, the cases in which one or more elements of the cluster centroid (i.e., the averaged storm characteristics) deviate significantly from the same elements in the REF are penalized.

As shown in Alg. 3.3, performance is determined based on the number of mutual elements (Eq. 3.5) and distance (Eq. 3.6) compared to the reference. Note that the performance evaluation is sequential and that we first check for the number of mutual events before considering the distance. We do this with the intention of prioritizing the result to be representative of the reference as much as possible, rather than risking it just being close to a potentially small subset of the reference.

Additionally, note that the algorithm is designed to identify the best clustering result with respect to REF from different subsets of the feature space. Hence, we argue that QSKM includes a feature selection. The final q features derived by the algorithm can be interpreted as the q most suited features from the feature space to characterize the reference. In Sect. 3.4, we will provide proof for this statement.

As already mentioned above, QSKM requires a few user-defined hyperparameters, which are:

- values for the number of subsets (J) and features (q),
- a value range for the number of clusters, i.e. $k = [k_1, \ldots, k_{\max}]$,
- the reference (REF).

The reference should consist of observations with similar properties. For example, in the context of this thesis REF consists of the 23 most severe windstorm events based on insured loss and highlighted by insurance experts (see Roberts et al., 2014). In this example, all windstorm events share the property of being "highly severe". The more similar the observations inside the reference, the more likely it is for QSKM to succeed in producing a representative cluster. Choosing appropriate values for J and q requires less expert knowledge than designing the reference. Still, these values should not be picked without consideration. Remember, J is part of the feature sampling and thus, in the context of finding the q most suited features, J should be sufficiently large such that $S_1^{n\times q}, \ldots, S_j^{n\times q}$ cover most (if not all) possible combinations of q features from $X^{n\times p}$. The selection of k is not directly shown in Alg. 3.3, but is included indirectly through the extended k-means algorithm (Alg. 3.2). When it comes to choosing the value range for k, it is less about determining the most likely number of clusters in the data and more about minimizing

computation time. The algorithm is designed to select the optimal number of cluster using silhouettes (Appendix A.1). Therefore, the value range can be arbitrary large. However, since each additional value of k increases the runtime of the algorithm, we suggest to start with a reasonably small value range k. For q, on the other hand, there is no rule of thumb. Therefore, we suggest to apply Alg. 3.3 for a set of q's. The results for the different q can be compared against each other using the cluster verification tools from Alg. 3.3. A practical example on how to chose J, q and k will be given in Section 3.4.1.

3.4 Application and validation of QSKM

3.4.1 Configuration of QSKM

The algorithm comes with three hyperparameters that have to be set by the user, namely the dimension of the feature space (q), the number of subsets (J) and a value range for the number of resulting clusters ($k = [k_1, \ldots, K]$). Each windstorm has 10 features (see Tab. 2.2). Here, we will iterate through a feature space with dimension 6 to 10, i.e. one application of QSKM for $q = 6$, one for $q = 7$ and so on. Remember, QSKM will draw J subsets of q features from the data, which is why J should be chosen large enough to cover most (if not all) possible combinations of q out of 10 features. For 6 out of 10 features there a 210 possible combinations. Since this is still a reasonable number of subsets, we decided to set J to all. In this case, the algorithm uses all possible combinations of subsets. In cases where the number of subsets would be to large, the user will have to set a number for J in which case the algorithm will randomly pick J different combinations. The hyperparameter k determines the possible number of clusters the algorithm should build and test. For example, $k \in [2, 3, 4]$ would mean that the clustering result can have 2-4 clusters. Here, we will set $k = 2, \ldots, 10$ to cover a wide range of potential clustering results.

In addition to the three hyperparameters, QSKM requires a reference (REF), i.e. a set of labels that classify one or more observations into a group. In the context of this thesis, REF has to be a set of windstorms that share at least one common characteristic. Here, we decided to use the top 23 windstorms highlighted by the XWS open access catalog of extreme European windstorms (Roberts et al., 2014). All of the 23 windstorms share the characteristic of being highly severe, which can be measured not only by different severity indices, but also by the large social and economic impacts they had at the time they occurred (see Roberts et al., 2014, for details). By selecting these 23 events as REF, we aim to create a single cluster that is representative of the top-end of extreme windstorms in Europe.

After setting the hyperparameters and REF, we apply QSKM to our tracked windstorms (see Sect. 2.3) for $q = 6, \ldots, 10$. Table 3.1 shows both our measures, the distance δ_j (Eq. 3.6) and maximum number of clustered events from the reference γ_j (Eq. 3.5), for different settings of q. Additionally, for the sake of comparison, we also include the same measures for a subset of events consisting of all windstorms that exceed the 90th percentile of the SSI. Note

TABLE 3.1: Comparison of different setups for QSKM, where q is the number of features QSKM had to chose. Shown are the distance (δ_j) of the cluster centroid to the reference centroid, the number of events from the reference (γ_j) in the cluster, the overall number of events in the cluster closest to the reference and the number of clusters (K). Additionally, the same measures are shown for all events exceeding the 90th percentile of the SSI.

Method	δ_j	$\gamma_j(\leq 23)$	No. of events	K
$QSKM_{q=10}$	119.3	18	590	2
$QSKM_{q=9}$	86.1	18	555	2
$QSKM_{q=8}$	97.7	20	580	3
$QSKM_{q=7}$	53.1	20	526	3
$QSKM_{q=6}$	171.3	21	681	3
$SSI_{90\%}$	439.0	12	224	-

that $q = 10$ means that QSKM is not forced to pick a subset of windstorm characteristics, as their are only 10 available, and thus reduces basically to a traditional application of k-means (Alg. 3.2) that checks the clustering results against REF. From Tab. 3.1 we can see how our two measures complement each other to provide a full picture of the optimization. If we would only look at γ_j, we would be tempted to select $q = 6$ as the best result or even think about further reducing q as Tab. 3.1 indicates a inverse relationship between both variables. However, δ_j has its minimum at $q = 7$ and maximum at $q = 6$. A closer look into the clustering result for $q = 6$ shows that this configuration has approximately 100 and 150 more events compared to $q = 7$ and $q = 8$, respectively, while also having 3 clusters. Therefore, events from REF are more likely to end up in the same cluster without having to be near the cluster center. Yet, it is also not enough to consider δ_j as the only relevant measure, as unlike γ_j the relationship between q and δ_j is not linear (see δ_j for $q = 8$ and $q = 9$ in Tab.3.1). Ins summary, our two measures indicate a trade-off in the configuration setup, where not all events from REF can be represented by one cluster without it becoming too general or diffuse.

As we have mentioned earlier, we prioritize γ_j over δ_j, since we want the result to be as representative as possible for REF. However, the configuration $q = 6$ showed that for maximizing γ_j, we end up having a less representative result. Thus, this type of priority only holds true within the respective setup (e.g., comparing the results of 6 random features against the results of 6 other random features). For a comparison between different configurations of QSKM, both measures have to be treated as a trade-off. Therefore, based on the results in Tab. 3.1, we decided to select $q = 7$ for QSKM, since this configuration includes almost all event from REF, while also having the lowest δ_j of all tested setups.

Finally, it is worth mentioning that all configurations of QSKM outperform the option of just selecting all events above a SSI threshold. This does not mean that the SSI is unrepresentative for extremely severe windstorm events. However, it indicates that the value alone is not necessarily enough

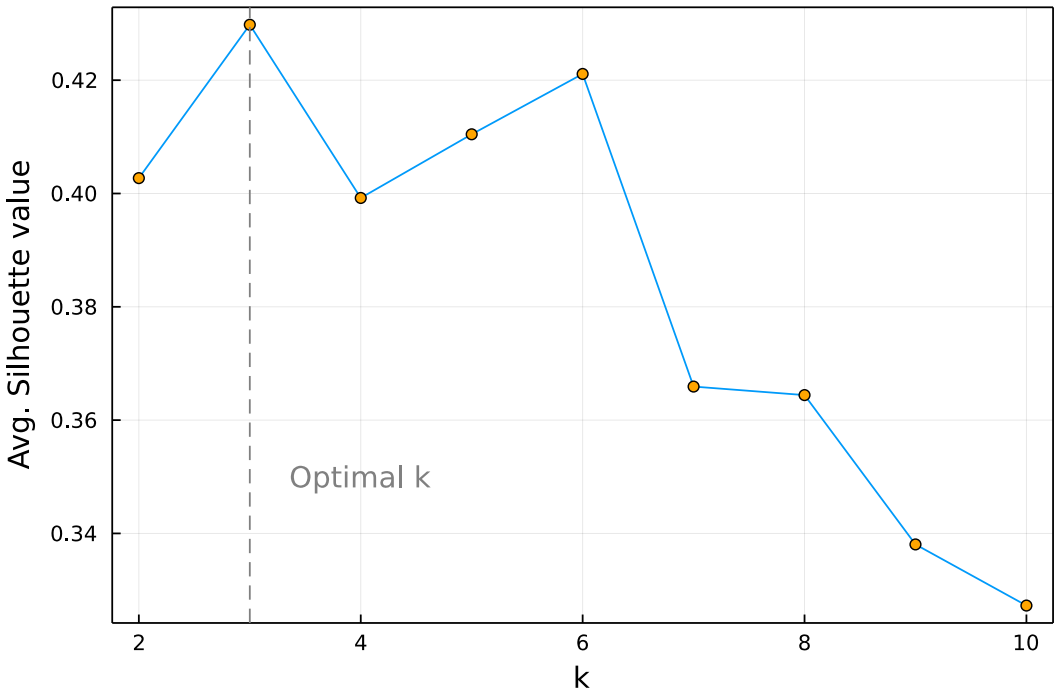

FIGURE 3.1: Averaged silhouette values for $k = 2, \ldots, 10$. The selected k is marked with a gray dashed line.

to identify them.

3.4.2 Analysis of the QSKM output

Now that we have selected $q = 7$ as the most suitable feature space, we can start evaluating QSKM. The algorithm has multiple outputs alongside the clusters, each of them provides additional information about the clustering result. Among them, the most interesting are the averaged silhouette values (Appendix A.1) for the selected range of k, the q features picked by QSKM and the cluster centers. Below, we will take a closer look at all three to get a better understanding of the clusters, and then move on to a direct comparison of the clustering results against REF.

According to the averaged silhouette values in Fig. 3.1, the optimal number of clusters for our setup is $k = 3$. However, the results are close to each other, as the values show only small differences between $k = 2$ and $k = 6$. From $k = 6$ onward, the values drop significantly. Considering that silhouettes are defined between -1 and 1, our maximum value of approximately 0.43 means that the majority of observation were assigned to the most suitable cluster. However, it still does not come close to the perfect value, so we have to expect some variance within the clusters which could affect our results at a later stage of this thesis.

A closer look at the cluster centers returned by QSKM (Fig. 3.2), shows what features define each cluster as well as the names of the seven features selected by the algorithm. From the ten initial features listed in Tab. 2.2, QSKM selected the seven that represent the intensity ($MEANV$), duration

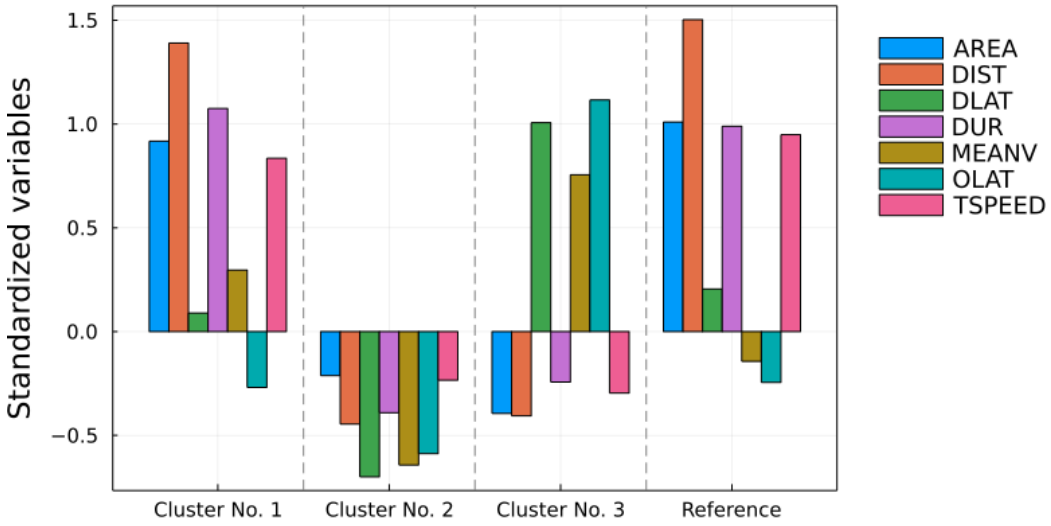

FIGURE 3.2: Standardized values (i.e. z-scores) of the cluster centers and the reference for the seven features selected by QSKM

(*DUR*), size (*AREA*), latitude of origin (*OLAT*) and destination (*DLAT*), travel speed of the event (*TSPEED*) and the travel distance (*DIST*). The individual cluster centeres indicated, that events in cluster No. 1 (hereafter denoted as CL1[2]) are primarily windstorms with long durations, larger size and high travel speed/distance, while those in CL3 show higher wind speeds and usually travel at higher latitudes (positive *OLAT* and *DLAT*). In contrast, events in CL2 have below-average values in all features representing magnitudes such as wind speed, size, or duration, and tend to occur at lower latitudes in the Northern Hemisphere (negative *OLAT* and *DLAT*). By comparing the resulting clusters to the center of REF, it becomes clear that CL1 is the cluster closest to REF. Both show strong similarities in almost all features. On average, only *MEANV* seems to be stronger for the events in CL1 than for those in REF.

Alongside the cluster centers in Fig. 3.2, the distribution of individual windstorm features (Fig. 3.3) further supports that CL1 is closest to REF. For *DUR* and *AREA* (Fig. 3.3a and b), CL1 and REF show very similar distributions, with the median and quartiles close to each other for most of the features. The features *DUR* and *AREA* of CL1 have both wider right tails, but are otherwise very good representations of REF. Features that are similarly well represented are *OLAT*, *DLAT*, *DIST* and *TSPEED* (see Appendix B.1). Even for features that are not from the seven picked by QSKM such as *OLON*, CL1 shows an overall better representation of REF than CL2 or CL3 (Fig. 3.3d). Only for *MEANV* it is hard to argue whether CL1 or CL2 is closer to REF. Still, Fig. 3.3 shows that QSKM is indeed able to define a representative cluster of events with respect to a given reference. To even further support this claim, we want to not only point out how close certain features of CL1 are to REF, but also how different the distributions of CL2 and CL3

[2]For simplicity, all cluster will be referred to as CL plus their respective number, i.e. cluster No. 3 is CL3 and so on.

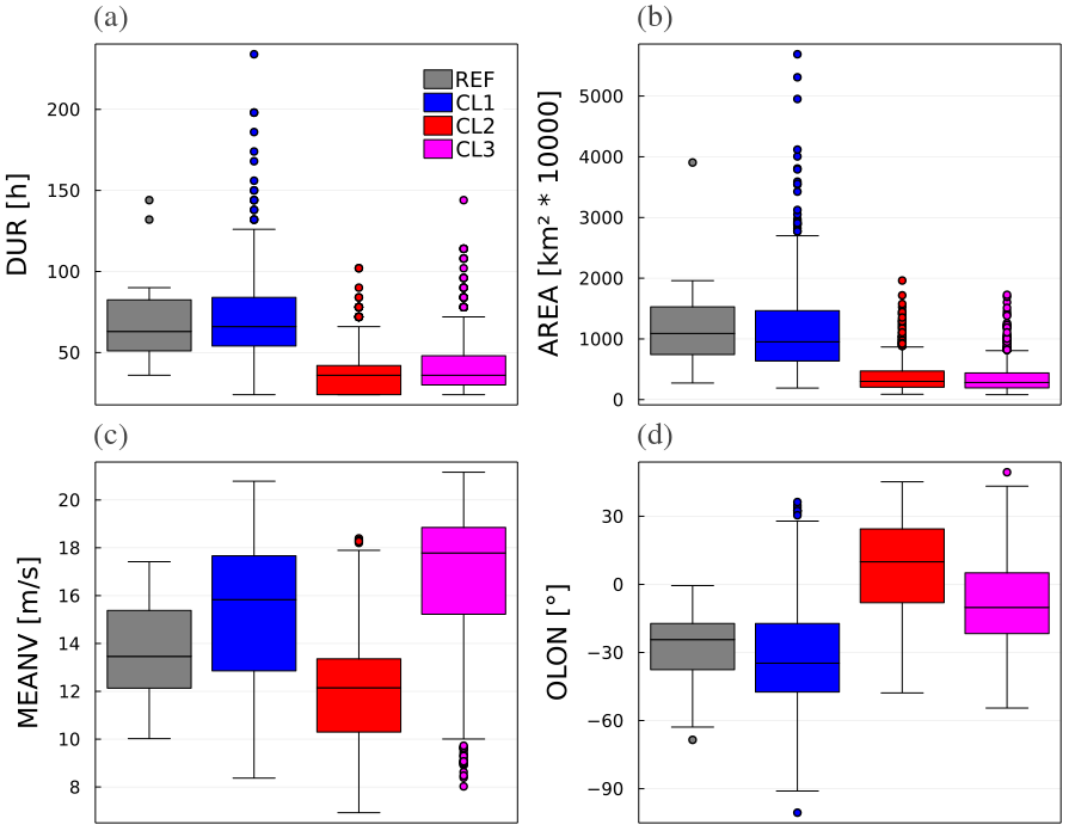

FIGURE 3.3: Boxplots of (a) duration, (b) size, (c) averaged wind speed and (d) longitude of origin for the elements in the three clusters of QSKM (CL1-CL3) and the reference (REF).

look in comparison. Especially for *DUR* and *AREA*, large parts of the distribution of CL2 and CL3 show significant smaller magnitudes compared to REF.

To summaries the characteristic differences between the clusters that are displayed in Fig. 3.2 and 3.3, Tab. 3.2 shows averaged values of various windstorm features for each cluster. On average, the wind speed of the events does not differ as much between the clusters as we would have expected from Fig. 3.2 or 3.3. Mean, minimum and maximum values of wind speed are very similar (less than 2 m/s) for CL1 and CL3, and even CL2 falls only

TABLE 3.2: Averaged values for different windstorm characteristics representing the intensity, size and duration of the events inside a certain clusters (No. 1–3). Peak is not an average, but the highest measured wind speed in the respective cluster.

No.	Wind speeds [m/s]			Duration [h]	Area [$km^2 \cdot 10^4$]	Peak [m/s]	SSI
	Mean	Min.	Max.				
1	15.18	8.37	21.03	77.70	1175.32	30.10	10.60
2	12.01	5.87	17.06	43.10	377.36	23.44	3.95
3	16.75	8.14	21.92	46.52	362.34	31.37	2.36

behind by approximately 3-4 m/s in the respective statistical measures. The peak wind speed measured in each cluster follows a similar pattern, only that the gap is larger between CL2 and the other two. Still, there is a considerable difference between a potential physical peak of more than 30 m/s and one below 25 m/s. However, even larger differences can be found for the duration, size and SSI. On average, windstorms in CL1 are 30 hours longer, have thrice the size and a significantly larger SSI than events from CL2 and CL3.

To obtain an overview of the primarily active regions of each cluster, Fig. 3.4 shows the respective storm tracks over Europe. Events from CL3 are primarily active over the northern parts of Europe and the Atlantic (Fig. 3.4c), with only a few windstorm entering Central Europe. In contrast, events from CL2 are more active in Central Europe and the Mediterranean region (Fig. 3.4b), with a noticeable higher track count in the latter. Both of the active regions for CL2 and CL3 could already be derived from $OLAT$ and $DLAT$ in the cluster centers (Fig. 3.2). However, it is still surprising how clear the regional differences are between both clusters. On the other hand, the tracks for CL2 fall in between and mix with the two previous cluster (Fig. 3.4a), where the majority of events can be found in the central parts of Europe and the Atlantic, but also, in a smaller amount, in the northern and southern parts of Europe. In comparison to REF (Fig. 3.4d), the higher track count of CL1 in Central Europe aligns with the results of the previous comparisons and further completes the picture of CL1 as a good representative for our extreme and severe windstorm events of the past.

Additionally, Fig. 3.4 also contains the number of events in each cluster. Almost half of our tracked windstorms are classified as CL2. This large amount of events, plus the weak wind speeds, small size and short duration shown in Tab. 3.2 and Fig. 3.2, would suggest that CL2 includes the overall least extreme events. However, this contradicts with the averaged SSI in Tab. 3.2, which is higher than the SSI of CL3. The remaining windstorm are almost fairly split into CL1 and CL3, where CL3 has only around 140 events more than CL1.

3.5 Summary and conclusion

In this chapter, we provided a comprehensive introduction of QSKM algorithm. We started by giving a detailed overview of a traditional k-means clustering approach, discussed its advantages as well as methodological flaws. In this way, we motivated our transition from k-means to QSKM, or, less generally, our transition from an regular unsupervised approach to a semi-supervised approach to clustering.

After its introduction, we applied QSKM to our tracked windstorms from Sect. 2.3 to illustrate its performance and evaluate the clustering. The results showed that

- only seven out of our ten windstorm features are necessary for the clustering ($MEANV$, DUR, $AREA$, $OLAT$, $DLAT$, $TSPEED$, $DIST$),
- QSKM is able to define a cluster similar to REF (that cluster is CL1),

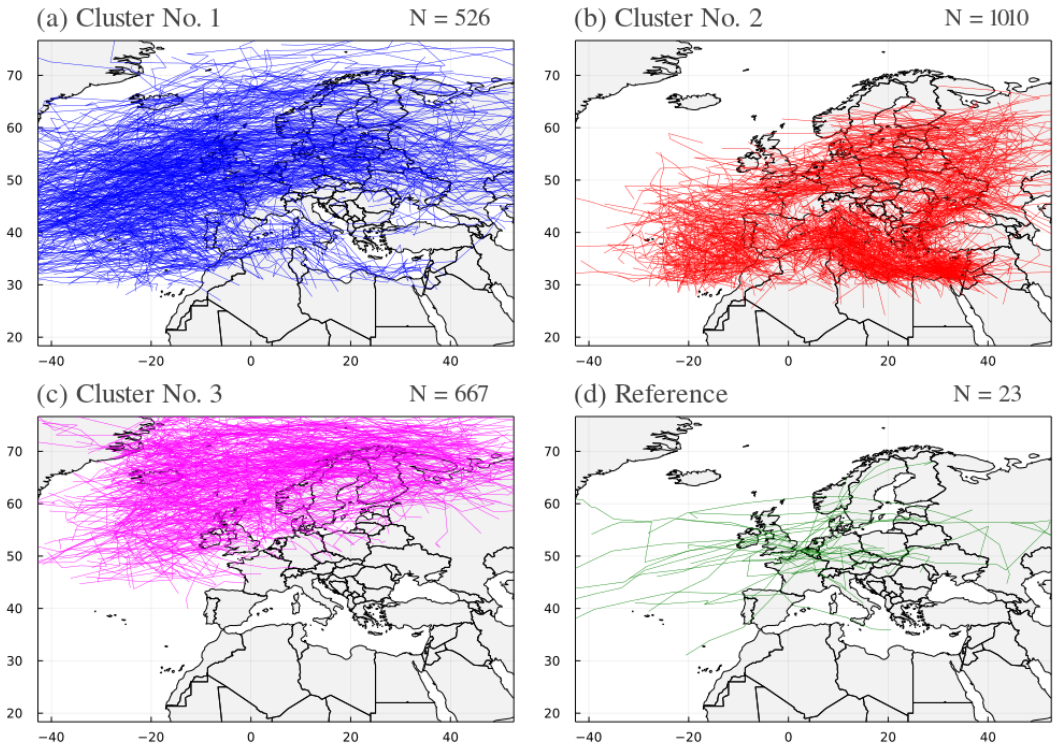

FIGURE 3.4: Windstorm tracks of (a-c) the different cluster and (d) the reference. The number in the top right is the number of tracks in the respective plot.

- there are 3 different classes of European winter windstorms, each of them having different prominent characteristics.

The further study of CL1 and REF indicated that highly severe and major events define themselves through a large size, long durations and far above average travel speed and distances. These type of events primarily affect Central Europe, but can also extend, in smaller numbers, towards the Mediterranean region or northern parts of Europe. In Chapter 4, we will comeback to the characteristics of CL1 to study their driving mechanisms and how they have change over the years.

In conclusion, the results are a good display of QSKMs advantages over a regular clustering approach. Thanks to QSKMs ability to cluster the data with respect to a reference, we were able to define a class of windstorms that met our requirements for a relevant subset of extreme events that includes both historically severe windstorms and a large number of events with similar characteristics. Additionally, the algorithm provided us with a tool to select relevant features from a pool of rather general characteristics of hazards. Arguably, a traditional k-means approach would have already produced a similar cluster from all of our ten defined characteristics (see QSKM$_{q=10}$ in Tab. 3.1). However, the cluster would have neither been as representative of REF as ours, nor would k-means have been able to identify key characteristics that determine the clustering result.

The only two disadvantages of QSKM are (*i*) the increased complexity of the algorithm and (*ii*) the high impact some hyperparameters can have on the

performance. The first point refers mainly to the significant increase in computation time due to a large number of additional k-means iteration caused by the feature sampling (step 1 in Alg. 3.3). Considering how the algorithm is designed, it is generally desirable to iterate through all possible combinations of a feature subset. However, dependent on the size of the feature space and the desired subset, the number of combinations can easily lead to an unreasonable high computation time. For example, on a Intel Core^{TM} i5-8250U CPU at 1.60GHz $\times$ 8 using the high performance programming language Julia (Bezanson et al., 2017), the computation time for 6 out of 10 features (i.e. $J = 210$ subsets) was approximately 7 minutes. If we extend this example to a simple case of 6 out of 13 features, we already end up with more than 1700 possible combinations and a projected run time of almost a hour.

The second disadvantage is a direct consequence of the first one. Since the computation time can increase significantly dependent on the dimension of the feature space, the choice of hyperparamters such as q (number of selected features from the feature space) and J (number of samples drawn from all possible combinations) can have a high impact on the clustering performance. To save computation time, users might test an insufficient number of setups for q and J, which could cause QSKM not to converge to its optimal result.

A semi-supervised approach comparable to QSKM, was proposed by Bair and Tibshirani (2004) who aimed to utilize gene expression data and the clinical data for the purpose of identifying subtypes of cancer. In their study, they used the clinical data as reference to identify a list of genes that correlate with the clinical variable of interest, to which they than apply an unsupervised clustering technique. The key methodological difference between the method of Bair and Tibshirani (2004) and QSKM is that they reduce their feature space based on the reference, i.e. the clinical variable of interest, before the actual clustering. Therefore, their method depends heavily on the quality of the reference and its ability to correctly represent the desired subset within the data. In contrast, QSKM only provides measures of how well a reduced feature space represents the reference. Theses measure are independent and thus allow a comparison between different configurations of QSKM. Hence, in QSKM the clustering is used to reduce the feature space.

In summary, we are convinced of QSKMs ability to produce a good subset of windstorms from our data that is representative of the top 23 windstorms highlighted by the XWS open access catalog of extreme European windstorms (Roberts et al., 2014). The presented results not only showed that events from CL1 and REF have remarkably similar characteristics, but also that the other two clusters display significant different magnitudes for multiple characteristics. Unlike regular unsupervised approaches, our clustering algorithm is able to select a subset of relevant characteristics from a pool of general characteristics of hazards, while also optimizing the clustering results towards a given reference. Therefore, our method has a methodological advantage compared to other clustering techniques.

Chapter 4

Evaluation of European windstorm classes

In Chapter 3, we introduced QSKM, a semi-supervised clustering method, and used it to identify three different windstorm classes (i.e. clusters) based on our in Chapter 2 defined windstorm characteristics and with respect to a reference catalog of severe European windstorms. Among those three classes, CL1 is the class closest to our chosen reference (20 out 23 windstorm events) and therefore represents a subset of potentially severe European windstorms. We also found that the three classes are governed by seven characteristics that describe the intensity, size, duration, travel speed and distance, as well as region of origin and impact.

In this chapter, the windstorm classes are further analyzed not only in terms of their characteristics and impacts on Europe, but also in terms of their connection to prominent large-scale atmospheric processes and how each of the classes may have changed over time. Therefore, this chapter primarily focuses on answering our second and third research question (Sect. 1.4). We start by providing a comprehensive overview of the potential impact, occurrence and temporal development of a windstorm from each of the respective classes (Sect. 4.1). Afterwards, for the purpose of understanding and quantifying the influence of large-scale atmospheric process on the type of windstorm, a random forest classification model (Appendix. A.2) is used to link the windstorm classes to characteristics of their parent cyclone, European teleconnection patterns and measures of baroclinic instability (Sect. 4.2). Finally, Sect. 4.3 covers a trend analysis where we investigate past changes in the occurrence of our windstorm classes. The chapter ends on a summary and discussion of the results (Sect. 4.4).

4.1 Impact, occurrence and temporal characteristics of the windstorm classes

In Sect. 3.4.2, we have already given a brief summary of the windstorm classes, their characteristics and how they differ from each other. We showed that CL1 primarily contains exceptionally large windstorm events with a long duration and high wind speed, and that it is the cluster closest to our reference. On the other hand, CL2 and CL3 consists of comparably smaller

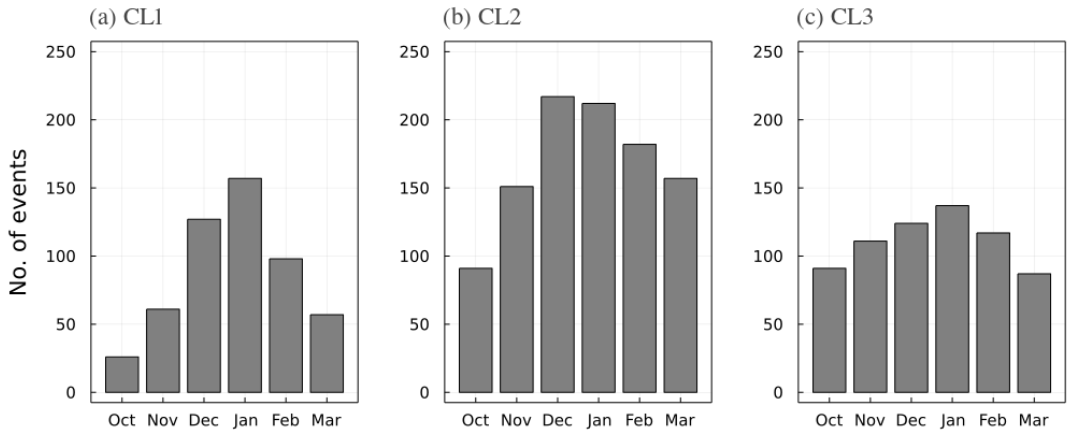

FIGURE 4.1: Number of occurred windstorm events in the respective months of our extended winter period October to March for the period 1981-2017, shown for (a) CL1, (b) CL2 and (c) CL3.

events with a shorter duration, where CL2 events have an overall lower wind speed, and CL3 events partially exceed the wind speeds in CL1 (see Tab. 3.2).

Here, however, we want to investigate our derived windstorm classes with respect to their occurrence and temporal structure, as well as quantify their impact on Europe using our characteristics from Tab. 2.2. We aim to provide a comprehensive overview of the inherent characteristics of our windstorm classes and how they compare to each other.

4.1.1 Occurrence

Regardless of the class, most of the windstorm events occur in December and January (Fig. 4.1), with the only notable difference being that CL1 and CL3 have more events in January than December, while CL2 has slightly more events in December. However, for CL2 and CL3 (Fig. 4.1b,c), the difference in the number of events in January and December is too small to assume or prove statistical significance. Overall, the event distribution is similar for CL1 and CL2 (Fig. 4.1a and b, respectively). In both classes windstorm events are considerably less frequent in October to November than for December, especially for CL1, where both months combined have not as much windstorm events than December alone. After they peak in December to January, the event numbers decline slowly towards March. Again, this decline is much more rapid for CL1 than for CL2. The event distribution for CL3 (Fig. 4.1c) is flatter from October through January, with a difference of less than 50 events between both months.

The differences in event numbers for October and November between CL3 and the other two are most likely due to the southward transition of the polar front in the winter months. As already discussed in Sec. 1.1, the polar front is a region of strong baroclinicty along which cyclones develop and intensify. In summer, the polar front can be found further north due to a weaker temperature gradient between warm subtropical and cold polar air masses. In October and November, the polar front is still in transition

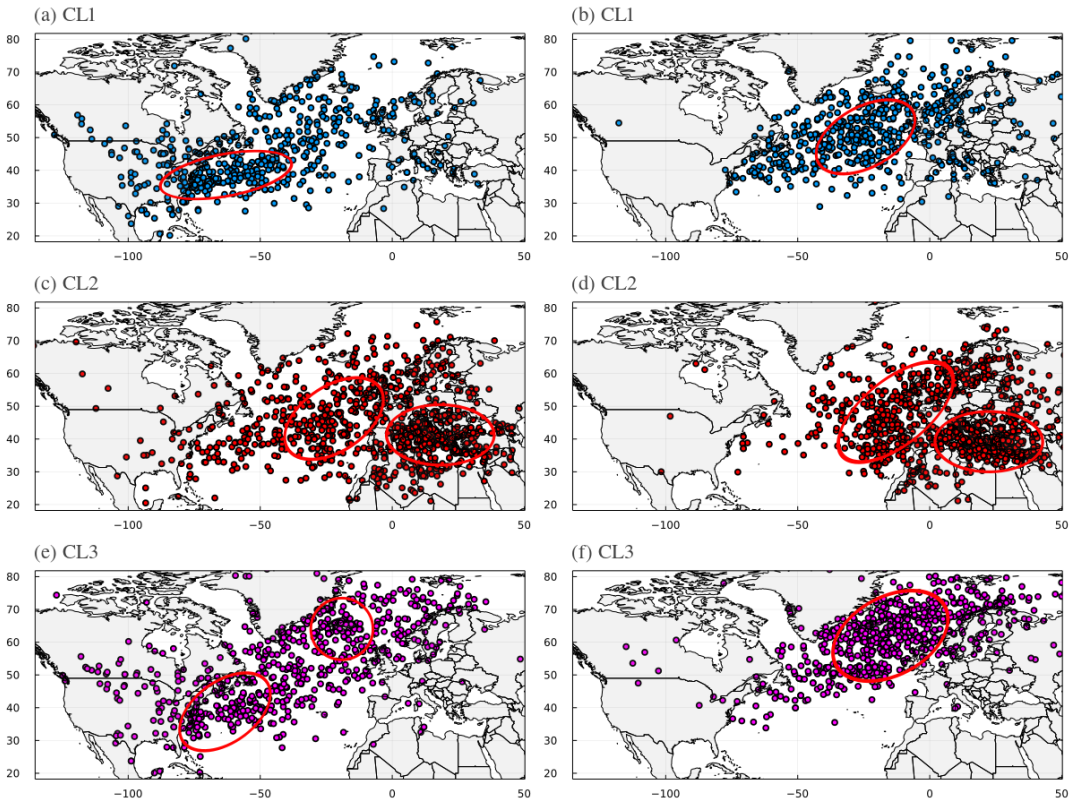

FIGURE 4.2: Locations of (a,c,e) cyclogenesis and (b,d,f) strongest intensification for (a,b) CL1, (c,d) CL2 and (e,f) CL3. The red circles mark regions where events cluster.

towards the south and therefore windstorm events are more likely to occur in the northern parts of Europe, which usually results in events of CL3 (see Fig. 3.4).

Analogously, the strength of the polar front could explain the differences in event numbers for the early winter months between CL1 and CL2. On average, the events of CL1 are larger and more intense than the events of CL2 (see Tab. 3.2). Therefore, it is likely that events of CL1 require a stronger polar front to even develop at all. The possible implications of a weaker or stronger polar front on the class of the occurring windstorm is further explored in Sect. 4.2.

Additionally to the temporal distribution of the occurrence of our windstorm classes, Fig. 4.2 shows the spatial distribution of two key processes of their parent cyclones, namely the cyclogenesis and intensification. The region of cyclogenesis is here defined as the coordinates of the cyclone center at the first time step of its tracking. Analogously, the region of intensification is defined as the coordinates of the cyclone center at the time step with the strongest decrease in minimum core pressure.

Although the regions of cyclogenesis and intensification are scattered all over of the North Atlantic, Europe and even the United States, there are areas where they cluster for each of the classes and thus indicate preferred regions for both processes. For events of CL1, the cyclogenesis (Fig. 4.2a) seems to start primarily along the east coast of the United States and the western parts

of the North Atlantic in general. The intensification (Fig. 4.2b) then usually peaks in the middle or eastern parts of the North Atlantic. A completely different picture can be seen for the events of CL2, where cyclogenesis primarily starts in the eastern parts of the North Atlantic and Mediterranean region (Fig. 4.2c) and intensification seems to happen shortly after along the west coast of Europe or the eastern parts of the Mediterranean, respectively (Fig. 4.2d). The regions of cyclogenesis for CL3 events (Fig. 4.2e) are loosely scattered across the North Atlantic, with only smaller clusters in the western parts of the North Atlantic, near the east coast of the United States, and around Iceland. The preferred region of intensification, on the other hand, is characterized by basically a single large cluster between southeast of Greenland and Europe (Fig. 4.2f).

Our results from Fig. 4.2 are consistent with the findings of Moran (2019), who have shown that windstorms with large footprints in the European region (i.e., events as in CL1) often originate along the east coast of the United States, while those with comparatively smaller footprints (i.e., events as in CL2) are more likely to originate in the western parts of the North Atlantic. Analogously, Moran (2019) have also shown that events with a larger footprint tend to have their intensification phase over the western parts of the North Atlantic, while windstorms with a small footprint often further intensify near the European coast or in the Mediterranean region.

4.1.2 Temporal development

For the purpose of understanding the temporal development of a windstorm event from our identified classes, Fig. 4.3 shows the averaged 10 m wind speed and size of the windstorm over its lifetime, as well as the respective minimum core pressure of the parent cyclone, for each of the classes. The boxes in each panel are calculated from all events of the respective class at the given time step. For each box, at least 10 events were used for the calculation. Later time steps that could not reach this threshold were removed from the figure.

When comparing the averaged wind speed (Fig. 4.3a-c), one can see that the events of CL1 show an overall drop in magnitude for the median with increasing lifetime, while the median for CL2 and CL3 hardly changes. This is most likely due to fact that most CL1 events develop in the North Atlantic, a region with low surface friction, and transition over land, a region with high variation in surface friction, during their lifetime, while the events of CL2 and CL3 often remain over the respective water body over which they originated (see windstorm tracks in Fig. 3.4). Therefore, CL1 events have their highest wind speeds in the early hours of their lifetime and are probably the most dangerous when they hit land. There is also a small decrease in the median wind speed for CL3 events (Fig. 4.3c) between 42-60 hours accompanied by an increase in variance. This suggest that most CL3 events disappear in this time frame, which is further supported by the averaged event duration of 46.52 hour for CL3 in Tab. 3.2.

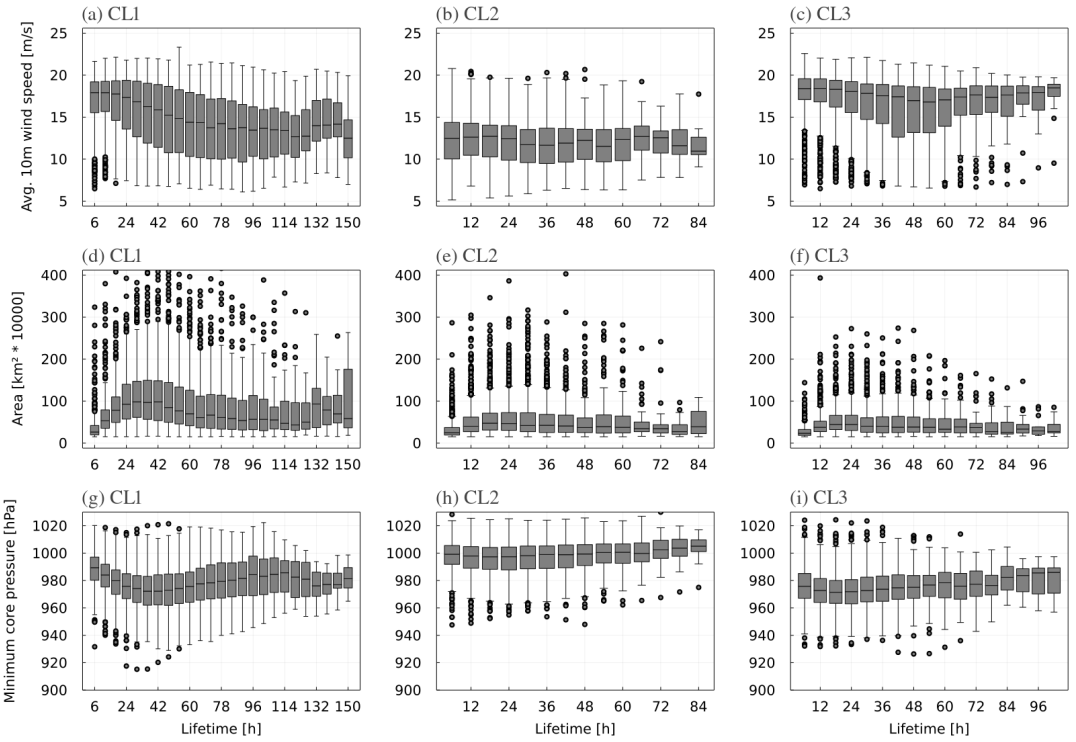

FIGURE 4.3: Temporal development of the windstorm events of (a,d,g) CL1, (b,e,h) CL2 and (c,f,i) CL3 over its lifetime, characterized by (a-c) the averaged 10 m wind speed, (d-f) the size of the event and (g-i) the minimum core pressure of the respective parent cyclones. Each box is calculated from at least 10 events.

For the size of the windstorms (Fig. 4.3d-f), again, the magnitudes vary stronger among CL1 events, but the overall temporal structure is more comparable between the classes. Windstorms of CL2 and CL3 usually reach their maximum size within the first 18-24 hours. For CL1 events, this time frame shifts to 30-42 hours. The fact that they reach their maximum size later than those of CL2 and CL3 suggest that CL1 windstorms are either longer effect by cyclone growth factors such as the jet stream or upper-level divergence, or spend more of their lifetime over areas with low friction such as the ocean. This would also explain the overall higher severity of CL1 events. Afterwards, the events continuously decrease in size (see median line in Fig. 4.3d-f). Both, CL1 and CL2 (Fig. 4.3d and e, respectively), show an increase in windstorm size in the later time steps, which in both cases can be traced back to exceptional long and large events that skew theses boxes towards larger magnitudes.

Figure 4.3g-i indicates that the minimum core pressure of the parent cyclone is inversely related to the windstorm size. Meaning, for CL2 and CL3, the minimum core pressure reaches its minimum between 18-24 hours, and for CL1 at approximately 30-42 hours. Afterwards, the minimum core pressure continuously increases until the cyclone disappears. These changes are again much more prominent for CL1 than for the other two classes.

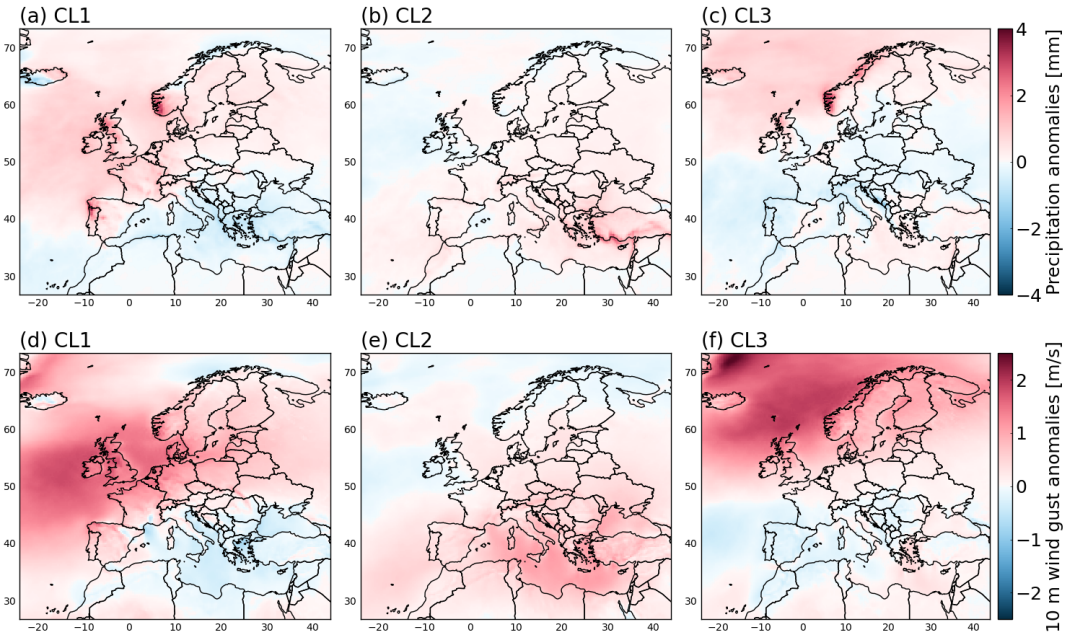

FIGURE 4.4: Averaged anomalies in (a-c) precipitation and (d-f) 10 m wind gusts in case of the occurrence of an windstorm events of (a,d) CL1, (b,e) CL2 or (c,f) CL3 for each grid box. The anomalies are calculated with respect to the baseline 1981-2010.

4.1.3 Impact

Windstorms are defined as extreme winds that are strong enough to cause damage to trees or buildings in the affected area (Pielke, 2007), but sometimes they are also accompanied by heavy precipitation. Due to the major socio-economical impact that extremes of both variables can produce, it is important to look at their potential magnitudes in case of a windstorm event. Local impacts are of particular interest in this regard, since wind and precipitation values averaged over the storm object are not representative of what is happening in a particular country or region. Therefore, Fig. 4.4 shows maps of anomalies for the daily maximum 10 m wind gusts and precipitation amount, averaged over the days a windstorm occurred within the respective grid boxes, for each of the three classes. The daily anomalies were calculated with respect to the climate baseline period 1981-2010.

Independent of the meteorological variable, the positive anomalies can be associated with the tracks of the respective windstorm class (compare to Fig. 3.4). Windstorm events of CL1 cause above-average precipitation and wind gusts in central and parts of northern Europe (see Fig. 4.4a and d), especially in the United Kingdom, which not only has positive anomalies in precipitation but also by far the strongest wind gusts anomalies compared to any other European land mass. Other maxima in precipitation can be found at the south coast of Norway and the coast of Portugal. For CL2 events, there are weak precipitation anomalies in central and south Europe (Fig. 4.4b). Stronger precipitation anomalies can be found at the southern coast of Turkey. Positive anomalies in 10 m wind gust are also weaker compared to CL1 and CL3 events and are primarily distributed across central

and southern Europe, where the largest anomalies can found across Mediterranean Sea (Fig. 4.4e). For CL3 events, larger than average precipitation amounts can be found over the North Atlantic, which primarily affects Scotland, Iceland, and the Scandinavian region, especially the west coast of Norway, which shows the largest precipitation anomalies (Fig. 4.4c). Weaker, but still positive, precipitation anomalies can also be found in the western parts of the Mediterranean Sea. The same applies for 10 m wind gust (Fig. 4.4f), except that the positive anomalies in the northern parts of Europe extend further south and now also affect the whole of the United Kingdom and the North and Baltic Seas.

Independent of the class, the negative anomalies in precipitation and wind gusts are found in the opposite cardinal direction to the positive anomalies, i.e. if there are mainly positive anomalies in the north, then most anomalies in the south are negative (and vice versa). A possible explanation for this pattern could be the North Atlantic Oscillation (NAO) and its influence on the windstorm occurrence of Europe. During its positive phase, the NAO can be associated with an increase in windstorm activity over central Europe and the eastern Mediterranean region (Donat et al., 2010; Nissen et al., 2010) and a decrease in the western parts of the Mediterranean (Nissen et al., 2010). Similar applies for precipitation. Hence, the anomaly patterns for CL1 (Fig 4.4a,d) and CL3 (Fig 4.4c,f) resembles the windstorm activity described in the literature for the positive NAO phase. However, there is no link between a negative NAO phase and an increase in windstorm activity over the eastern parts of the Mediterranean region and we therefore lack an explanation for the anomaly patterns for CL2.

Another component to the potential impact of a windstorm is the duration of the event and the available response time for citizens in the affected area. Therefore, Fig. 4.5 shows maps of windstorm duration and speed of onset, i.e. the time required for the event to reach its peak (here defined as the time between the first and highest exceedance of the 98^{th} percentile). Each grid box in Fig. 4.5 shows averaged numbers of hours the grid box is affected (Fig. 4.5a-c) in case of the occurrence of an event at the very same location and how fast this event usually reaches its peak (Fig. 4.5d-f). It should be noted that not all windstorm classes affect the entire study area and therefore in some areas the speed of onset becomes zero just because no event of the studied class occurred in the respective grid boxes (e.g., south Europe in Fig. 4.5f for CL2). However, these grid boxes are easy to identify, because the same grid boxes show no duration in the respective figures for the duration of an event as well.

In general, the local windstorm durations (Fig. 4.5a-c) peak over the water bodies southwest of Spain, the Mediterranean Sea, and north of Iceland, while over land the longest durations can be primarily found in central and eastern Europe. Due to their overall longer lifetime in comparison to events of CL2 and CL3, CL1 events are responsible for the majority of these peaks across Europe. Outside the peaks, there seems to be a southwest to northeast gradient in duration for all three classes, where coastal regions tend to have shorter durations compared to the inland, which is particularly interesting

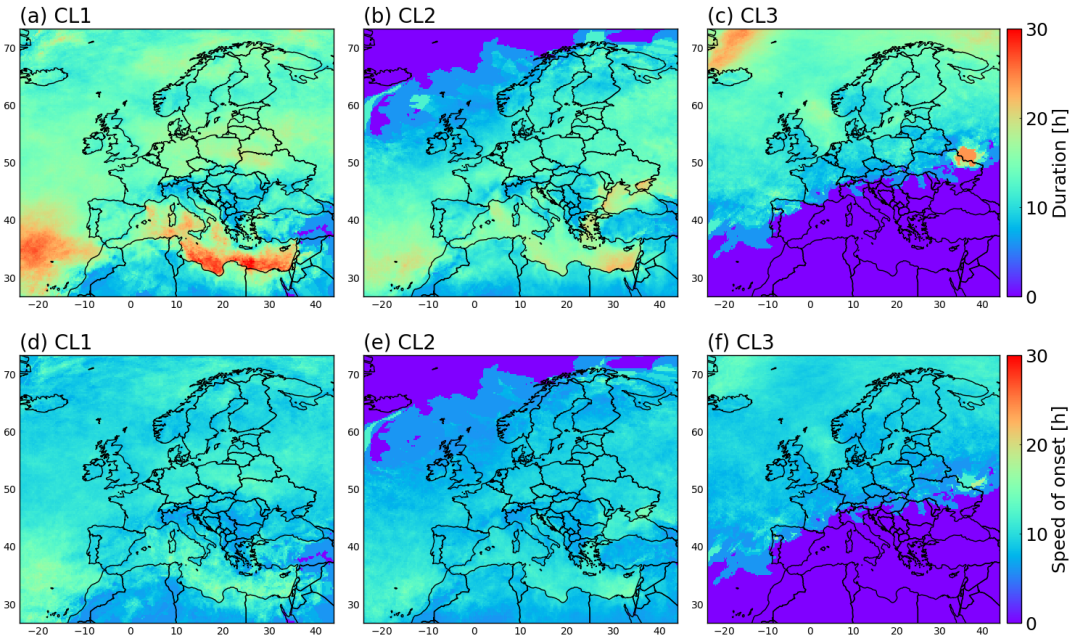

FIGURE 4.5: Averaged (a-c) duration and (d-f) speed of onset in case of the occurrence of an windstorm event of (a,d) CL1, (b,e) CL2 or (c,f) CL3 for each grid box. When a box is 0 (purple), no event occurred at this grid box.

in the case of CL1, where precipitation and wind gusts show a completely opposite pattern (Fig. 4.4a,d). This shows that although different regions are affected by the same type of events, they still face different aspects of their characteristics. For example, in a direct comparison of the United kingdom and Germany, the United Kingdom would be stronger impacted by CL1 events with respect to wind gusts and precipitation, but all within a shorter duration. However, that said local windstorm durations above 20 hours are rare and thus the respective regions in Fig, 4.5a-c are probably the result of a few very long storms.

The speed of onset (Fig. 4.5d-f) mirrors the patterns of the duration for each of the three classes, and shows that, in general, the longer an area is affected by a windstorm, the later it will experience its maximum wind speed. This rather simple proportional relationship between duration and speed of onset could be due to the size of the event. Imagine the windstorm at a given time as a cell with a gradient of wind speed, with high wind speeds in the center and comparably low wind speeds at the edges. Given this scenario, a grid box would first experience the lower wind speeds at the edges, which would continuously increase with time as the windstorm center travels towards the grid box. The larger the event, the longer this process would take. Overall, the maximum wind speed does not seem to be a good variable for defining the speed of onset. Given the interpretation of the speed of onset as a measure of available response time, it would be better if we had the time between the forecast of the event and the date of impact. However, this is outside the scope of this study.

4.2 Identification and quantification of the large-scale atmospheric processes behind the windstorm classes

In this section, we aim to identify and quantify the impact of large-scale atmospheric drivers such as the North Atlantic oscillation or the jet stream on the windstorm classes we previously derived from the ERA5 windstorm tracks using QSKM (see Sect. 3.4). In the following subsections, we present a set of features for various large-scale atmospheric factors associated in the literature with the development, occurrence and intensification of European winter windstorms (Sect. 4.2.1). Afterwards, we train a random forest classifier (Appendix A.2) between our windstorm classes and the aforementioned features. With the help of the random forest model, we can identify and quantify the impact of large-scale features on the windstorm classes. Overall, our goal is to answer our second research question (see Sect. 1.4) and identify atmospheric drivers that are important for our windstorm classes and to what extent they help to shape them.

4.2.1 Motivation and calculation of large-scale atmospheric drivers

A key challenge in windstorm classification is the selection, design, and calculation of features that serve as proxies for our desired large-scale atmospheric drivers. The pool of potential candidates ranges from indices describing modes of teleconnections such as the North Atlantic or Arctic oscillation to the intensity or location of the jet stream, or the depth, magnitude and location of atmospheric depressions. However, identifying suitable predictors is only the first task, calculating them is the second. Calculating the selected predictors is not trivial and requires careful considerations of what information we want to extract from the various atmospheric drivers. In the following, we will give a brief motivation and description of our selected predictors. A summary of the variable names and a short description of our selected predictors is given in Tab. 4.1.

Teleconnenction patterns of the Northern Hemisphere

The influence of teleconnection patterns on European winter windstorms has been extensively investigated and demonstrated in previous studies (e.g. Pinto et al., 2009; Donat et al., 2010; Walz et al., 2018). Given the influence of teleconnections on the occurrence of windstorms, it makes sense to include them in our classification model to see if they are also determinant for the class of windstorms as well.

We include five teleconnetion patterns that are influential for weather and climate variability in Europe, namely the North Atlantic Oscillation (NAO), East Atlantic pattern (EA), East Atlantic/Western Russia pattern (EA/WR), Scandinavia pattern (SCA) and Polar/Eurasia pattern (POL). The patterns

TABLE 4.1: Abbreviates of the large scale features used in our classification model and a short description of their definition.

Variable names	Description
NAO	North Atlantic Oscillation pattern index
EA	East Atlantic pattern Index
EA/WR	East Atlantic/Western Russia pattern index
SCA	Scandinavia pattern Index
POL	Polar/Eurasia pattern Index
PMIN	Minimum core pressure of matched cyclone
LAT_P, LON_P	Latitude/Longitude of PMIN
LAT_O, LON_O	Latitude/Longitude of cyclone origin
JET	Wind speed anomalies at 250 hPa, Avg. around footprint center
PJET	Footprint position w.r.t max. wind speed at 250 hPa
EGR	Eady growth rate between 850-500 hPa
EPT	Equivalent-potential temperature at 850 hPa

and their respective indices are estimated following the approach of the Climate Prediction Center[3] (CPC). Their procedure is an extension of the work from Barnston and Livezey (1987) and aims to isolate the primary teleconnection patterns using rotated principal component analysis of monthly mean standardized 500 hPa geopotential height anomalies over the Northern Hemisphere for a baseline climate period 1950-2000. The monthly patterns are projected onto daily data to create daily time series of pattern indices.

For the classification model, we calculate the daily pattern indices and average them over the duration of the respective windstorm event. A key difference in this thesis is that we isolate the desired patterns from the ERA5 data (Sect. 2.1), rather than the NCEP reanalysis (Kalnay et al., 1996) used by the CPC, for the baseline climate period 1981-2010. For the monthly indices, this only had a minor impact. A comparison of our estimated indices and those of the CPC showed an average correlation of 96% between them. The daily indices, on the other hand, showed almost no correlation. To a certain degree, this was to be expected, because ERA5 is a newer product with a significantly larger temporal and spatial resolution than NCEP, which should primarily impact the smaller time scales such as hours and days. Since our windstorms are tracked from ERA5, we are inclined to rather rely on our indices than those of the CPC.

Characteristics of the parent cyclone

In Section 2.3, we assigned each of the tracked windstorms to a parent cyclone to examine whether and how their characteristics affect the classification. Dependent on the study field or interest, cyclones have a similar arbitrary amount of characteristics as their footprints, ranging from statistical

[3]A detailed explanation of their methodology is given under https://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml

properties such as the size, duration and intensity, to more complex information that describe their development or explosiveness.

For our classification, we decided to specifically focus on three character traits of the parent cyclone: its intensity, location and origin. In general, the minimum core pressure (p_{min}) of the cyclone is a good proxy for its intensity. The lower the core pressure, the deeper and more extreme the cyclone usually is. The cyclone identification and tracking algorithm used in this thesis (see Sect. 2.2.2) comes with a p_{min} for each time step of the cyclone track. Here, we take the smallest p_{min} as proxy for the cyclone intensity, i.e.

$$\text{PMIN} = \min\left(p_{\text{min},1}, \ldots, p_{\text{min},t}\right), \tag{4.1}$$

where t is the lifetime of the cyclone. For the location we will use the respective longitude (LON_P) and latitude (LAT_P) of PMIN. By doing so, we hope to add more complexity to PMIN. For example, we do not want that a very deep cyclone in the West Atlantic has the same weight in the model than a similar cyclone in the East Atlantic, which is closer to (and part of) our study area.

The origin refers to the region of cyclogenesis. Severe cyclones that show an explosive intensification usually originate in the West Atlantic (Gray and Dacre, 2006; Allen et al., 2010). However, Dacre and Gray (2009) also point out that type C cyclones (Deveson et al., 2002), which are cyclones with strong upper-level forcing but a very weak low-level baroclinicity, dominantly origin in the Mid to East Atlantic. Furthermore, Fig. 4.2 already showed that the classes have preferred regions of cyclogenesis. Therefore, a distinction between the origin of the different cyclone types should significantly improve the model quality. Same as for Fig. 4.2, we will use the longitude (LON_O) and latitude (LAT_O) of the first time step of the cyclone track as proxy for its origin. Note that this is not the same origin as defined for windstorms in Sect. 2.4, since there can be a considerable time lag of multiple hours between the first detection of the cyclone and the windstorm (see Moran, 2019).

All of the above defined characteristics are directly or indirectly part of the output from the cyclone tracking (Sect. 2.2.2) and therefore non of them needs to be calculated. However, this also means that, unlike most other features described in this section, they are calculated from ERAINT (Sec. 2.1) and not ERA5.

The jet stream

The jet stream is a meandering, long and narrow band of high-speed winds that typically flow eastward in the middle and upper troposphere. In the Norther Hemisphere, the jet stream is related to the polar front, a region of strong baroclinicty which plays a major role in the development and intensification of cyclones.

Numerous studies have shown that the jet stream is an important contributor in the development of deep cyclones and extreme windstrom events in the North Atlantic. For example, Pirret et al. (2017) and Priestley et al. (2017)

showed that the relative positioning of the cyclone to the jet stream has a major influence on its deepening and propagation. Their results are consistent with several other studies that have shown that cyclones experience rapid deepening in the left jet exit (e.g. Baehr et al., 1999; Ulbrich et al., 2001), a region of intense upper-air divergence. Furthermore, due to the physical relationship to the polar front, the strength of the jet stream is a good proxy for baroclinic instability, since both are directly related to large-scale meridional temperature differences between the subtropics and the polar region (e.g. Eady, 1949).

Given its influential role, we include the strength of the jet stream (JET) and the relative positioning of the footprint to the jet steam (PJET) in our classification. Here, JET is calculated as the averaged wind speed anomalies ($v^{'}$) at 250 hPa in a 1000 km radius (r) around the windstorm center at his highest severity, i.e. the time step of the maximum SSI ($t_{\max}$):

$$\mathrm{JET} = \frac{1}{|\boldsymbol{v}^{'}|} \sum_{i} \boldsymbol{v}_{\boldsymbol{i}}^{'}, \quad i = 1, \ldots, |\boldsymbol{v}^{'}|, \tag{4.2}$$

where $\boldsymbol{v}^{'} = \{v_{r,t}^{'} | r \leq 1000km \text{ and } t = t_{\max}\}$ and $|\boldsymbol{v}^{'}|$ is the number of elements in $\boldsymbol{v}^{'}$. The wind speed anomalies are estimated with respect to the baseline climate period 1981-2010 using daily averages of 6 hourly fields.

We define PJET as the difference between the latitudes of maximum wind speed at 250 hPa (LAT_v) and the windstorm center (LAT_c), both again at the time step of the maximum SSI

$$\mathrm{PJET} = \mathrm{LAT}_{v,t_{\max}} - \mathrm{LAT}_{c,t_{\max}}. \tag{4.3}$$

This way, PJET indicates not only the relative distance between jet stream and windstorm, but also whether the event is located north (PJET < 0) or south (PJET > 0) of the jet.

Eady growth rate and equivalent-potential temperature

The maximum Eady growth rate (σ; Lindzen and Farrell, 1980) is a measure frequently used to describe baroclinic instability in the atmosphere and is defined as

$$\sigma = 0.31 f |\partial v / \partial z| N^{-1}, \tag{4.4}$$

where f is the Coriolis parameter, $|\partial v / \partial z|$ the vertical wind shear, and N the Brunt-Väisälä frequency. Barolclinic instability is widely excepted as the primary mechanism by which cyclones develop in the midlatitudes and σ has been successfully used as a diagnostic tool for baroclinicty in various studies (e.g. Ulbrich et al., 2001; Pinto et al., 2009; Willison et al., 2015). Here, we include σ in our classification model to study to what degree baroclinic instability is responsible for different types of windstorms. Following Pinto et al. (2009) and Moran (2019), we calculate σ from 6-hourly fields for the atmospheric layer between 850-500 hPa and apply a 3-day running mean over the times series to account for the high-frequency variability of σ. Afterwards,

the time series is averaged to daily means at each grid box. Same as JET (see Sect. 4.2.1), we calculate the model value (EGR) as a local average in a 1000 km radius (r) around the windstorm center on the day where the SSI peaks ($t_{\max}$):

$$\mathrm{EGR} = \frac{1}{|\sigma|}\sum_i \sigma_i, \quad i = 1, \ldots, |\sigma|, \tag{4.5}$$

where $\sigma = \{\sigma_{r,t} | r \leq 1000km \text{ and } t = t_{\max}\}$ and $|\sigma|$ is the number of elements in σ.

The equivalent-potential temperature (θ_e) is the temperature that an air parcel would reach if all the water vapor it contains were to condense, releasing its latent heat in the process, and it were adiabatically brought to a standard reference pressure (here 1000 hPa). Latent heat is a crucial component in the lifting process of an air parcel. The more moisture the air parcel contains, the more latent heat it can release and θ_e increases. Therefore, θ_e is a good indicator for atmospheric instability, where higher values of θ_e indicate less stable the atmosphere. Pinto et al. (2009) studied the contribution of θ_e to cyclone intensification by comparing non-extreme and extreme cyclones and found that the latter generally had higher θ_e values. It will be interesting to see whether or not θ_e is also indicative for a certain class of windstorm. We calculate θ_e at 850 hPa according to a formula given by Bolton (1980) using 6-hourly fields and average the results to a time series of daily means at each grid box. For our model, we again use the local average in a 1000 km radius around the windstorm center on the day where the SSI peaks, i.e.

$$\mathrm{EPT} = \frac{1}{|\boldsymbol{\theta_e}|}\sum_i \boldsymbol{\theta_{e,i}}, \quad i = 1, \ldots, |\boldsymbol{\theta_e}|, \tag{4.6}$$

where $\boldsymbol{\theta_e} = \{\theta_{e,r,t} | r \leq 1000km \text{ and } t = t_{\max}\}$ and $|\boldsymbol{\theta_e}|$ is the number of elements in $\boldsymbol{\theta_e}$

Other potential predictors

In addition to the predictors listed in Tab. 4.1, there are numerous other potential candidates that we could include in our classification model. For example, Walz et al. (2018) studied the impact of 20 different Large-scale features on serial-clustering of windstorms for different regions of Europe. Their list of features includes features such as teleconnection patterns, sea ice cover, land and sea temperature differences, and various other indices for a number of oscillation patterns. Pinto et al. (2009) investigated not only the influence of the jet stream, EGR and latent heat on cyclone development, but also the effects of upper-air divergence. Moran (2019) studied the role of land-sea contrasts and gradients of sea surface temperature as cyclonic growth factors in the West Atlantic.

However, for reasons of model complexity and in light of our research questions, we decided to reduce the number of features and include only variables that are either directly related to, or generally known to be involved in, the occurrence, propagation, development or intensification of European

TABLE 4.2: Confusion matrix for the RF model tested on the last 30% of the data, i.e. the last 661 of 2203 windstorm events. Also shown are the precision and recall for each class.

	Ground truth			
Predicted	CL1	CL2	CL3	Precision
CL1	95	25	30	0.63
CL2	40	269	9	0.85
CL3	27	14	152	0.79
Recall	0.59	0.87	0.80	

windstorms. Table 4.1 contains features that are commonly used in the literature to describe all of the above properties. In addition, we have added cyclone characteristics as diagnostic tool for the relationship between the windstorm and its parent cyclone.

4.2.2 Model tuning and verification

We fit a random forest (RF; Appendix A.2) classification model to the windstorm clusters identified by QSKM (see Sect. 3.4) using the the features from Tab. 4.1. For this task, we use the RF classifier of the *MLJ* package (Blaom et al., 2020), a package for composable machine learning for the programming language Julia (Bezanson et al., 2017). We tune the classifier with respect to three hyperparameters, namely the number of trees (between 100-500), the maximal depth of a tree (between 4-20) and the number of features used at each split (between 4-10). The different possible combinations of hyperparameters are each tested by 3-fold cross validation using the *accuracy* (Appendix A.3, Eq. A.7), the fraction of correct predictions, as a performance metric. The accuracy is a commonly used metric in machine learning to get an overview of the general performance of the model. The complete tuning process is performed on the first 70% of the data (1542 out of 2203 windstorm events). This split is done without shuffling of the events so that the results can be reproduced. After the best hyperparameters are determined by tuning, we use the tuned classifier to predict the classes of the remaining 30% of the data (661 of 2203 storm events). All results in this section are based on the performance of the tuned RF classifier in predicting last 30% of the windstorms.

The confusion matrix (Appendix A.3) in Tab. 4.2 shows that the tuned RF classifier performs well with an overall accuracy of 78% correctly predicted windstorm classes. However, Tab. 4.2 shows that the disparity in model performance is rather large between the classes. The *recall* (Appendix A.3, Eq. A.10), the fraction of a correctly predicted class out of all observations of

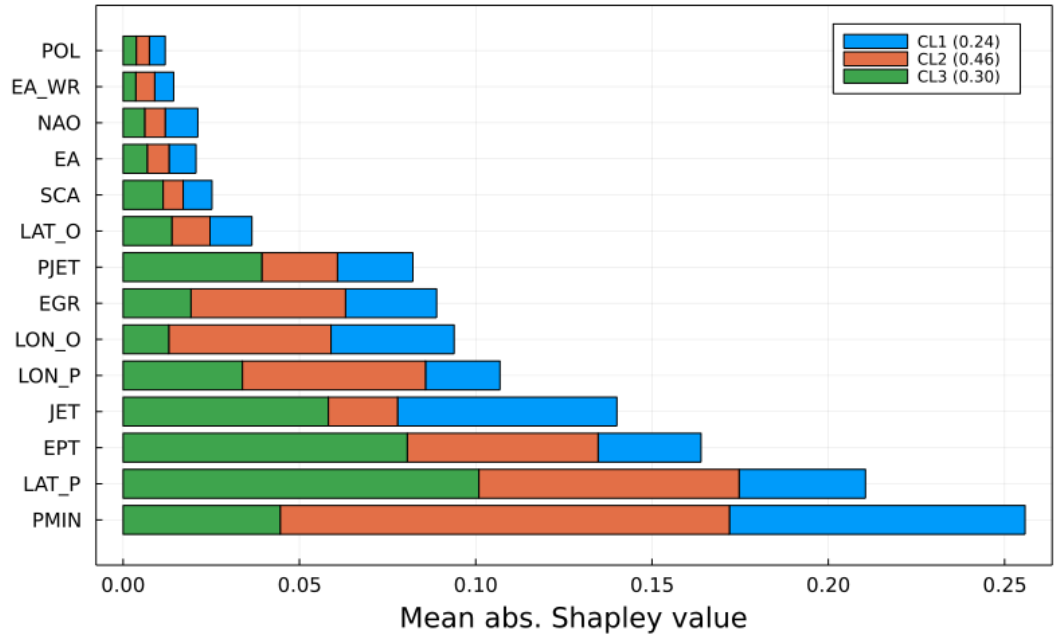

FIGURE 4.6: Feature importance for our features from Tab. 4.1 derived from Shapley values. The larger the mean absolute (abs.) Shapley value, the more important feature. The different colors show the importance of the feature for each class. Combined they show the global feature importance. The bracketed numbers in the legend is the baseline, i.e. the average prediction, of each class.

the respective class, for CL2 and CL3 is by 87% and 80%, respectively, compared to only 59% for CL1. Similar the *precision* (Appendix A.3, Eq. A.9),the fraction of a correctly predicted class out of all model predictions of the respective class, for CL1 is only at 63%, while it goes up to 79% and 85% for CL3 and CL2, respectively. Both metrics indicate that windstorm events of CL2 and CL3 are very well partitioned in the model tuning and can be correctly predicted with high confidence. CL1 events, on the other hand, seem hard to discriminate from those of CL2 and CL3, as precision and recall are considerably lower for CL1 than for the other two.

4.2.3 Quantifying the contribution of large-scale atmospheric drivers

Using Shapley values, a metric for quantifying the contribution of a feature to the model prediction of an observation, we can derive the feature importance from our tuned RF classifier (Sect. 4.2.2) by calculating the mean absolute Shapley value for each feature and class (see Fig. 4.6). Note the baseline values, i.e. the averaged probability, of each class in the legend. Those are important, because in case of a classification scenario Shapley values measure the contribution of features in terms of added value on top of the averaged probability of the class. For a better understanding of this, assume the prediction of the probability $\hat{P}$ of an event y being part of CL1 given an observed feature space x. The Theory behind Shapley values assumes that

$\hat{P}(y = \text{CL1}|X = x)$ can be defined as the averaged probability $\overline{P(y = \text{CL1})}$ of predicting CL1 plus the individual contributions of the features $\theta_1, \ldots, \theta_p$, i.e.

$$\begin{aligned}\hat{P}(y = \text{CL1}|X = x) = \overline{P(y = \text{CL1})} + S_{\theta_1}(y = \text{CL1}|X = x) \\ + \ldots + S_{\theta_p}(y = \text{CL1}|X = x),\end{aligned} \tag{4.7}$$

where $-1 \leq S \leq 1$ is the contribution, i.e. added probability. Shapley values help us to assign a value to each $S(\theta)$. For a more comprehensive description of the Shapley values see Appendix A.4 or Molnar (2019).

According to Fig. 4.6, the most important large-scale atmospheric drivers for our model are JET, LAT_P, EPT and PMIN. Since both the jet stream and atmospheric instability play major roles in the development of cyclones, it is not surprising that JET and EPT are important to our model. It also makes sense that PMIN, as a proxy for the cyclone intensity, is a strong indicator for our windstorm classes, as more intense cyclones are more likely to produce severe windstorms, and our classes differ in severity (see SSI in Tab. 3.2). It is interesting that LAT_P has so much weight in our model, but can be explained through the fact that our windstorm classes have regions where they are primarily active (see Fig. 3.4). CL1 and CL3 are primarily active in Central and Northern Europe, while events of CL2 usually occur in the Mediterranean. Therefore, it is likely that LAT_P serves as threshold for the model whether an event is more likely to occur in the south (e.g. CL2) or further north (e.g. CL1 or CL3).

Large-scale atmospheric drivers that play a surprisingly minor role are the teleconnections and LAT_O. In the literature, teleconnection patterns are often associated with windstorm occurrence (e.g. Pinto et al., 2009; Donat et al., 2010). However, our model indicates that they have only little influence on what type of windstorm occurs. For LAT_O, on the other hand, the reason seems to be that the origin of the windstorms is too widely scattered across the meridians, independent of the class (compare to Fig. 4.2a,c,e). This stands in contrast to LON_O, where at least the differentiation between east and west seems to matter.

From Fig. 4.6 we can also deduce how important the features are for each classes by looking at their individual contribution to the global feature importance. For example, PMIN plays a larger role in the prediction of CL1 and CL2 events (mean abs. Shapley value of approximately 0.09 and 0.13, respectively) compared to CL3 events (approximately 0.04). Overall, the prediction of CL1 events seems to be guided primarily by PMIN and JET, which are the two features with the highest individual importance for CL1. Besides these two, most other features have a similar weight in the prediction of CL1 events (outside the previously mentioned teleconnection patterns and LAT_O). Analogously, the most influential features for CL2 are PMIN, LAT_P and EPT, while for CL3 its LAT_P, EPT and JET.

The Shapley values can also be used to show the feature dependence, i.e. how the average model prediction changes with the magnitude of a feature. Figure 4.7 shows the feature dependence for the four features with highest

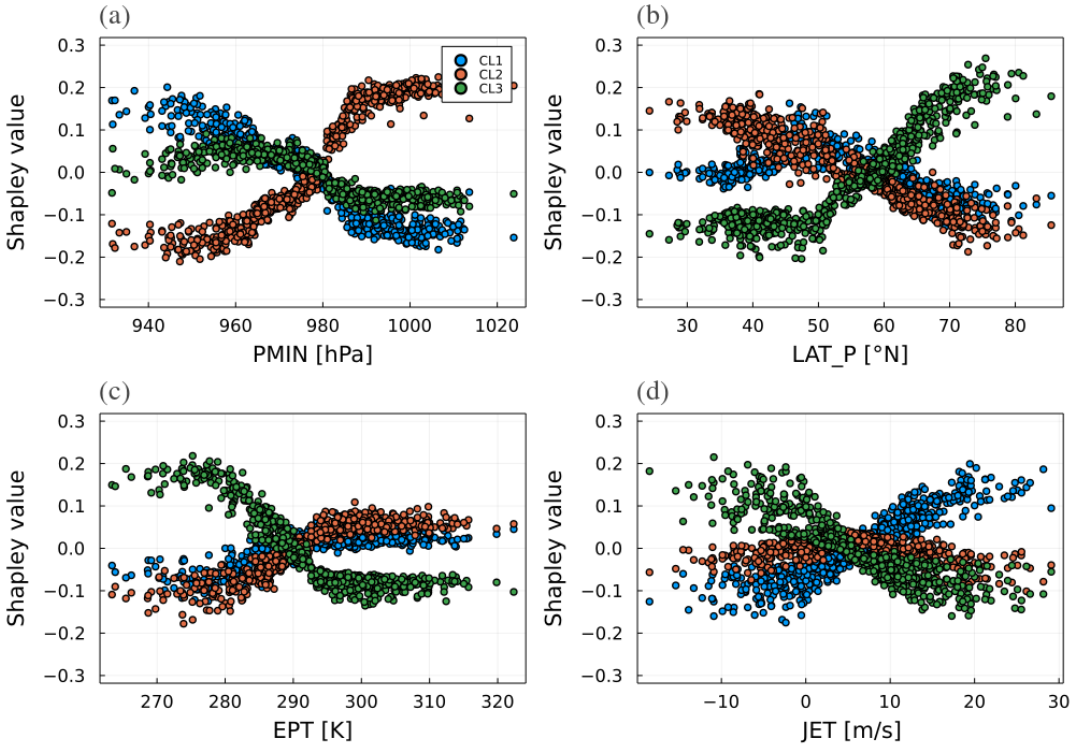

FIGURE 4.7: Feature dependence for (a) JET, (b) LAT_P, (c) PMIN and (d) EPT, the four features with the highest feature importance. The dots are the respective Shapley value of a track and class in our test sample. Again, the Shapley values are estimated with respect to the average model prediction of the class (see Fig. 4.6).

feature importance (Fig. 4.6), namley JET, LAT_P, EPT and PMIN. Each of the dots is a Shapley value, i.e. feature contribution, of the respective feature. Overall, each class in every panel has 661 dots; one for each of the 661 tracks in the test sample. Together, they represent the dependency curve of the respective feature and class.

For PMIN (Fig. 4.7a) the dependence curves are similar for CL1 and CL3, where the Shaply values increases with decreasing PMIN. Hence, both types of events are more likely to have a parent cyclone with a minimum core pressure of below 980 hPa. As PMIN decreases (increases), the dependence curve of CL1 splits from CL3, and the occurrence of a CL1 event becomes increasingly (decreasingly) likely. The complete opposite can be seen for the dependence curve of CL2. Here, the probability of a CL2 events increases with an increasing PMIN, while pressure values below 980 hPa are unlikely for this type of event. Cyclones with a low core pressure are often related to severe windstorms and therefore, considering CL1 and CL3 events are events with higher wind speeds in general (see Tab. 3.2), it is not surprising that the occurrence of these two types of events are more likely with decreasing PMIN. This is especially true for CL1, which is designed to represent historical severe events.

Similar, but mirrored, dependency curves can be seen for LAT_p (Fig. 4.7b) and EPT (Fig.4.7c). For LAT_p, the dependence curves of CL1 and CL2 show a similar behavior, where with increasing latitude both types of event get

more and more unlikely. A key difference between CL1 and CL2 is the location of their maximum, which is between 45-50°N for CL1 and 30-40°N for CL2. Like CL2 in Fig. 4.6a, the dependency curve of CL3 shows the opposite behavior to CL1 and CL2, where the occurrence of a CL3 events is more likely for higher values of LAT_P. However, this tendency only starts at 50°N. Before this threshold, the Shapley values stays between -0.2 and -0.1. As already mentioned, for EPT the orientation of the curves are almost mirrored to LAT_P. Events of CL1 and CL2 are more likely for higher EPT values, while those of CL3 are more likely for lower values. Both features, LAT_P and EPT, are well suited to distinguish CL3 events from those of CL1 and CL3 and the reason for this is that apparently both features have a close relationship to the region in which CL3 events are primarily active. CL3 events usually occur in the polar regions above 60°N (see Fig. 3.4). The Shapley values of LAT_P for CL3 are starting to be positive around 60°N. Around the same latitude, the polar air mass begins, a large volume of air defined by its low temperatures compared to the midlatitudes. Therefore, since the air temperature is proportional to EPT, it is not surprising that lower EPT values are clear indicator for CL3 events.

JET (Fig. 4.7d) shows almost linear dependence curves for all three classes, where each of them has different slope. Small values of JET, i.e. values between -15 and 5 m/s, correspond to positive Shapley values for CL3 and negative values for CL1. After the 5 m/s threshold this relationship swaps to positive Shapley values for CL1 and negative values for CL3. The Shapley values for CL2 are between -0.1 and 0, with a maximum at approximately 5 m/s, for the full range of JET. Overall, Figure 4.7d suggest that a jet stream with higher wind speeds compared to the climatological mean favors the development or occurrence of windstorm events typical of CL1. Note, that we calculate JET using wind speed anomalies at 250 hPa (Tab.4.1). Therefore, Fig.4.7a is not a contradiction to the results from Tab. 3.2, where larger wind speeds are a prevalent characteristic of events from CL3 rather then CL1 or CL2. Here, small values of JET only means that the events where either far away from the jet stream, i.e. outside the 1000 km radius around the windstorm center that was used for averaging, or that the jet stream was weaker compared to the climatological average.

The remaining 10 features start to become less and less distinguishable with decreasing importance (see Appendix B.2). As we already suspected from Fig. 4.6, features for the teleconnection patterns such as NAO, POL or EA and LAT_O seem to have close to no visible impact on the model prediction (Appendix B.2e-j). The features LON_P and LON_O (Appendix B.2a,b) are weak but clear indicators for CL2 and can be summarized as the further east PMIN and the origin are, the more likely the event belongs to CL2. Also interesting to note, weaker values of EGR are indicative for CL2 events (Appendix B.2c). The dependence curve of PJET shows that in the case of the windstorm occurring far south of the jet stream (PJET < -10), it is likely to be an event of CL3.

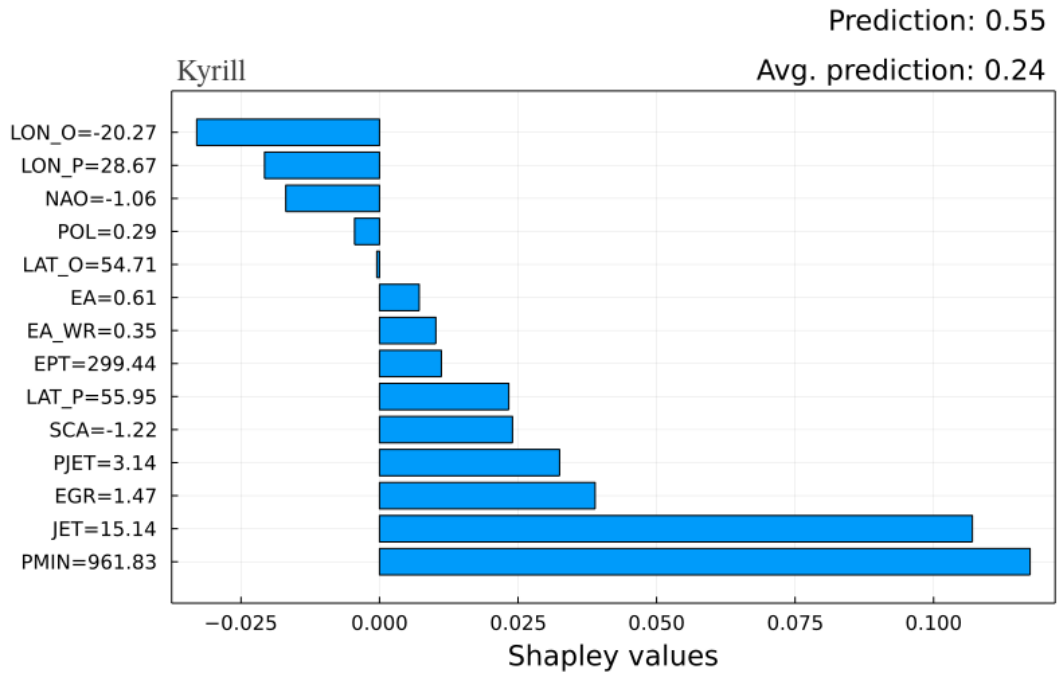

FIGURE 4.8: Contribution of each large-scale feature on the model prediction of the class for Kyrill. The contribution is given with respect to the average (avg.) probability of 0.24. In the upper-right corner we also show the actual predicted probability of the event belonging to CL1.

4.2.4 Case studies: Daria, Lothar and Kyrill

Finally, to extend on our findings from Sect. 4.2.3 and to gain a more comprehensive understanding of the interaction between the individual large-scale atmospheric drivers and extreme windstorm events, we conduct a case study of three historical severe windstorms, namely Daria, Lothar and Kyrill. All three events are from our reference and thus part of CL1. Together, they are the three most severe windstorm events of the past 40 years, measured on insured loss (Roberts et al., 2014). In the following, we focus on quantifying and understanding of the relationship between our large-scale features (Tab. 4.1) and these events. For a detailed description of the development and impact of the three individual events, we refer to McCallum (1990) for Daria, Ulbrich et al. (2001) for Lothar, and Fink et al. (2009) for Kyrill.

The Shapley values derived from the prediction of the class for Kyrill (Fig. 4.8), shows that Kyrill had a very deep parent cyclone with a minimum core pressure of approximately 960 hPa and was influenced by a comparably strong jet stream. Both were the dominant factors for the model prediction, with the next closest feature being EGR with less than half the Shapley value of JET. The value for PJET show us that with a difference of only 3.14°N, the center of Kyrill was north of and very close to the jet stream, which might be reason for the magnitude of JET. Interesting is the impact of NAO, POL, $\mathrm{LON_O}$ and $\mathrm{LON_P}$ on the prediction. The Shapley values of all four features are negative, which means that their magnitude is characteristic for one of

the other classes but not CL1. This also the reason why the predicted probability of 0.55 is low compared to these of Daria and Lothar (Appendix B.3 and B.4).

Daria and Lothar (Appendix B, Fig. B.3 and B.4) show larger Shapley values and predicted probabilities, with 0.89 and 0.95 respectively, than Kyrill. Both windstorms had a intense parent cyclone with a minimum core pressure below 950 hPa and were influenced by a closely located and strong jet stream. Again, the features PMIN and JET are stronger than for Kyrill. They both originated in the West Pacific (Daria: $\mathrm{LON_O} = -83.39$; Lothar: $\mathrm{LON_O} = -74.55$) and reached their highest intensity close to the European continent (see $\mathrm{LON_P}$ and $\mathrm{LAT_P}$ in Appendix B, Fig. B.3 and B.4). Hence, both events are the poster example of a windstorm event that originated near the U.S. East coast and intensified on their way across the Atlantic. This stands in contrast to Kyrill, which first occurred in the East Atlantic (Fig. 4.8, $\mathrm{LON_O}$).

Overall, Fig 4.8 and Fig. B.3-B.4 show that despite the difference in origin and intensification, our model is still able to correctly classify Kyrill. However, this difference might be an indication for why our model struggles to correctly predict CL1 events in general (see *precision* and *recall* in Tab. 4.2). Clearly, although the extreme windstorms in our reference have comparable characteristics they do not all have a similar cyclogenesis and intensification phase, and those difference might have been carried over in the construction of our classes. This means that some of the events in CL1 are comparable in their development to Kyrill, others are more comparable to Daria or Lothar, while some might even be completely different from the three aforementioned events. If those events now lack strong indicators such as JET or PMIN, features that were decisive for the prediction of Kyrills' class, it is possible that they are incorrectly classified. Therefore, it is not likely that additional or different features in our classification model would improve the model performance with regard to CL1 events. Rather, a further partitioning of CL1 would be required to account for events with similar characteristics, but different origin, development and intensification. However, this is outside the scope of this dissertation as our focus lies on windstorm characteristics.

4.3 Trend analysis of European windstorm activity

Finally, we close this chapter on a study of European windstorm activity. In Sect. 1.2 we summarized that reanalysis data indicates an overall increase in cyclone and windstorm activity over Europe, but that there are substantial difference in the magnitude, or sometimes even the sign, of these trends in the literature due to variations in the cyclone/windstorm data caused by the tracking scheme, definitions of extremes and/or the underlying reanalysis data (Ulbrich et al., 2009). Here we hope to add to or to further clarify the current understanding of trends in European windstorms by estimating trends from our windstorm classes, providing information on potential changes in the activity of certain windstorms types.

In Sect. 4.3.2, we describe the methodology used to estimated the trends, while the actual results are presented in Sec. 4.3.3. Trends are estimated for all tracks and for each class of tracks individually, to provide a comprehensive trend analysis of European windstorms from the ERA5 reanalysis.

4.3.1 Windstorm count data

We define windstorm activity as the number of days on which a windstorm event occurred per extended winter period. To account for regional differences, we analyze windstorm activity for each grid box individually. Thus, for each grid box x we calculate the time series

$$Y_{x,t} = (Y_{x,1981/1982}, \ldots, Y_{x,2016/2017}), \tag{4.8}$$

where each $Y_{x,u} \in \mathbb{N}$ is the count of days with an windstorm event for the respective extended winter period. For example, $Y_{10,1981/1982}$ would be all days with an windstorm event at the 10th grid box for the months October to March in 1981 to 1982. Since we have 37 years of windstorms, we have 36 complete extended winter period and thus $Y_{x,t}$ consists of 36 values at each grid box.

Note that we do not distinguish between events in $Y_{x,t}$, i.e. if a consecutive number of windstorm days at a grid box is caused by a single windstorm, they are still counted individually. We believe that this is the most honest representation of local windstorm activity, since from perspective of decision makers or local residents only the actual number of days with a hazardous events matters, not the event itself. However, since it is relevant information whether the increasing/decreasing number of windstorm days is caused by longer or more frequent windstorm events, we define a second time series

$$Z_{x,t} = (Z_{x,1981/1982}, \ldots, Z_{x,2016/2017}), \tag{4.9}$$

where each $Z_{x,u} \in \mathbb{N}$ is the count of windstorm event for the respective extended winter period at a grid box x.

As already mentioned above, we intent to estimate trends not only for the combined sum of windstorms, but also for the individual classes. In cases, where we estimate trends for a class of windstorms, the definition of our time series (Eq. 4.8 and 4.9) does not change, except that they are now only counting events from the investigated cluster.

Lastly, from Fig. 2.2 we could already see that some grid points experienced much more windstorm events than others. Since we reduce the time series down to 36 values, the overall event density at a grid point should not be a problem as long as there were a few windstorm events for the majority of the winter periods. However, since we intent to also estimate trends for the windstorm classes, the event density at some grid points could fall below a critical value where we have to many years with no events to estimate a reliable trend. Therefore, we decided to only include grid points in the trend estimation with values above zero for at least 20 out of 36 winter periods. Again, this applies to all studied trend scenarios. That means in cases where

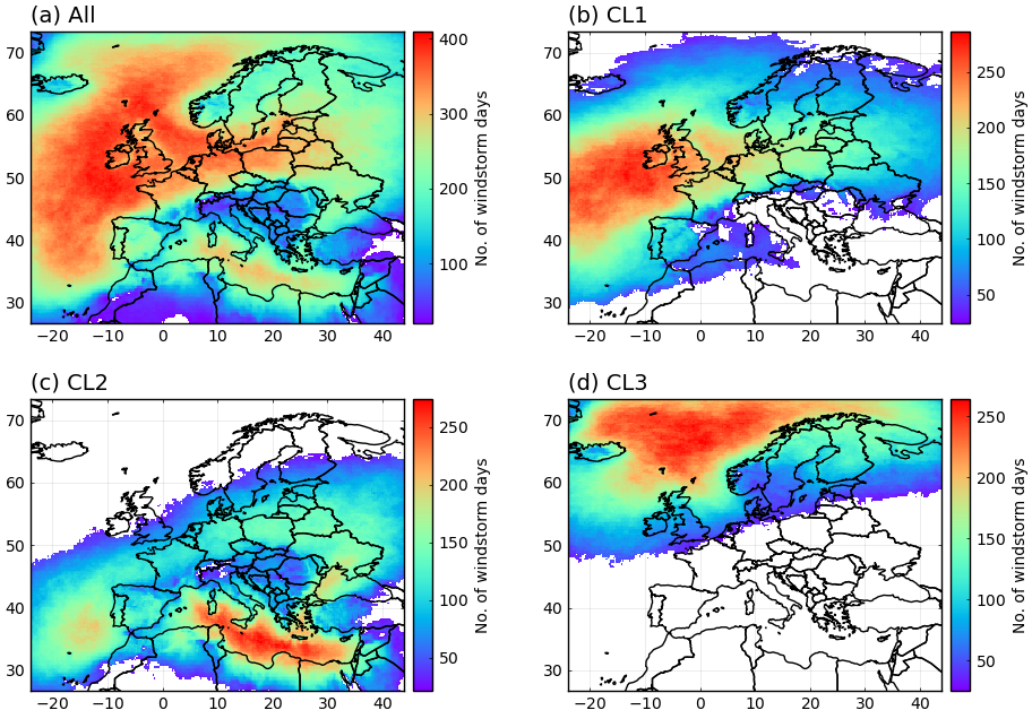

FIGURE 4.9: Grid boxes and their respective number of windstorm days that are left after the selection process for (a) all, (b) CL1, (c) CL2 and (d) CL3. Trends are only estimated at grid boxes that are not white.

we estimate trends for CL1, the respective grid boxes also need windstorm events of CL1 for at least 20 winter periods . Figure 4.9 shows the map of all grid boxes that are left for the individual trend scenarios after the selection process.

4.3.2 Methodology for trend estimation

Various approaches can be found in the literature to capture trends in windstorm activity. For example, Neu et al. (2013), Nissen et al. (2010) and Befort et al. (2016) used least square regression on winter cyclone counts. Donat et al. (2011c) also used a least squares regression, but for two different measures of storm intensity, namely gale days and wind speed percentiles, and tested for significance with the Mann-Kendall test (Kendall, 1948). Paciorek et al. (2002), on the other hand, used Poisson regression (McCullagh and Nelder, 1999) on regional counts of winter cyclones, with an additional focus on intense cyclone (minimum core pressure below 970 hPa).

For our trend estimation, we follow the approach of Paciorek et al. (2002), with slight modifications regarding the Poisson regression model and how we define our windstorm counts (see Sect. 4.3.1). We decided for a Poisson regression in favor of the more commonly used least square regression, as we believe it is the more suitable model to describe windstorm counts. Our main argument against least square regression is that windstorm counts

are discrete measures and therefore cannot be described using a continuous probability distribution such as the Normal distribution. However, normally (or continuously) distributed residuals are one of the key assumptions of least square regression for hypothesis testing, i.e. for testing for significance. The Poisson distribution, on the other hand, has been frequently used to describe count data and access trends in various studies across different research fields (e.g. Thompson and Sorte, 2008; Achcar et al., 2011; Le et al., 2021).

In our Poisson regression model, the windstorm counts $Y \in \mathbb{R}^n$ are assumed to be a Poisson random variable with mean

$$\mathrm{E}\left[Y(t)\right] = \exp(\alpha + \beta t), \tag{4.10}$$

where $\alpha \in \mathbb{R}$ is the intercept, $\beta \in \mathbb{R}^n$ the regression coefficient and $t \in \mathbb{R}^n$ the independent variable, which is here the trend and therefore denoted as t. An important consequence of assuming a Poisson process for our observations is that the variance of Y equals the mean of Y. However, this a fact that non of our time series can fulfill, since the variance of Y is always larger than its mean. The case of the variance of Y being larger than its mean is called overdispersion. Overdispersion does not necessarily influences the estimate of β, but generally leads to underestimated standard errors of β, which are essential for calculating the p-value. Underestimated standard errors lead to smaller p-values and thus more time series might show a significant trend than is actually the case. To prevent this, we include a dispersion term following McCullagh and Nelder (1999) which adjusts the standard errors for overdispersion before the p-values are estimated. For a comprehensive discussion of the overdispersion problem and how a dispersion term can be calculated and included in the model, we refer to Hilbe (2014).

The for overdispersion adjusted Poisson regression model is fitted for the time series discussed in Sect. 4.3.1 for each grid box. We test the estimated trends (i.e. β) for significance using the z-test (see Berger and Casella, 2001). Trends below a significance level of $\alpha = 0.1$ are considered to be significant. All the necessary tools for the modeling are included in the *GLM* package (`https://juliastats.org/GLM.jl/stable/`) for the programming language Julia (Bezanson et al., 2017).

4.3.3 Estimated trends within the windstorm cluster

Figure 4.10 shows the regional trends in the number of windstorm days (hereafter denoted as β_D^u, where u is the respective time series) for our different time series (i.e. all events or only those from the clusters). When all events are included in the trend analysis (Fig. 4.10a), β_D^{All} shows contrasting magnitudes between the southern and central to northern parts of Europe, where all trends above (below) 40°N are negative (positive). The strongest decrease in β_D^{All} can be found in the northern parts of France and Southwest Germany, while the strongest increase lies in a small cluster south of Italy. Overall, the clusters of negative trends are larger and further spread across Europe than their counterparts in the south. For CL1 (Fig. 4.10b), β_D^{CL1} is

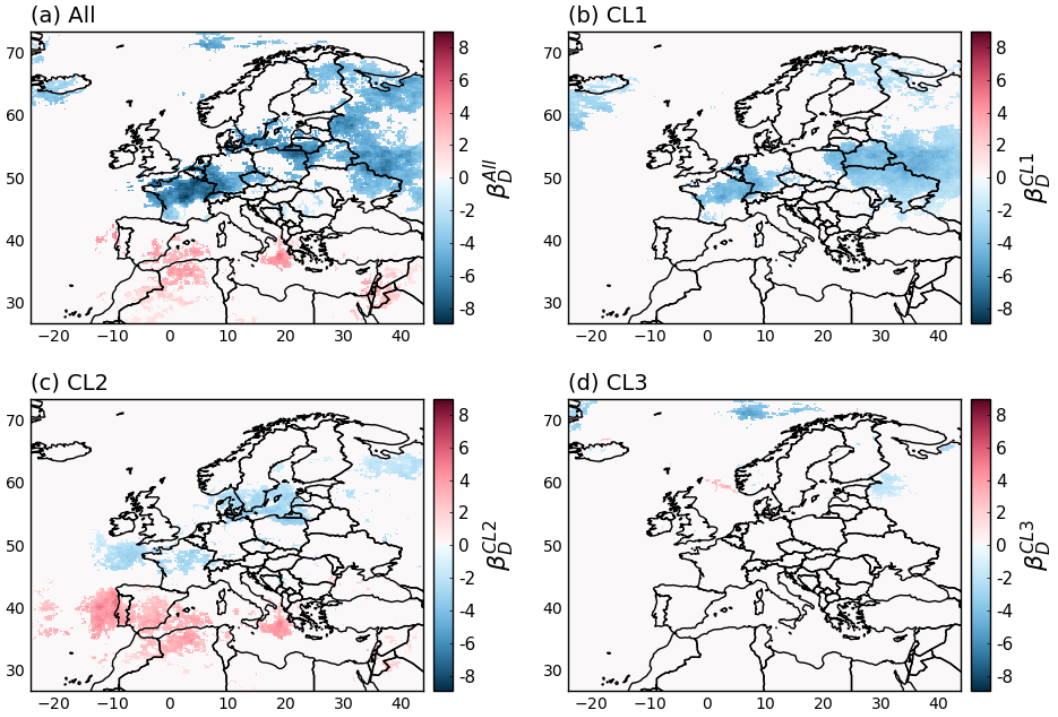

FIGURE 4.10: Regional trends in the number of windstorm days in units of (36 winter periods)$^{-1}$ for (a) all windstrom events and only events of (b) CL1, (c) CL2 and (d) CL3. Show are only grid boxes with a significant trend (significance level of $\alpha = 0.1$).

exclusively negative and can be only found in Central to Northern Europe. There a four clusters of β_D^{CL1}: over Northern France and Southwest Germany, Belarus/Ukraine, Iceland and the northern parts of Finland. Each of these can be found in similar version in Fig. 4.10a, when all events are considered. The trends in CL2 (Fig. 4.10c) are similar divided than for all events. A large cluster of positive β_D^{CL2} can be found over western parts of the Mediterranean and smaller batches south of Italy and the Levant region. Decreasing trends are found in similar location than for all and CL1 (Fig. 4.10a,b) with an additional small cluster west of the coast of France. There are almost no significant trends in CL3 (Fig. 4.10), besides very small batches of β_D^{CL3} north of Scandinavia, Russia and the east coast of Greenland, and an even smaller cluster of positive β_D^{CL3} west of the coast of Norway.

Overall, the trends in Fig. 4.10 indicate an increase in windstorm days in the Mediterranean and decrease for central and the northern parts of Europe. The decrease seems to be mostly due to the decrease in windstorm days from CL1 and partly CL2, while the increase in the south is solely due to an increase in windstorm days in CL2. Events of CL3 seem to have little to no impact on the full picture of trends in European windstorm days besides a small batch of negative trends north of Scandinavia. The decrease in windstorm activity in the central and northern parts of Europe is consistent with the results of Paciorek et al. (2002),Nissen et al. (2010) and Wang et al., 2011, and contradicts the results of Donat et al. (2011c) and Neu et al. (2013).

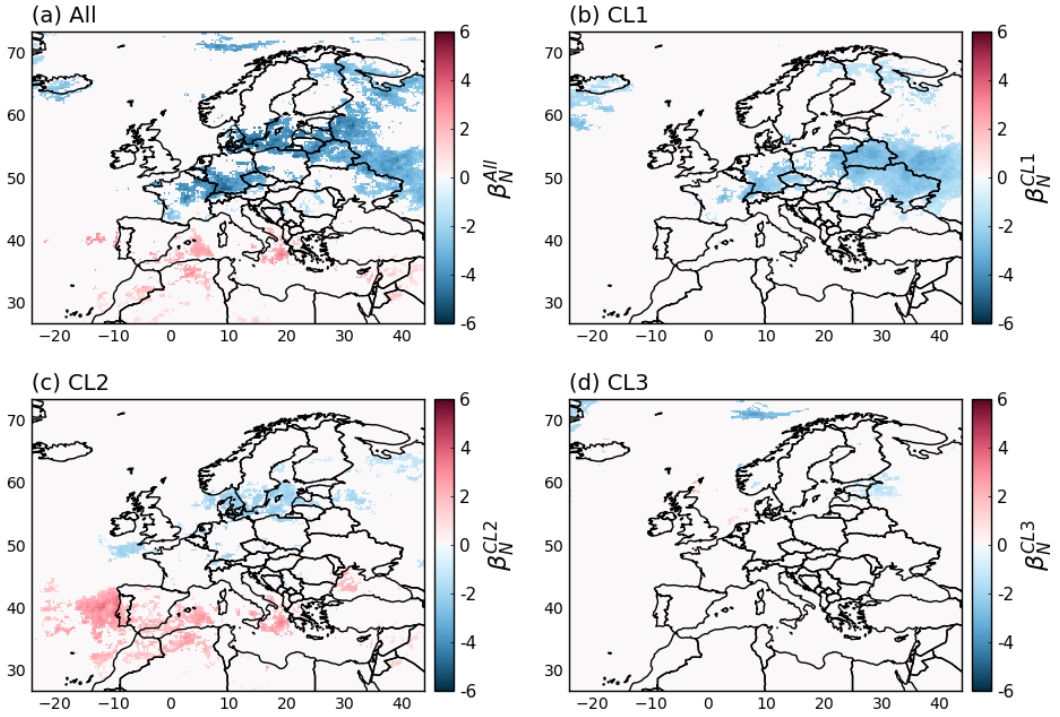

FIGURE 4.11: Regional trends in the number of windstorm events in units of (36 winter periods)$^{-1}$ for (a) all windstrom events and only events of (b) CL1, (c) CL2 and (d) CL3. Show are only grid boxes with a significant trend (significance level of $\alpha = 0.1$).

Similar, the positive trends around Spain and Portugal can only be partly found in Paciorek et al. (2002), but are otherwise contradicting the results in the literature (e.g. Nissen et al., 2010; Neu et al., 2013).

To complement the results from Fig. 4.10 and to identify whether changes in European windstorm activity are due changes in the overall event duration or rather their frequency, Fig. 4.11 shows trends in the number of windstorm events across Europe (hereafter denoted as β_N^u, where u is the respective time series). Both, Fig. 4.10 and 4.11, show the same trend patterns across all studied time series, with only small difference for individual grid points. Therefore, it can be assumed that the trends in European windstorm activity are caused by changes in windstorm frequency. However, one does not exclude the other. The strong link between windstorm days and number of events is favored by our tracking setup, since we only use 6-hourly data and events have to be at least a day long. Therefore, it makes sense that changes in windstorm numbers have a stronger influence on the overall trend behavior, since a single additional event can already add multiple days to our time series.

4.4 Summary and discussion

In this chapter we investigated our derived windstorm classes with respect to their impact, occurrence and temporal evolution (Sect. 4.1), as well as their

connection to large-scale atmospheric processes (Sect. 4.2) and observed climate trends (Sect. 4.3). We provided an overview of their origin, potential impact, and the relationship between the windstorm and its parent cyclone. Many of our results further support the findings of previous studies. Only a few results contradict the results from the literature. Here, we provide a summary of our findings, and further discuss our conclusions.

We found that windstorm events in CL1 primarily affect central and northern parts of Europe with larger than average precipitation amounts and wind gusts (Fig. 4.4a,d), and that they usually occur in December, January or February (Fig. 4.1a). A typical parent cyclone of these events origins along the east coast of the United States and intensifies over the mid and eastern parts of the North Atlantic (Fig. 4.2). Their minimum core pressure often falls below 970 hPa (Fig. 4.7c), especially in the case of severe windstorm events such as Kyrill, Lothar or Daria (see Fig. 4.8 and Fig. B.3-B.4). Windstorm events of CL1 can be associated with an unusually strong jet stream (Fig. 4.7a) and are typically found north of the strongest jet wind speeds (Fig. B.2a). In case of the occurrence of a CL1 type event, local areas are affected between 15-20 hours, where peak wind gusts are to be expected within the first 10 hours (Fig. 4.5a,d). Our trend analysis showed that in the period 1981-2017 the number of CL1 events has decreased for central and eastern Europe (Fig. 4.11). Positive trends could not be found within our study area.

Windstorm events of CL2, on the other hand, usually affect the Mediterranean region, although anomalies in precipitation and wind gust are small compared to CL1 and CL3 (Fig. 4.4b,e). Again, the largest number of windstorm events are found for the winter months December to February. Yet, unlike CL1 and CL3, CL2 events are more common in December instead of January (Fig. 4.1). Their parent cyclones usually form over the mid and western parts of the North Atlantic or the Mediterranean Sea (Fig. 4.2c). Dependent on the origin, the cyclones intensify along the European coast or further east in the Mediterranean (Fig. 4.2d). Their core pressure rarely falls below 990 hPa and increases only slowly during its lifetime (Fig. 4.3h). This behavior makes the core pressure one of the strongest feature for distinguishing CL2 type events from those of CL1 or CL3 (see Fig. 4.7a). On the local scale, windstorm events of CL2 affect the Mediterranean region up to 20 hours and more (Fig. 4.5), which is surprising considering that CL2 events are the shortest on average (see Tab. 3.2). This suggest that windstorms in the Mediterranean region are slow moving events, which leads to longer local durations regardless of an overall shorter lifetime. The trend analysis showed that the overall number of CL2 events has increased in the western parts of the Mediterranean region but decreased for central Europe between 1981-2017 (Fig. 4.11).

Typical CL3 type windstorm events primarily occur in the northern parts of Europe, especially the North Atlantic and coastal regions are affected by positive anomalies in precipitation and wind gusts (Fig. 4.4c,f). Again, the most active season are the winter months December to February. However, compared to CL1 and CL2, the difference between the number of events in October and December is small (Fig. 4.1c). A typical parent cyclone of

theses events origins in western parts of the North Atlantic and further intensify eastwards between the coast of Greenland and the United Kingdom (Fig. 4.2e,f). Similar to CL1, their minimum core pressure lies between 970-980 hPa, but experiences much less variance across its lifetime (Fig. 4.3i). Due to the fact that CL3 events are most active in the northern parts of Europe, the lowest values of the minimum core pressure can be found above 60°N (Fig. 4.7b). In general, these type of events can be associated with lower values of latent heat release in the atmosphere (Fig. 4.7d) and a weaker jet stream (Fig. 4.7a). In cases of the occurrence of a CL3 type event, local areas are affect for 8-15 hours, with peak wind gusts between 5-10 hours (Fig. 4.5c,f). Our trend analysis showed very small cluster of negative trends in windstorm days and number of events were found in Russia and north of the Scandinavian coast (Fig. 4.10 and 4.11). Otherwise, CL3 has the least climate trend signals of all three classes.

Overall, we showed that the most important large-scale atmospheric processes are related to the strength and relative positioning of the jet stream, the characteristics of the parent cyclone (i.e. core pressure and positioning) and the equivalent-potential temperature, which could suggest a strong influence of latent heat release (Fig. 4.6). However, each class was differently affected by changes in their magnitude or location. We also showed that large-scale teleconnection patterns such as the NAO or POL have no impact on the class of windstorm (Fig. 4.6 and B.2), which is an interesting result considering how influential some of these patterns are for the occurrence of European windstorms (e.g. Pinto et al., 2009; Donat et al., 2010; Walz et al., 2018). Our trend analysis showed that windstorm activity decreased over large parts of central and north Europe, but increased in the Mediterranean. These trends are consistent with the findings of some previous studies, but also contradicts the findings of other studies. This shows the disparity in information that can be caused by the use of different data, methods and/or investigated time period.

Chapter 5

Synopsis

The two main objectives of this dissertation were to classify European winter windstorms based on their inherent characteristics such as duration, size, and intensity, and to evaluate the windstorm classes with the goal of identifying key characteristics, understanding the driving large-scale atmospheric mechanisms, quantifying impacts, and estimating trends. From those two objectives, we derived four research questions (see Sect. 1.4), which ultimately defined the structure and methodology of this thesis. In this last chapter, we summarize the key answers to our goals and research questions, critically evaluate the methods we used and developed, and discuss possible improvements (Sect. 5.1). Finally, we provide an outlook on potential future studies that could build on our shown results (Sect. 5.2).

5.1 Conclusion

To answer our first research question and main goal of this dissertation, whether it is possible to construct objective and universal applicable windstorm classes based on their characteristics, we developed and applied the semi-supervised clustering technique QSKM. This novel clustering approach is designed to partition windstorm events and compare the resulting classes with a reference, in our case the XWS open access catalog of extreme European windstorms (Roberts et al., 2014). During the clustering, QSKM selected the most appropriate characteristics to represent the reference from a set of characteristics that we derived from the literature of the windstorm and natural hazard community (see Tab. 2.2). The result were three classes of European winter windstorms, one of which showed exceptional similarities to the given reference (Fig. 3.2), even for characteristics that were not part of the seven characteristics (AREA, DUR, MEANV, DLAT, OLAT, DIST, TSPEED) filtered by QSKM (Fig. 3.3 and Fig. B.1). Throughout this thesis, we referred to this class as CL1. Windstorms of this class are characterized by their exceptional size, lifetime, travel speed and distance (Fig. 3.2). Large SSIs and high wind speeds (see Tab. 3.2) further underlined that QSKM successfully distinguished a subset of extreme and potentially dangerous windstorm events.

Overall, we believe that we achieved our goal of creating transparent and reproducible windstorm classes. Our proposed methodology maintains a high degree of objectivity throughout the classification. While the reference and characteristics are arguably subjective and provided by the user, they

are merely constraints as the final selection of characteristics and the class construction itself is unsupervised. Furthermore, since the classes were constructed from very basic properties, their inherent characteristics, they are suitable for a range of different studies. At least one of the classes resembles observed historical extreme events in their characteristics and can therefore be regarded as an extension to the list or stand-alone catalog of extreme European windstorm events. This alone makes our data interesting for every user that requires a list of extreme European winter windstorms. The fact that the class contains more than 500 events further supports this claim, as this a sample size that current windstorm catalogs cannot provide (compare to Roberts et al. (2014) or Stucki et al. (2014)).

Given our second objective, we undertook a comprehensive analysis to gain further insight into the origin and nature of the windstorms classes. The aim of this analysis was to assess their potential impacts, occurrence, and temporal characteristics, and to understand and quantify the driving large-scale atmospheric processes that determine classification. For the latter, we used a random forest classification model and Shapley values, two powerful statistical tools that are commonly used in machine learning for knowledge discovery. Combining the two tools allowed us to directly measure feature importance and dependence, revealing which of our selected large-scale features were decisive in the classification and how they interacted with the windstorm classes. We concluded the evaluation of the classes with a trend analysis, using Poisson regression to identify past trends.

As we already gave a comprehensive summary and conclusion in Sect. 4.4, we will not further discuss the result for our second objective in this part of the dissertation, but rather point out a few key findings. Overall, the evaluation further underlined the nature of CL1 as a subset of especially extreme events as we found several indicators that already have been discussed in the literature such as deep parent cyclones with a core pressure below 970 hPa, primarily originating in the West Atlantic, a strong and close jet stream to support development and intensification. Furthermore, the evaluation showed the strong influence of the parent cyclone on the windstorm class. Three of the five most important features relate to the positioning and strength of the parent cyclone (Fig. 4.6) and each of them influences the classes in a different way. Another key results was that according to our trend analysis the windstorm activity decreases for central and northern Europe and increases in parts of southern Europe.

The primary source of uncertainty, that has to be accounted for in the results of both of our objectives, are the characteristics and features. In both cases, the clustering and random forest model, we motivated and argued for the used features (see Sect. 2.4 and 4.2.1, respectively), often citing their use and impact in other windstorm studies. Still, the range of characteristics and features that can be included in this type of study is wide and their potential impact on the results hard to estimate. Furthermore, all of the features and characteristics have to be regarded as estimates or proxies of a real event or atmospheric situation. Therefore, their performance can vary dependent on the way they are calculated. In this regard, the evaluation indicated that

two of our features (under-)overperform due to the way we calculated them, namely PJET and EPT. As a feature that is meant to represent the relative positioning of the windstorm and jet stream and thus the impact of the left jet exit on windstorm intensification, PJET had are comparable low impact in the model (Fig. 4.6). Our definition of PJET (Eq. 4.3) might have been to simple to really represent such an important component. EPT, on the other hand, played an important role in the model. However, it is arguable how much of its impact just comes from the relationship to the air temperature, considering the fact that EPT is most influential for events in the northern parts of Europe, but indifferent for events in central and southern Europe.

5.2 Outlook

In the case study of Kyrill, Lothar and Daria (Sect. 4.2.4), we have already mentioned that although the events in CL1 have similar characteristics, they can show notable differences in their origin, development and other characteristics, and that those differences might be the reason behind the low predictability of CL1 events in our random forest model (see Tab. 4.2). For example, with regard to our case study, while Kyrill, similar to Daria and Lothar, originated near the US east coast, it developed a secondary cyclone in the area of the occlusion point several days later in the East Atlantic (Fink et al., 2009). Due to its favorable positioning to the jet stream, this secondary cyclone (usually referred to as Kyrill II) continued to grow, eventually exceeding the core pressure anomaly of the parent cyclone and ultimately causing the devastating destruction that made Kyrill famous in the windstorm community. The cyclone matching introduced in Sect. 2.2.3 assigned the windstorm to Kyrill II, which can be see from the LON_O in Fig. 4.8. Regardless of whether this is considered to be correct or false, it does show the potential differences within CL1 due to differences in development or other characteristics. Exploring those differences might reveal additional sub-classes of windstorms in CL1 and, consequently, improve our classification model and understanding of European windstorms.

Another potential expansion of our work is the application of our methodology to windstorm tracks from climate projections. Climate change is an important topic of increasing urgency and while trends and changes in windstorms and cyclone activity have been extensively studied (see Sect. 1.2 and 4.3), the possible occurrence of new types of windstorms is still a less explored field. Our methodology could help to fill this gap. The study could be conducted in two different ways: for one, the same characteristics that have been identified in this thesis can be used to cluster the projected windstorms, or second, QSKM is applied to the projected windstorms using the same reference and the full list of characteristics as shown in Tab. 2.2. Both cases would answer the question of whether new classes are formed under climate change conditions. In the first case, however, the additional aim is to detect shifts in the distribution of characteristics between classes, while the second case covers the scenario in which previously ignored characteristics gain importance in the future.

A third possible application of our results are the use of windstorm classes in impact studies or storm loss models. In the course of this dissertation, we have shown that CL1 is representative of extreme European winter windstorms in every aspect we tested. Therefore, CL1 is the optimal sample for studying the impact of windstorms on Europe. Especially considering that the class contains more than 500 events, a sufficient sample size for a variety of impact studies. Furthermore, the class can be used to quantify the variance in impact between events with different characteristics, rather than just simple differentiation through scores such as the SSI.

Appendix A

Statistical methods and measures

A.1 Silhouettes

Silhouettes is a method for evaluating and interpreting the quality of clustering. The method is designed to be independent of the clustering technique, and therefore can be applied to any desired clustering output from any given clustering method. This independence is especially helpful for comparing the quality of multiple k-means clustering results for different k's against each other, since the number of clusters has no effect on the silhouette values.

The silhouette value s for a data point i is defined as

$$s(i) = \frac{b(i) - a(i)}{\max\left(a(i), b(i)\right)}, \tag{A.1}$$

where $a(i)$ is the averaged distance d from the ith data point to the other points in the same cluster C_k:

$$a(i) = \frac{1}{|C_k| - 1} \sum_{j \in C_k, i \neq j} d(i, j), \tag{A.2}$$

where $|C_k$ is the number of observation in C_k, and $b(i)$ the smallest averaged distance of the ith data point to all points in any other cluster:

$$b(i) = \min_{k' \neq k} \frac{1}{|C'_k|} \sum_{j \in C'_k} d(i, j). \tag{A.3}$$

Equation A.1 ranges from $-1 \leq s(i) \leq 1$, where a high (low) value indicates that the ith data point is well (poorly) matched to its own cluster while also being poorly (closely) matched to the neighboring clusters.

Each individual $s(i)$ contains information about how well a single object fits into its assigned cluster. Therefore, we can derive the overall clustering quality by averaging all $s(i)$. If their are a lot of objects with a high s, the average will be high itself and the clustering configuration can be considered to be appropriate. On the other hand, if the average is small, one should consider adding/removing clusters from the configuration.

A.2 Random Forests

Random forests (RF) is a supervised machine learning method based on decision trees. As such, the modeling process involves stratifying or segmentation of the feature space into regions representative of the target values. One of the main advantages of RF, and the reason why the model is used, is that due to its simple design, RF can be used for classification and regression purposes without requiring major changes to the theory or evaluation tools. In this section, we will give a brief overview of the RF model and how it functions. For a more comprehensive description of RF or tree-based methods in general, we suggest Hastie et al. (2017) or Gareth et al. (2021).

Since RF is based on decision trees, it is beneficial to look into the concept of decision trees first. As already mentioned above, the idea of decision trees, and therefore of all tree-based methods, consists of splitting the feature space into several regions. The splitting process is recursive and starts from the so-called root, which is the first node. Nodes are split points at which the data is segmented into two partitions based on the most suited feature value. Popular methods for determining suitability is through the reduction of variance in each split, in case of regression, or the gini index (see Gareth et al., 2021), in case of classification. The splitting process continuous for each partition until some stopping rule is applied. The resulting endpoints of the tree are called leafs. Mathematically, leafs express the relationship between the target prediction $\hat{f}(X)$ and the features X:

$$\hat{f}(X) = \sum_{m=1}^{M} c_m I\left\{X \in R_m\right\}, \tag{A.4}$$

where M is the number of regions (i.e. partition) R and I the identity function, which is 1 if X is in the region R_m and 0 otherwise. The constant c_m is the target value for R_m. In case of a RF regression model, an estimate for c_m is given by the mean value of all observations of the target value in R_m. For a classification model, it is the majority vote of all classes in R_m.

The advantages of decision trees are that they are simple, easily interpretable and able to handle qualitative (i.e. categorical) features without requiring any type of transformation. They can be visualized in a convenient flow structure that maps the full decision process, split criteria and feature importance (for example, Fig. A.1).

However, a common problem with decision trees is their strong tendency to overfit the data, resulting in models that have low bias but also high variance. This tendency makes them non-robust to the point that small changes in the data can cause a significantly different tree structure. In other words, in a real machine learning application, predictions from decision trees suffer from high-variance as the predicted value is highly dependent on the subset of data the model was trained on. Therefore, decision trees are best suited for describing finite data such as the data used in Fig. A.1, but more elaborate methods are required for accurate predictions.

A popular method for dealing with the high variance of decision trees is

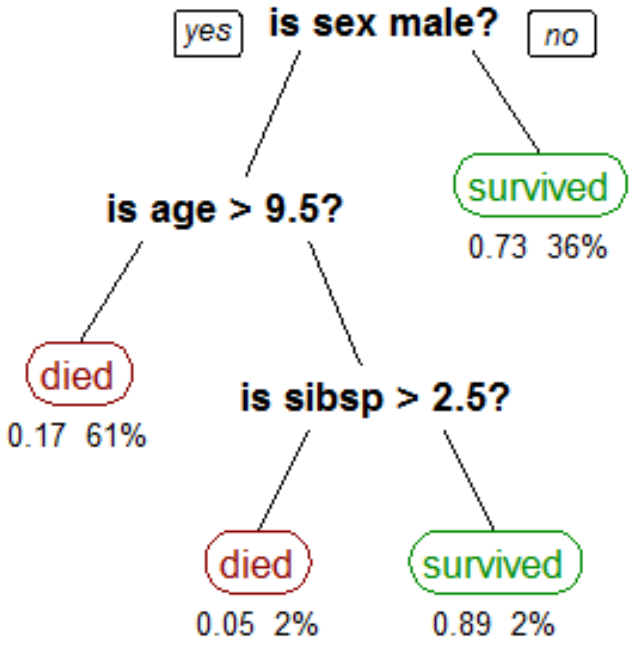

FIGURE A.1: Example decision tree showing survival chances of passengers on the Titanic based on 3 feature: *sex* (categorical - male or female), *age* (continuous) and number of siblings on board (*sibsp*; continuous). People with a good survival chance where either female (probability of 0.73, 36% of the passengers) or young boys from smaller families (probability of 0.89, 2% of the passengers). Source: Stephen Milborrow, https://commons.wikimedia.org/w/index.php?curid=14143467

a process called bagging. In bagging, bootstrap is used to create B different samples of the data, to each of which a decision tree is fitted. If we are now interested in a prediction $\hat{f}(x)$ for a instance x, we can calculate $\hat{f}_1(x), \ldots, \hat{f}_B(x)$ using the previously fitted trees, and average them in order to obtain a single low-variance prediction:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}_b(x) \tag{A.5}$$

To motivated this process mathematically, assume N independent and identically distributed (i.i.d) random variables $X_1, \ldots, X_N$, each with variance σ^2. The Bienaymé formula (Bienaymé, 1855) states that the variance of the mean ($\overline{X}$) can be described as

$$\text{Var}\left(\overline{X}\right) = \text{Var}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{1}{N^2}\sum_{i=1}^{N} \text{Var}\left(X_i\right) = \frac{1}{N^2} N\sigma^2 = \frac{\sigma^2}{N}. \tag{A.6}$$

Hence, the variances of the average decreases as N increases. In the context of bagging this means, that given a large enough B the variance of the prediction should decrease.

Random Forests is an improved version of Bagging, i.e., the method uses bootstrap sampling and averaging (Eq. A.5) for its prediction, but in addition also fixes a methodological problem arising from bootstrapping the data. Remember, for Eq. A.6 to hold true we assume i.i.d random variables. Since each tree is build from a poll of the same data, the trees are identically distributed. However, since they are build with respect to a number of features, they may not necessarily be independent. For example, consider a very strong feature within the data. In such a case, there is a high chance that each tree starts by splitting the data based on that feature alone, making them all correlated (i.e. not independent). To resolve this, Breiman (2001) proposed to decorrelate the trees by forcing the algorithm to chose a random subset of features as split candidates at each split. In other words, while building the tree, the algorithm is not allowed to consider the full set of features when it choses the best suited candidate for splitting the data. This prevents a very strong feature from dominating and leads to uncorrelated trees from each bootstrap sample. Therefore, RF is strictly better and should be preferred over simply bagging of decision trees in a machine learning application.

A.3 Confusion matrix

The confusion matrix is a popular verification tool for classification models and aims to provide a general overview on the model performance. In general, the confusion matrix is a $N \times N$ matrix, where N is the possible number of observed outcomes. On one axis of the matrix are the labels that the model predicted, and the other axis are the actual observed labels (the so-called ground truth). Table A.1 shows the structure of a confusion matrix using the example of a classification of the state of aggregation of water with three possible outcomes: gaseous, solid and liquid. Independent of N, on the diagonal from top left to bottom right are always the cases where the model predicted the same outcome as was observed in the ground truth. Above and below the diagonal are the cases where the model predicted the wrong outcome. The cases above the diagonal are often referred to as false positives or false alarms, while the cases below the diagonal are often referred to as false negatives or misses.

A reason why the confusion matrix is so popular is because of the various model performance metrics that can be derived from the distribution of cases in the matrix, with the three most commonly know metrics being accuracy, precision, and recall. The accuracy is the fraction of correct predictions, i.e.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}, \tag{A.7}$$

and thus a metric for quantifying the overall model performance. The higher the accuracy, the better the model. For example, the accuracy of the classifier

TABLE A.1: Confusion matrix showing the performance of a classifier trained to recognize the state of aggregation of water. Also shown are examples for the column-wise (row-wise) calculation of the precision (recall).

	Ground truth			
Predicted	Gaseous	Solid	Liquid	Precision
Gaseous	12	9	8	$\frac{12}{12+9+8}$
Solid	8	38	14	$\frac{38}{38+8+14}$
Liquid	17	12	150	$\frac{150}{150+17+12}$
Recall	$\frac{12}{12+8+17}$	$\frac{38}{38+9+12}$	$\frac{150}{150+8+14}$	

used in Tab. A.1 is

$$\text{Accuracy} = \frac{12 + 38 + 150}{268} = 0.746, \tag{A.8}$$

or approximately 75%, when convert into percentages. However, the accuracy alone can be misleading. In the classification problem shown in Tab. A.1 the number of correct predictions of *Liquid* are far greater than the number of correct predictions for *Gaseous* or *Solid*. In such a case, the accuracy would be governed by the skill of the model to correctly predict *Liquid*. Therefore, precision and recall are often used alongside the accuracy to get a more complete picture of the model performance. The precision is defined as the fraction of a correctly predicted class out of all model predictions of the respective class, i.e.

$$\text{Precision} = \frac{\text{Number of correct predictions of the class}}{\text{Total number of predictions of the class}}. \tag{A.9}$$

The recall, on the other hand, is the fraction of a correctly predicted class out of all observations of the respective class, i.e.

$$\text{Recall} = \frac{\text{Number of correct predictions of the class}}{\text{Total number of observation of the class}}. \tag{A.10}$$

Thus, precision and recall complement the accuracy by providing additional information about the model performance with respect to a given class. Examples for the calculation of precision and recall are given in Tab. A.1.

Although precision and recall are similar in their design, they both have fundamentally different interpretations. Precision quantifies the performance of the model in terms of how well it predicted a class on average, while recall indicates how often the model failed to detect a class on average. For example, assume we have a high precision of 0.84 and relatively low recall of 0.50 for a class X. In this example, the model was correct in 84% of the cases it predicted X, but only managed to correctly identify 50% of X. In other

words, the model would be very good for half the observation in the class, but ignored the other half completely.

A.4 Shapley values

In science, the interest in a statistical model often lies not in its ability to precisely estimate a target variable, but rather in the possibility to evaluate and quantify the contribution of certain features to the model prediction. In case of a linear model, calculating the individual contribution is rather easy. Consider a linear model prediction for a given instance x:

$$\hat{f}(x_1, \ldots, x_p) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p, \tag{A.11}$$

where each x_j with $j = 1, \ldots, p$ is a feature value and β_j is the corresponding regression parameter (i.e. weight) gained from fitting the model. The contribution ϕ_j of the j-th feature on the prediction $\hat{f}(x)$ can be calculated as

$$\phi_j(x) = \beta_j x_j - \mathbf{E}(\beta_j \mathbf{X_j}) = \beta_j x_j - \beta_j \mathbf{E}(\mathbf{X_j}). \tag{A.12}$$

Equation A.12 is also known as situational importance (Achen, 1982). It is the difference between what the feature j contributes when its value is x_j and what it is expected to contribute. If ϕ_j is positive (negative), than the feature has a positive (negative) contribution. If it is 0, then the feature has no contribution at all.

However, not every statistical model provides a convenient interpretation tool such as the one shown in Eq. A.12, but rather resemble a black box. This means that the design of the model and the role of certain features in it are often not interpretable. Typical examples for such models are random forest or neural networks.

For models like this, we can use the Shapley value to quantify and interpret the contribution of each individual feature. The Shapley value was first introduced by Shapley (1953) and is a measure from coalitional game theory, where it is used to assign payouts to players depending on their contribution to the total payout. It was later adopted to machine learning as a model-agnostic tool to quantify the impact of a feature based on its contribution to the prediction. In this section, we give a short overview of the Shapley value and introduce the interpretation tools that can be derived from it. For a more comprehensive discussion of the Shaply value, we suggest Molnar (2019).

Similar to Eq. A.12, the goal of the Shapley value is to explain the contribution of a feature as the difference between the prediction $\hat{f}(x)$ and the average prediction. The Shapley value itself gives us the contribution of a feature value to a model prediction $\hat{f}(x)$, weighted and summed over all

possible feature value combinations:

$$\phi_j(x) = \overbrace{\sum_{S \subseteq \{1,\ldots,x_p\} \setminus \{x_j\}}}^{\text{A}} \underbrace{\frac{|S|!(p-|S|-1)!}{p!}}_{\text{B}} \underbrace{(\Delta(S \cup \{x_j\}) - \Delta(S))}_{\text{C}}, \tag{A.13}$$

where S is the feature subset, p the number of features and Δ the prediction for feature values in set S that are marginalized over features that are not included in set S:

$$\Delta = \int \hat{f}(x_1, \ldots, x_p) d\mathbb{P}_{x \notin S} - E_X(\hat{f}(X)). \tag{A.14}$$

For a better understanding of Eq. A.13, let us breakdown the function to its individual components. Part A is the sum over all possible combinations, i.e. coalitions, of the features minus the feature of interest x_j. Part B is the weight, defined by the length of the subset $|S|$ and p. In part C, the marginal contribution of the feature value is calculated as the difference between the model prediction minus the average prediction of the feature set S with and without x_j .

Since the number of coalitions increases exponentially as the number of features increases, computing an exact solution for the Shapley value becomes a computational challenge. Therefore, Štrumbelj and Kononenko (2014) propose an approximation with Monte-Carlo sampling:

$$\phi_j(x) = \frac{1}{M} \sum_{m=1}^{M} \left(\hat{f}(x^m_{+j}) - \hat{f}(x^m_{-j}) \right), \tag{A.15}$$

where M is the number of possible coalitions and m an indicator for the coalition. Here, $\hat{f}(x^m_{+j})$ is the model prediction with x_j and $\hat{f}(x^m_{-j})$ the same prediction without it.

Example: bike rental data

The concept and theory behind the Shapley values can be difficult to understand, so it is beneficial to look at an example of how these values are used and interpreted. For our example, we look at bike rental data from the company Capital-Bikeshare in Washington D.C., an openly available dataset from the UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset). Table A.2 shows a list and short description of the variables in the dataset. The information about the weather and season were added by Fanaee-T and Gama (2014). A minor preprocessing of the data was done by Molnar (2019).

Our goal is to understand how the rented number of bikes (*cnt* in Tab. A.2) varies dependent on weather, day or season. For this purpose, we train a random forest regression model (Appendix A.2) to predict the number of rented bikes per day. Our features are all variables from Tab. A.2 except *cnt*, which is our target variable.

TABLE A.2: Variable names and their description for the bike rental dataset.

Short name	Long name	Description
cnt	Bike count	Number of bikes rented that day
season	Season	Spring, Summer, Fall or Winter
yr	Year	The year, either 2011 or 2012
mnth	Month	Months from January to December
holiday	Holiday	Holiday, yes or no
weekday	Weekday	Day of the week
workingday	Working day	Working day, yes or no
weathersit	Weather situation	Misty, good or rain/snow/storm
temp	Temperature	Mean daily temperature
hum	Humidity	Mean daily relative humidity
windspeed	Wind speed	Mean daily wind speed in 10 m
days_since_2011	Days since 2011	No. of days since 2011

Remember, Shapley values explain the contribution of a feature as the difference between the model prediction of a data instance and the average prediction of all instances (see Eq. A.13 and A.14). Therefore, it is meaningful to first look at the Shapley values for the prediction of a single day. Figure A.2 shows the estimated Shapley values of each feature for the 12th October in 2011. For this day, the model predicts 2603 rented bikes, while the average prediction for all days is 4518 rented bikes. The Shapley values explain how each feature contributes to the deficit of -1915 rented bikes. According to Fig. A.2, the biggest contributers are the humidity, the weather situation, temperature, wind speed, and the number of days since 2011 as proxy for a trend. The high humidity and rainy weather of that day are the main reason behind the reduced number of rented bikes. Only the temperature on this day has a meaningful positive impact on the number of rented bikes.

Although a local interpretation of the feature contribution, as shown in Fig. A.2, can be interesting and also serves as a "sanity check" for the fitted model, a global interpretation is often more informative, especially when the general relationship between the target and a certain feature is of interest. For this purpose, Lundberg and Lee (2017) introduced SHAP (short for SHapley Additive exPlanations), a framework for the global interpretation of the feature contribution based around the aggregation of Shapley values. Two of the more important global interpretation tools from their framework are visualization tools for the feature importance and dependence. The SHAP feature importance is defined as the average absolute Shapley value of all predicted instances:

$$I_j = \frac{1}{N} \sum_{n=1}^{N} |\phi_j|. \tag{A.16}$$

The interpretation is rather simple: the larger the absolute Shapley value, the more important is the feature. Figure A.3 shows the feature importance for the bike dataset. Contrary to what the local Shapley values (Fig. A.2) might suggest, the most important feature is the trend (*days_since_2011*), followed

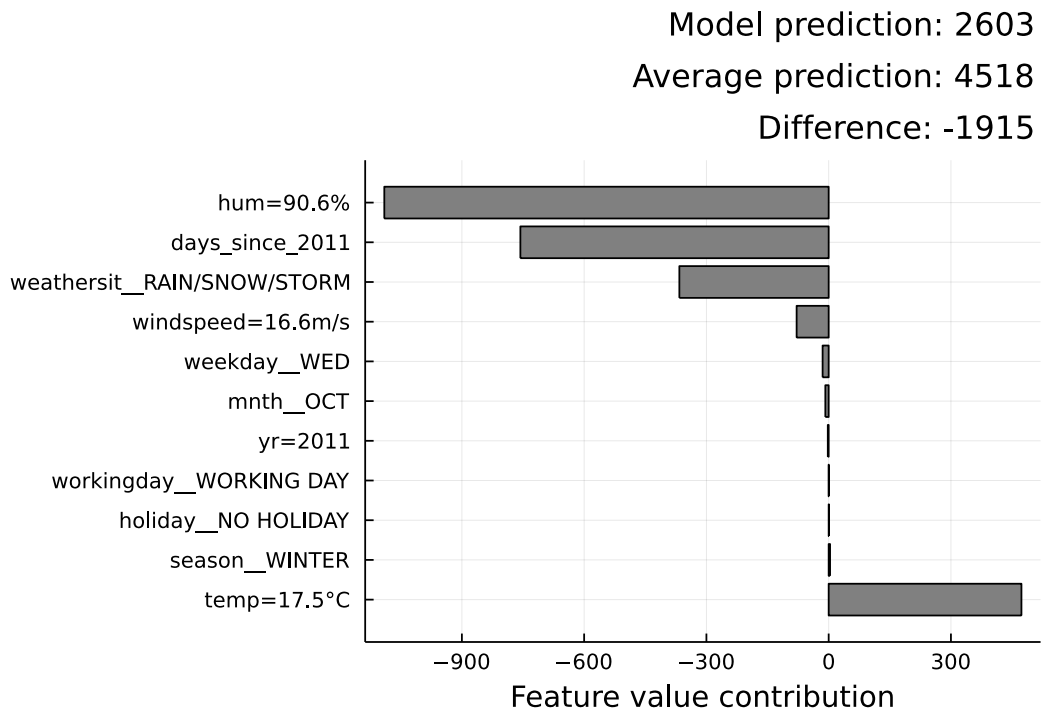

FIGURE A.2: Shapley values (i.e. feature value contribution) for the 12th October, 2011. The model prediction for this day is 2603 rented bikes; average prediction for all days is 4518 rented bikes.

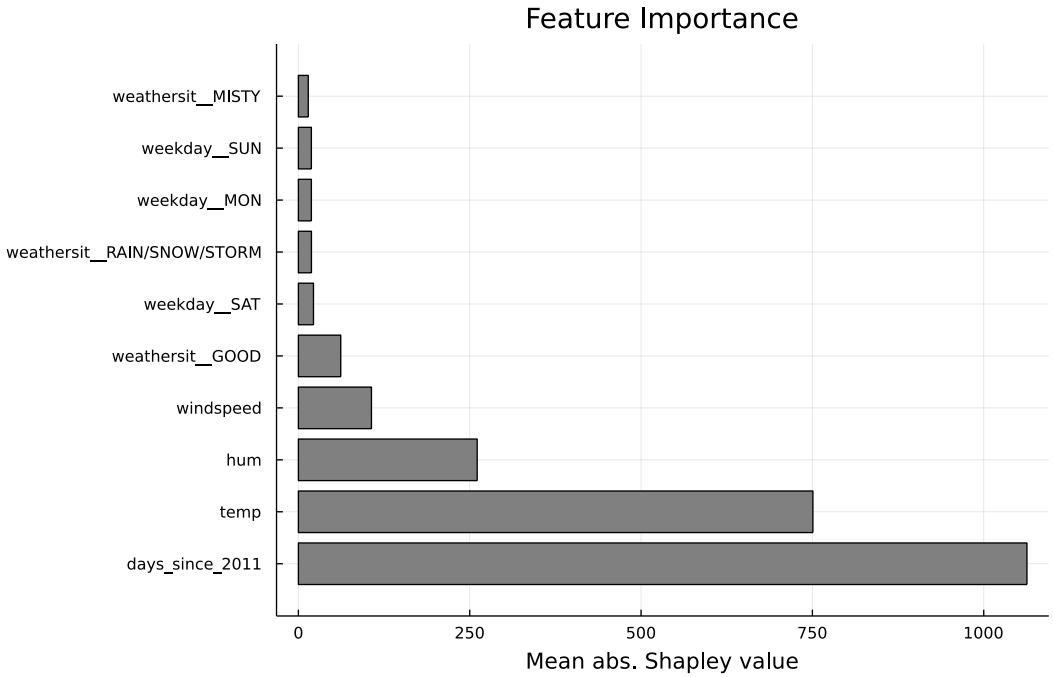

FIGURE A.3: SHAP feature importance for the bike dataset calculated from Eq. A.16 for the features in Tab. A.2

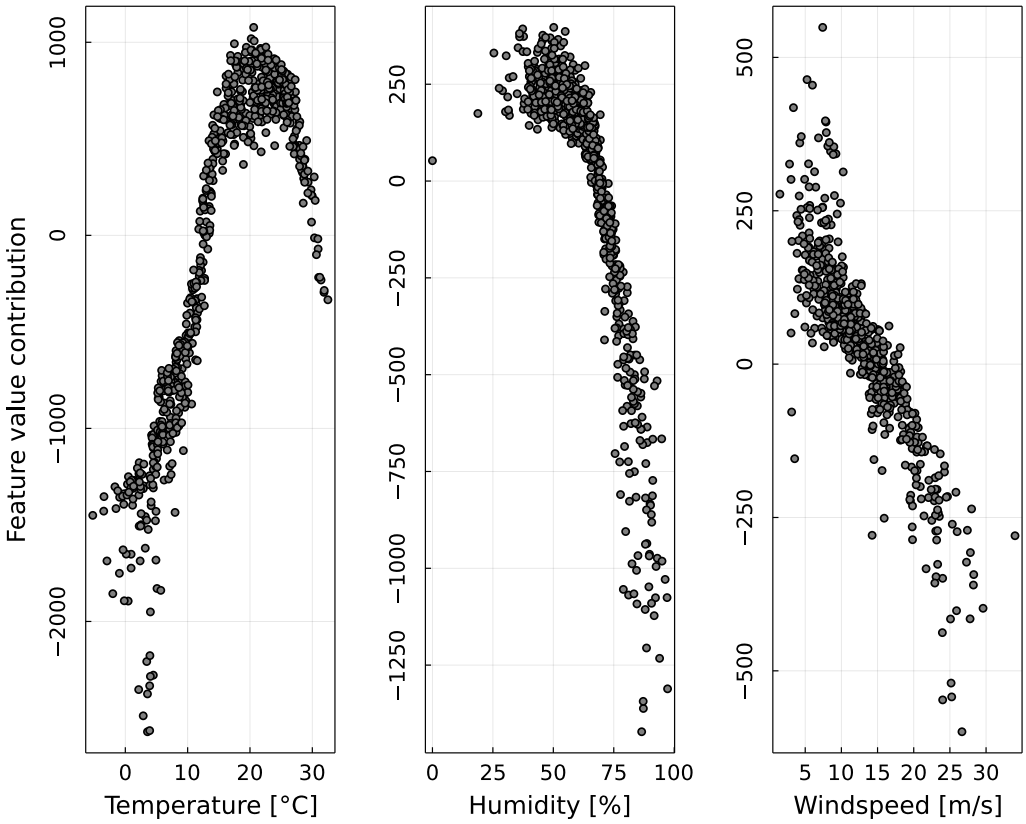

FIGURE A.4: SHAP feature dependence for the bike dataset. The dependence is visualized as the local Shapley value of all instances against their respective feature value.

by the temperature, humidity and wind speed. The high Shapley values for the trend come from a large increase in rented bikes from 2011 to 2012 (not shown). To get an understanding of how the other three features interact with the number of rented bikes, we can look at the SHAP feature dependence in Fig. A.4. As you can see, the SHAP feature dependence is nothing more than the local Shapley values plotted against their respective feature values. Therefore, the plot provides a good overview of how features contribute to the prediction dependent on their magnitude.

Appendix B

Supplement

B.1 Supplemental material to Section 1.3.2

TABLE B.1: Characterization of the different types of windstorms as shown in Dreveton et al. (1998).

Origin	Atlantic Ocean					Mediterranean Sea	
Direction of the flow	West					South	
Location of the low pressure	British Isles		Biscay			Iberian Peninsula	Italy-Corsica
Direction of the winds	North-west to west		South-west to south			South-east	North-east
Affected areas	All the country	Half North	Half North	All the country		South-eastern part	South-eastern part
Centre of low pressures	One centre	One center	One center	One center	A complex area of low pressures	One center	One center
Associated phenomena	Cold thunderstorms, strong shower with hail or snow	Showers and sometimes thunderstorms, mid temperatures	Mild temperatures	Mild temperatures and heavy rain	Mild temperatures, close to average, thunderstorms with hailstones	Heavy rain, sometimes snow	Snow
Type of Storm	N	W	SW	S	TH	E	NE

B.2 Supplemental material to Section 3.4

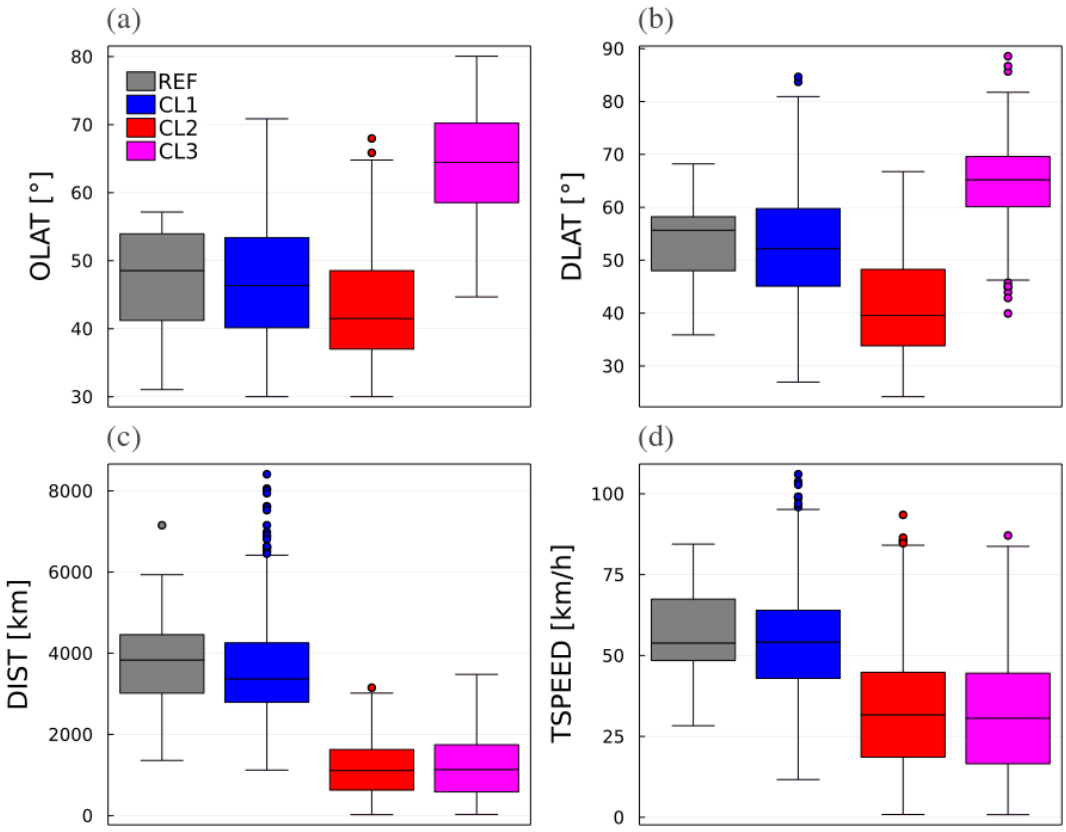

FIGURE B.1: Same as Fig. 3.3, but for OLAT, DLAT, DIST and TSPEED.

B.3 Supplemental material to Section 4.2.3

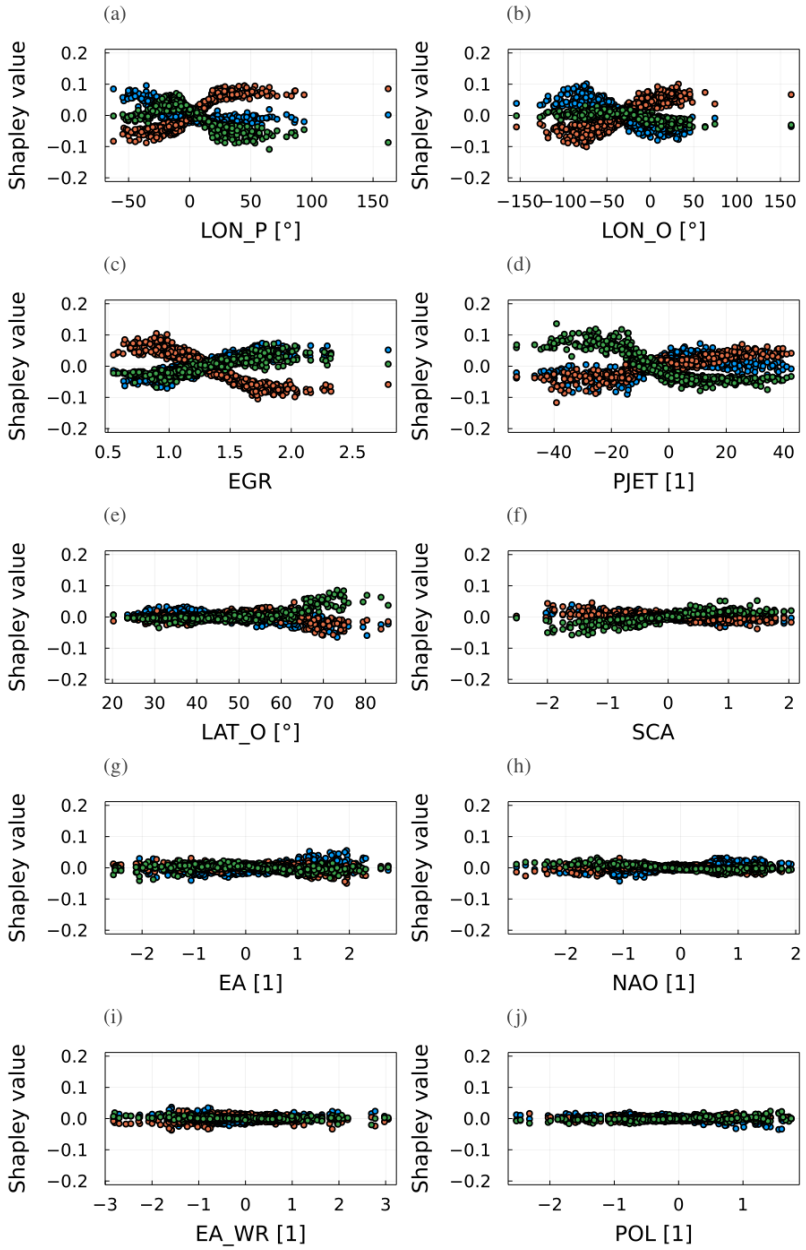

FIGURE B.2: Same as Fig. 4.7, but for (a) PJET, (b) LON_O, (c) LON_P, (d) SCA, (e) EGR, (f) LAT_O, (g) EA, (h) NAO, (i) EA/WR adn (j) POL.

B.4 Supplemental material to Section 4.2.4

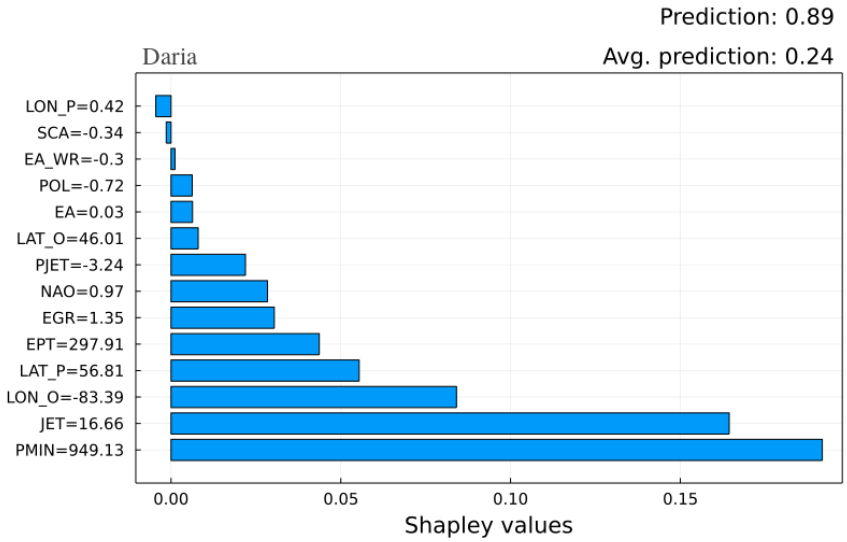

FIGURE B.3: Same as Fig. 4.8, but for the windstorm Daria.

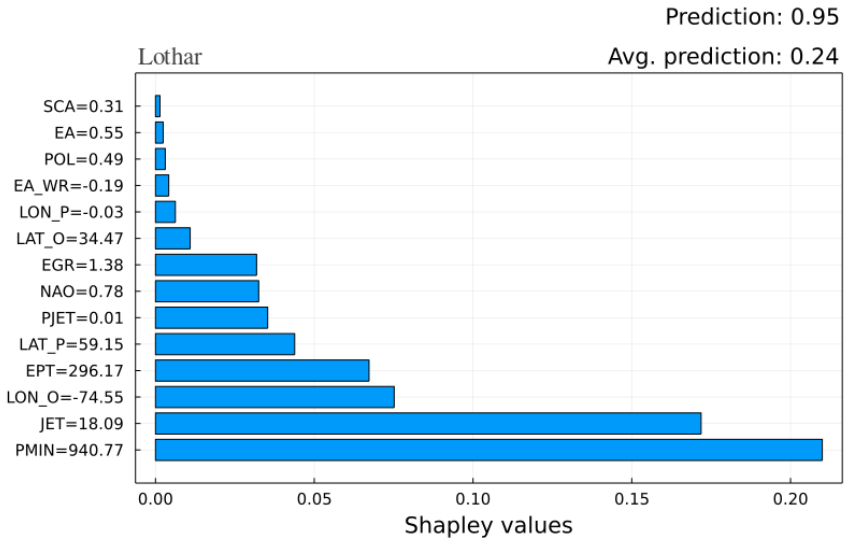

FIGURE B.4: Same as Fig. 4.8, but for the windstorm Lothar.

Bibliography

Achcar, J. A. et al. (2011). "Use of Poisson spatiotemporal regression models for the Brazilian Amazon Forest: malaria count data". In: *Revista da Sociedade Brasileira de Medicina Tropical* 44 (6), pp. 749–754. DOI: 10.1590/S0037-86822011000600019.

Achen, C. (1982). *Interpreting and Using Regression*. Vol. 3. SAGE Publications, Inc., p. 152. DOI: 10.4135/9781412984560.

Allen, J. T., A. B. Pezza, and M. T. Black (2010). "Explosive Cyclogenesis: A Global Climatology Comparing Multiple Reanalyses". In: *Journal of Climate* 23 (24), pp. 6468–6484. DOI: 10.1175/2010JCLI3437.1.

Arthur, D. and S. Vassilvitskii (2007). "K-means++: the advantages of careful seeding". In: *18th Annual ACM-SIAM symposium on Discrete algorithms*.

Baehr, C. et al. (1999). "Dynamical Characterization of the FASTEX cyclogenesis cases". In: *Quarterly Journal of the Royal Meteorological Society* 125 (561), pp. 3469–3494. DOI: 10.1002/qj.49712556117.

Bair, E. and R. Tibshirani (2004). "Semi-Supervised Methods to Predict Patient Survival from Gene Expression Data". In: *PLoS Biology* 2 (4). Ed. by Todd Golub, e108. DOI: 10.1371/journal.pbio.0020108.

Barnston, A. G. and R. E. Livezey (1987). "Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns". In: *Monthly Weather Review* 115 (6), pp. 1083–1126. DOI: 10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2.

Befort, D. J. et al. (2016). "Different long-term trends of extra-tropical cyclones and windstorms in ERA-20C and NOAA-20CR reanalyses". In: *Atmospheric Science Letters* 17 (11), pp. 586–595. DOI: 10.1002/asl.694.

Berger, R. L. and G. Casella (2001). *Statistical Inference*. 2nd ed. Duxbury Press, pp. 1–688.

Berrisford, P. et al. (2011). *The ERA-Interim archive Version 2.0*. ECMWF, pp. 1–23.

Bezanson, J. et al. (2017). "Julia: A Fresh Approach to Numerical Computing". In: *SIAM Review* 59 (1), pp. 65–98. DOI: 10.1137/141000671.

Bienaymé, I. (1855). "Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés". In: *Comptes rendus de l'Académie des sciences*, pp. 309–317.

Blaom, A. et al. (2020). "MLJ: A Julia package for composable machine learning". In: *Journal of Open Source Software* 5 (55), p. 2704. DOI: 10.21105/joss.02704.

Bolton, D. (1980). "The Computation of Equivalent Potential Temperature". In: *Monthly Weather Review* 108 (7), pp. 1046–1053. DOI: 10.1175/1520-0493(1980)108<1046:TCOEPT>2.0.CO;2.

Breiman, L. (2001). "Random Forests". In: *Machine Learning* 45, pp. 5–32. DOI: 10.1023/A:1010933404324.

Catto, J. L. (2016). "Extratropical cyclone classification and its use in climate studies". In: *Reviews of Geophysics* 54 (2), pp. 486–520. DOI: 10.1002/2016RG000519.

Catto, J. L. et al. (2012). "Relating global precipitation to atmospheric fronts". In: *Geophysical Research Letters* 39 (10), pp. 1–6. DOI: 10.1029/2012GL051736.

Compo, G. P. et al. (2011). "The Twentieth Century Reanalysis Project". In: *Quarterly Journal of the Royal Meteorological Society* 137 (654), pp. 1–28. DOI: 10.1002/qj.776.

Dacre, H. F. and S. L. Gray (2009). "The Spatial Distribution and Evolution Characteristics of North Atlantic Cyclones". In: *Monthly Weather Review* 137 (1), pp. 99–115. DOI: 10.1175/2008MWR2491.1.

Dacre, H. F. and J. G. Pinto (2020). "Serial clustering of extratropical cyclones: a review of where, when and why it occurs". In: *npj Climate and Atmospheric Science* 3 (1), p. 48. DOI: 10.1038/s41612-020-00152-9.

Della-Marta, P. M. et al. (2009). "The return period of wind storms over Europe". In: *International Journal of Climatology* 29.3, pp. 437–459. DOI: 10.1002/joc.1794.

Deveson, A. C. L., K. A. Browning, and T. D. Hewson (2002). "A classification of FASTEX cyclones using a height-attributable quasi-geostrophic vertical-motion diagnostic". In: *Quarterly Journal of the Royal Meteorological Society* 128 (579), pp. 93–117. DOI: 10.1256/00359000260498806.

Donat, M. G. et al. (2010). "Examination of wind storms over Central Europe with respect to circulation weather types and NAO phases". In: *International Journal of Climatology* 30 (9), pp. 1289–1300. DOI: 10.1002/joc.1982.

Donat, M. G. et al. (2011a). "Future changes in European winter storm losses and extreme wind speeds inferred from GCM and RCM multi-model simulations". In: *Natural Hazards and Earth System Science* 11 (5), pp. 1351–1370. DOI: 10.5194/nhess-11-1351-2011.

Donat, M. G. et al. (2011b). "High-resolution refinement of a storm loss model and estimation of return periods of loss-intensive storms over Germany". In: *Natural Hazards and Earth System Science* 11 (10), pp. 2821–2833. DOI: 10.5194/nhess-11-2821-2011.

Donat, M. G. et al. (2011c). "Reanalysis suggests long-term upward trends in European storminess since 1871". In: *Geophysical Research Letters* 38 (14), pp. 1–6. DOI: 10.1029/2011GL047995.

Dreveton, C., B. Benech, and S. Jourdain (1998). "Classification of windstorms over France". In: *International Journal of Climatology* 18 (12), pp. 1325–1343. DOI: 10.1002/(SICI)1097-0088(1998100)18:12<1325::AID-JOC274>3.0.CO;2-D.

Eady, E. T. (1949). "Long Waves and Cyclone Waves". In: *Tellus* 1 (3), pp. 33–52. DOI: 10.3402/tellusa.v1i3.8507.

Fanaee-T, H. and J. Gama (2014). "Event labeling combining ensemble detectors and background knowledge". In: *Progress in Artificial Intelligence* 2 (2-3), pp. 113–127. DOI: 10.1007/s13748-013-0040-3.

Fink, A. H. et al. (2009). "The European storm Kyrill in January 2007: synoptic evolution, meteorological impacts and some considerations with respect to climate change". In: *Natural Hazards and Earth System Sciences* 9 (2), pp. 405–423. DOI: 10.5194/nhess-9-405-2009.

Fränti, P. and S. Sieranoja (2019). "How much can k-means be improved by using better initialization and repeats?" In: *Pattern Recognition* 93, pp. 95–112. DOI: 10.1016/j.patcog.2019.04.014.

Gade, K. (2010). "A Non-singular Horizontal Position Representation". In: *Journal of Navigation* 63 (3), pp. 395–417. DOI: 10.1017/S0373463309990415.

Gareth, J. et al. (2021). *Tree-Based Methods*. DOI: 10.1007/978-1-0716-1418-1_8.

GDV (2018). "Sturm - die rasende Gefahr". In: *Naturgefahrenreport 2018*. Gesamtverband der Deutschen Versicherungswirtschaft e. V., pp. 8–25.

Gonzalez, T. F. (1985). "Clustering to minimize the maximum intercluster distance". In: *Theoretical Computer Science* 38, pp. 293–306. DOI: 10.1016/0304-3975(85)90224-5.

Gray, S. L. and H. F. Dacre (2006). "Classifying dynamical forcing mechanisms using a climatology of extratropical cyclones". In: *Quarterly Journal of the Royal Meteorological Society* 132 (617), pp. 1119–1137. DOI: 10.1256/qj.05.69.

Handmer, J. et al. (2012). "Changes in impacts of climate extremes: human systems and ecosystems". In: *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 231–290.

Hastie, T., R. Tibshirani, and J. Friedman (2017). *Random Forests*.

Hersbach, H. et al. (2020). "The ERA5 global reanalysis". In: *Quarterly Journal of the Royal Meteorological Society* 146 (730), pp. 1999–2049. DOI: 10.1002/qj.3803.

Hewson, T.D. and U. Neu (2015). "Cyclones, windstorms and the IMILAST project". In: *Tellus A: Dynamic Meteorology and Oceanography* 67 (1), p. 27128. DOI: 10.3402/tellusa.v67.27128.

Hilbe, J. M. (2014). *Modeling Count Data*. Cambridge University Press. DOI: 10.1017/CBO9781139236065.

Höppner, C., M. Buchecker, and M. Bründl (2010). *Risk Communication and Natural Hazards*. CapHaz-Net WP5 Report, Swiss Federal Research Institute WSL.

IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]* IPCC, 151pp.

James, G. et al. (2013). "Classification". In: *An Introduction to Statistical Learning*. Vol. 103. New York: Springer Texts in Statistics, pp. 127–173. DOI: 10.1007/978-1-4614-7138-7_4.

Kalnay, E. et al. (1996). "The NCEP/NCAR 40-Year Reanalysis Project". In: *Bulletin of the American Meteorological Society* 77 (3), pp. 437–471. DOI: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kendall, M. G. (1948). *Rank correlation methods.* Griffin.

Klawa, M. and U. Ulbrich (2003). "A model for the estimation of storm losses and the identification of severe winter storms in Germany". In: *Natural Hazards and Earth System Science* 3 (6), pp. 725–732. DOI: 10.5194/nhess-3-725-2003.

Kloesgen, W. and J. M. Zytkow (2002). "The Goal of Classification". In: *Handbook of Data Mining and Knowledge Discovery*. Oxford University Press, pp. 254–258.

Kruschke, T. (2015). "Winter wind storms : identifcation, verification of decadal predictions, and regionalization". Freie Universität Berlin.

Le, D. D., R. L. Gonzalez, and J. U. Matola (2021). "Modeling count data for health care utilization: an empirical study of outpatient visits among Vietnamese older people". In: *BMC Medical Informatics and Decision Making* 21 (1), p. 265. DOI: 10.1186/s12911-021-01619-2.

Leckebusch, G. C., D. Renggli, and U. Ulbrich (2008). "Development and application of an objective storm severity measure for the Northeast Atlantic region". In: *Meteorologische Zeitschrift* 17 (5), pp. 575–587. DOI: 10.1127/0941-2948/2008/0323.

Leckebusch, G. C. et al. (2007). "Property loss potentials for European midlatitude storms in a changing climate". In: *Geophysical Research Letters* 34 (5), p. L05703. DOI: 10.1029/2006GL027663.

Lindzen, R. S. and B. Farrell (1980). "A Simple Approximate Result for the Maximum Growth Rate of Baroclinic Instabilities". In: *Journal of the Atmospheric Sciences* 37 (7), pp. 1648–1654. DOI: 10.1175/1520-0469(1980)037<1648:ASARFT>2.0.CO;2.

Lloyd, S. (1982). "Least squares quantization in PCM". In: *IEEE Transactions on Information Theory* 28 (2), pp. 129–137. DOI: 10.1109/TIT.1982.1056489.

Lundberg, S. and S. Lee (2017). "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems*.

Mailier, P. J. et al. (2006). "Serial Clustering of Extratropical Cyclones". In: *Monthly Weather Review* 134 (8), pp. 2224–2240. DOI: 10.1175/MWR3160.1.

McCallum, E. (1990). "THE BURNS' DAY STORM, 25 JANUARY 1990". In: *Weather* 45 (5), pp. 166–173. DOI: 10.1002/j.1477-8696.1990.tb05607.x.

McCullagh, P. and J. A. Nelder (1999). *Generalized linear models*. 2. ed., [R. Chapman & Hall/CRC.

McIlveen, R. (1991). *Beaufort scale*.

Molnar, C. (2019). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*.

Moran, T. (2019). "Factors affecting the severity of European winter windstorms", pp. 1–115. DOI: 10.17169/refubium-29369.

Munich RE (2002). "Winterstürme in Europa (II)". In: *Schadensanalyse 1999 - Schadenpotenziale*, 72pp.

— (2013). "NatCatSERVICE - Naturkatastrophen 1980 bis 2012 im Überblick". In: *TOPICS GEO 2012*, pp. 54–55.

Murray, R. J. and I. Simmonds (1991). "A numerical scheme for tracking cyclone centres from digital data. Part I: development and operation of the scheme". In: *Australian Meteorological Magazine* 39 (3), pp. 155–166.

Neu, U. et al. (2013). "Imilast: A community effort to intercompare extratropical cyclone detection and tracking algorithms". In: *Bulletin of the American Meteorological Society* 94 (4), pp. 529–547. DOI: 10.1175/BAMS-D-11-00154.1.

Nissen, K. M. et al. (2010). "Cyclones causing wind storms in the Mediterranean: characteristics, trends and links to large-scale patterns". In: *Natural Hazards and Earth System Sciences* 10 (7), pp. 1379–1391. DOI: 10.5194/nhess-10-1379-2010.

Paciorek, C. J. et al. (2002). "Multiple Indices of Northern Hemisphere Cyclone Activity, Winters 1949–99". In: *Journal of Climate* 15 (13), pp. 1573–1590. DOI: 10.1175/1520-0442(2002)015<1573:MIONHC>2.0.CO;2.

Pardowitz, T. et al. (2016a). "Estimating uncertainties from high resolution simulations of extreme wind storms and consequences for impacts". In: *Meteorologische Zeitschrift* 25 (5), pp. 531–541. DOI: 10.1127/metz/2016/0582.

Pardowitz, T. et al. (2016b). "Uncertainties in Forecasts of Winter Storm Losses". In: *Natural Hazards and Earth System Sciences Discussions* (July), pp. 1–18. DOI: 10.5194/nhess-2016-182.

Peña, J. M., J. A. Lozano, and P. Larrañaga (1999). "An empirical comparison of four initialization methods for the K-Means algorithm". In: *Pattern Recognition Letters* 20 (10), pp. 1027–1040. DOI: 10.1016/S0167-8655(99)00069-0.

Pfahl, S. and H. Wernli (2012). "Quantifying the Relevance of Cyclones for Precipitation Extremes". In: *Journal of Climate* 25 (19), pp. 6770–6780. DOI: 10.1175/JCLI-D-11-00705.1.

Pielke, R. A. (2007). *Windstorm*. https://www.britannica.com/science/windstorm. Accessed: 2022-03-17.

Pinto, J. G. et al. (2005). "Sensitivities of a cyclone detection and tracking algorithm: individual tracks and climatology". In: *Meteorologische Zeitschrift* 14 (6), pp. 823–838. DOI: 10.1127/0941-2948/2005/0068.

Pinto, J. G. et al. (2009). "Factors contributing to the development of extreme North Atlantic cyclones and their relationship with the NAO". In: *Climate Dynamics* 32 (5), pp. 711–737. DOI: 10.1007/s00382-008-0396-4.

Pirret, J.S. R., P. Knippertz, and T. M. Trzeciak (2017). "Drivers for the deepening of severe European windstorms and their impacts on forecast quality". In: *Quarterly Journal of the Royal Meteorological Society* 143 (702), pp. 309–320. DOI: 10.1002/qj.2923.

Poli, P. et al. (2016). "ERA-20C: An Atmospheric Reanalysis of the Twentieth Century". In: *Journal of Climate* 29 (11), pp. 4083–4097. DOI: 10.1175/JCLI-D-15-0556.1.

Priestley, M. D. K. et al. (2017). "Rossby wave breaking, the upper level jet, and serial clustering of extratropical cyclones in western Europe". In: *Geophysical Research Letters* 44 (1), pp. 514–521. DOI: 10.1002/2016GL071277.

Rex, D. F. (1951). "The Effect of Atlantic Blocking Action upon European Climate". In: *Tellus* 3 (2), pp. 100–112. DOI: 10.1111/j.2153-3490.1951.tb00784.x.

Roberts, J. F. et al. (2014). "The XWS open access catalogue of extreme European windstorms from 1979 to 2012". In: *Natural Hazards and Earth System Sciences* 14 (9), pp. 2487–2501. DOI: 10.5194/nhess-14-2487-2014.

Rodriguez, M. Z. et al. (2019). "Clustering algorithms: A comparative approach". In: *PLOS ONE* 14 (1). Ed. by Hans A Kestler, pp. 1–34. DOI: 10.1371/journal.pone.0210236.

Rousseeuw, P. J. (1987). "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20, pp. 53–65. DOI: 10.1016/0377-0427(87)90125-7.

Shapley, L. S. (1953). *17. A Value for n-Person Games*. DOI: 10.1515/9781400881970-018.

Simmonds, I., R. J. Murray, and R. M. Leighton (1999). "A refinement of cyclone tracking methods with data from FROST". In: *Australian Meteorological Magazine* 48 (SPEC. ISS.), pp. 35–49.

Steinley, D. and M. J. Brusco (2007). "Initializing K-means Batch Clustering: A Critical Evaluation of Several Techniques". In: *Journal of Classification* 24 (1), pp. 99–121. DOI: 10.1007/s00357-007-0003-0.

Stucki, P. et al. (2014). "A catalog of high-impact windstorms in Switzerland since 1859". In: *Natural Hazards and Earth System Sciences* 14.11, pp. 2867–2882. DOI: 10.5194/nhess-14-2867-2014.

Thompson, F. R. and F. A. La Sorte (2008). "Comparison of Methods for Estimating Bird Abundance and Trends From Historical Count Data". In: *Journal of Wildlife Management* 72 (8), pp. 1674–1682. DOI: 10.2193/2008-135.

Thorndike, R. L. (1953). "Who belongs in the family?" In: *Psychometrika* 18 (4), pp. 267–276. DOI: 10.1007/BF02289263.

Tibshirani, R., G. Walther, and T. Hastie (2001). "Estimating the number of clusters in a data set via the gap statistic". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (2), pp. 411–423. DOI: 10.1111/1467-9868.00293.

Ulbrich, U., G. C. Leckebusch, and J. G. Pinto (2009). "Extra-tropical cyclones in the present and future climate: A review". In: *Theoretical and Applied Climatology* 96 (1-2), pp. 117–131. DOI: 10.1007/s00704-008-0083-8.

Ulbrich, U. et al. (2001). "Three extreme storms over Europe in December 1999". In: *Weather* 56 (3), pp. 70–80. DOI: 10.1002/j.1477-8696.2001.tb06540.x.

Walz, M. A. et al. (2018). "Modelling serial clustering and inter-annual variability of European winter windstorms based on large-scale drivers". In: *International Journal of Climatology* 38 (7), pp. 3044–3057. DOI: 10.1002/joc.5481.

Wang, X. L. et al. (2011). "Trends and low-frequency variability of storminess over western Europe, 1878–2007". In: *Climate Dynamics* 37 (11-12), pp. 2355–2371. DOI: 10.1007/s00382-011-1107-0.

Willison, J., W. A. Robinson, and G. M. Lackmann (2015). "North Atlantic Storm-Track Sensitivity to Warming Increases with Model Resolution". In: *Journal of Climate* 28 (11), pp. 4513–4524. DOI: 10.1175/JCLI-D-14-00715.1.

Wu, X. et al. (2008). "Top 10 algorithms in data mining". In: *Knowledge and Information Systems* 14 (1), pp. 1–37. DOI: 10.1007/s10115-007-0114-2.

Yuan, C. and H. Yang (2019). "Research on K-Value Selection Method of K-Means Clustering Algorithm". In: *J* 2 (2), pp. 226–235. DOI: 10.3390/j2020016.

Zhang, Z. and B. A. Colle (2017). "Changes in Extratropical Cyclone Precipitation and Associated Processes during the Twenty-First Century over Eastern North America and the Western Atlantic Using a Cyclone-Relative Approach". In: *Journal of Climate* 30 (21), pp. 8633–8656. DOI: 10.1175/JCLI-D-16-0906.1.

Štrumbelj, E. and I. Kononenko (2014). "Explaining prediction models and individual predictions with feature contributions". In: *Knowledge and Information Systems* 41 (3), pp. 647–665. DOI: 10.1007/s10115-013-0679-x.