# Winter wind storms: Identification, verification of decadal predictions, and regionalization

DISSERTATION

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

am Fachbereich Geowissenschaften

der Freien Universität Berlin

vorgelegt von

Tim Kruschke

Berlin, Dezember 2014

# 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, 18. Dezember 2014
_____

# Abstract

Wind storms globally pose the most important natural hazard from a socio-economic perspective. For the European continent, it is especially winter storms related to synoptic-scale extra-tropical cyclones that often affect several countries at the same time bearing high risk of cumulative loss. Societal and economic stakeholders are interested in different aspects regarding these phenomena. On the one hand, (re-)insurance loss modeling requires high spatio-temporal resolution information for winter storms that happened in the past as well as physically consistent scenarios of storm events that could happen. On the other hand, socio-economic planning activities would benefit from any reliable information regarding the frequency of damage-prone storm events for the upcoming seasons and years.

The current thesis addresses three aspects in this context: (i) It further develops an objective impact-oriented identification scheme regarding such wind storms. (ii) State-of-the-art decadal climate forecasts are analyzed whether they can provide skillful predictions of Northern Hemisphere winter storm frequency. (iii) A statistical downscaling approach is developed, efficiently estimating high resolution surface gusts from coarse reanalysis and model data. All three topics are successfully tackled.

The objective identification procedure is advanced in several aspects, including a more sophisticated spatio-temporal tracking of identified storms. The actual revision of the scheme is applied to the *ERA-Interim*-reanalysis, yielding the first consistent global climatology of recent wind storm climate. In this context, it is shown that the algorithm is also suitable for other than its core target, that is extra-tropical winter storms. Properties of different storm types are compared, revealing several interesting facts. An exemplary result is the systematically higher translation velocity related to greater travel distances of winter storms over the North Pacific when compared to the North Atlantic, resulting into higher storm frequencies for many locations in that region although storm genesis is generally less frequent than over the North Atlantic.

After applying the identification scheme to current decadal climate forecasts based on one specific general circulation model, these are found to offer skillful predictions regarding the average winter storm frequency of the upcoming five to ten years for large parts of the northern-hemispheric extra-tropics. However, most of the current skill arises from external forcing of greenhouse gas and aerosol concentrations. This means that the

enormous effort of initialization provides very limited added value, in particular for the entrance regions of the stormtracks, the Eastern Mediterranean, and parts of the Arctic. Based on a novel approach for correcting potential model drifts and systematic combinations of various hindcast experiments, several different initialization strategies are compared in this respect. However, no significant differences are found in this context. All strategies considered yield the same level of prediction skill.

Statistical downscaling of European winter storm events diagnosed in any type of gridded data set – reanalysis, numerical weather prediction, or climate simulations such as the previously diagnosed decadal predictions – can be performed efficiently by means of the approach developed in this thesis. The method is based on a Stepwise Linear Regression scheme. Thorough analyses on optimal predictors yield a combination of wind magnitudes and vector components as well as their squares derived from a wider environment (up to 300 km) of the respective location to be most appropriate. The method outperforms all other approaches tested, including a recently published simple statistical model providing not only better estimations of the "average" winter storm gust, but also more appropriately discriminating the core storm field against areas featuring comparably weak gusts. The statistical downscaling results are complemented by a reasonable quantitative assessment of its uncertainties, accounting for the heteroscedasticity of the predictands, i.e. the surface gusts.

The synthesis of these three major efforts represents a substantial advance in research on winter wind storms over Europe and the Northern Hemisphere.

# Zusammenfassung

Sturmereignisse stellen global die, aus sozio-ökonomischer Perspektive, bedeutendste Form von Naturkatastrophen dar. Für Europa sind es insbesondere Winterstürme als intensive extra-tropische Zyklonen, die zumeist mehrere Länder gleichzeitig betreffen und damit ein hohes Risiko hinsichtlich sog. Kumul-Schadensereignisse darstellen. Verschiedene Aspekte bzgl. dieser Ereignisse sind von Relevanz für gesellschaftliche und wirtschaftliche Akteure. Einerseits erfordert beispielsweise die Schadensmodellierung von Rückversicherern räumlich und zeitlich hoch aufgelöste Informationen über vergangene Winterstürme und entsprechende physikalisch konsistente Szenarien von möglichen Ereignissen. Auf der anderen Seite könnten verschiedenste gesellschaftliche und ökonomische Planungen von verlässlichen Informationen über die Häufigkeit schadenträchtiger Sturmereignisse in den nächsten Monaten und Jahren profitieren.

Die vorliegende Dissertation beschäftigt sich erfolgreich mit drei Aspekten in diesem Zusammenhang: (i) Sie stellt einen wesentlichen Beitrag zur Weiterentwicklung eines objektiven Verfahrens zur Identifikation von potentiell schadenverursachenden Sturmereignissen dar. (ii) Dekadische Klimaprognosen werden hinsichtlich ihrer Vorhersagegüte in Bezug auf die Frequenz von Winterstürmen über der Nordhemisphäre hin untersucht. (iii) Sie stellt eine neues Verfahren zur statistischen Ableitung hoch aufgelöster bodennaher Windböen aus (räumlich) grob aufgelösten Reanalyse- und Modelldaten vor.

Das objektive Identifikationsverfahren wurde in mehrfacher Hinsicht verbessert, insbesondere wurde die raum-zeitliche Verfolgung identifizierter Ereignisse weiter entwickelt. Die aktuellste Version dieses Algorithmusses wird auf Daten der *ERA-Interim*-Reanalyse angewendet. Das Ergebnis ist die erste konsistente globale Analyse der Klimatologie potentiell schadenträchtiger Sturmereignisse. In dem Zusammenhang kann gezeigt werden, dass das (weiter-)entwickelte Verfahren auch für andere Sturmereignisse als extra-tropische Winterstürme sinnvoll anwendbar ist. Die Eigenschaften verschiedener Sturmtypen werden verglichen. Eine der wesentlichen dabei erzielten Erkenntnisse ist jene hinsichtlich systematisch höherer Zuggeschwindigkeiten und damit weiter zurückgelegter Distanzen von nordpazifischen Winterstürmen im Vergleich zum Nordatlantik, was dazu führt, dass in vielen Bereichen des Nordpazifiks für den gegebenen Ort die Frequenz von Winterstürmen höher ist, obwohl die um die Fläche bereinigte Anzahl entstehender Sturmereignisse geringer ausfällt als im Nordatlantikraum.

Die Anwendung des Identifikationsverfahrens für aktuelle dekadische Vorhersagen eines spezifischen Klimamodells ermöglicht die quantitative Untersuchung der Güte dieser Prognosen. Es zeigt sich, dass für große Teile der außertropischen Nordhemisphäre belastbare Vorhersagen bzgl. der durchschnittlichen Frequenz von Winterstürmen für die kommenden 5 bis 10 Jahre getroffen werden können. Allerdings entstammt ein Großteil dieser Vorhersagbarkeit den externen Antrieben von Treibhausgasen und Aerosolkonzentrationen. Der enorme Aufwand der jährlichen Initialisierung von Vorhersageensembles generiert nur für vergleichsweise wenige Regionen einen Zusatznutzen. Diese sind die Eingangsbereiche der beiden nordhemisphärischen *stormtracks*, sowie eine Region östlich des Mittelmeers und Teile der Arktis. Die Verwendung eines neuartigen Verfahrens zur Korrektur eventueller Modelldrifts und die systematische Kombination verschiedener *Hindcast*-Experimente erlaubt zudem den Vergleich unterschiedlicher Initialisierungsstrategien. Es zeigt sich allerdings, dass keine der bisher verwendeten Strategien deutliche Vorteile hinsichtlich der Vorhersagegüte gegenüber anderen generiert.

Europäische Wintersturmereignisse, die in beliebigen gegitterten Datensätzen (Reanalysen, numerische Wettervorhersagen oder Klimasimulationen, wie die zuvor behandelten dekadischen Prognosen) identifiziert wurden, können mithilfe eines neu entwickelten Verfahrens effizient statistisch regionalisiert werden. Das zugrundeliegende Verfahren ist jenes der Schrittweisen Linearen Regression. Gründliche Analysen hinsichtlich optimaler Prediktoren ergeben dabei, dass eine Kombination aus Windgeschwindigkeiten und der entsprechenden Vektorkomponenten sowie ihrer Quadrate die besten Ergebnisse liefert, sofern für einen gegebenen Ort auch die weitere Umgebung (bis zu 300 km) mit einbezogen wird. Dieses entwickelte Verfahren erzielt deutlich bessere Modellierungsergebnisse als alle anderen getesteten Verfahren, inklusive eines kürzlich veröffentlichten einfacheren Ansatzes. Dabei wird nicht nur die "durchschnittliche" Böe innerhalb von Winterstürmen besser regionalisiert, sondern auch eine angemessenere Abgrenzung des Kernsturmfeldes von Bereichen vergleichsweise schwacher Windgeschwindigkeiten erreicht. Die statistische Regionalisierung wird komplettiert durch eine quantitative Abschätzung der damit verbundenen Unsicherheiten, die der Heteroskedastizität der Prediktanten, also der Windböen, gerecht wird.

Insgesamt stellen diese drei durch die vorliegende Dissertation abgedeckten Themenbereiche einen substantiellen Fortschritt für die Forschung bzgl. Winterstürmen über Europa und der Nordhemisphäre dar.

# Contents

# 1. Preface

On 14 November 1854 – 160 years ago – an intense cyclone associated with hurricane force winds swept over the Black Sea and the Crimean Peninsula. It heavily affected the allied British, French, and Ottoman forces sieging Sevastopol as part of the Crimean War and destroyed substantial parts of their supply fleet as well as several battle ships, including the *Henri IV*, the "pride of the French navy". This catastrophic event caused the French Minister of War assigning the famous astronomer Urbain Le Verrier to study this storm. Le Verrier gathered meteorological observations for the week 11–16 November from approx. 250 European stations. Analyzing that data, he found that the storm could be tracked while intensifying along its path from the Atlantic across Europe (Landsberg, 1954; Lindgrén and Neumann, 1980). He even stated that a warning could have been issued one day in advance, given the available data in time:

> "En apprenant à Vienne que la tempête avait sévi à telle heure sur les côtes de France, à telle heure à Paris, à telle heure à Munich, et toujours en augmentant d'intensité, ne pouvait-on prévoir qu'elle allait atteindre la mer Noire ?"
> (Le Verrier, 31 January 1855, towards the French Academy of Science; according to Fiero, 1991, pp. 110–111)

Consequently, Le Verrier successfully wrote a proposal to Napoleon III – the Emperor of the Second French Empire – regarding a project for storm warnings, based on the possibility to exchange observational data in a timely manner by using the Empire's telegraphic lines. He initiated the *Bulletin International de l'Observatoire de Paris* released in 1858 with daily observations from 18 cities (4 outside of France) and – since 1863 – synoptic weather maps (Monmonier, 1999). Basically, this early (manual) tracking of the destructive "Storm of Balaklava" was the impetus for the first governmental meteorological forecasting service (Landsberg, 1954).

## 1.1. Wind storm identification

Until today – 160 years later – wind storms pose the globally most important natural hazard from the socio-economic perspective. Such events were responsible for approx. 720 bnUS\$ of insured losses in the period 1980–2013 which is equivalent to ~56% of all insured losses related to natural catastrophes (Munich RE Group, 2014b). The wind storm type being most relevant differs with the region considered. For Europe, it is winter storms related to synoptic scale extra-tropical cyclones that are associated with by far the largest economic impacts (Munich RE Group, 2008a; Handmer et al., 2012). If North America or East Asia is considered, tropical cyclones (called "hurricanes" or "typhoons" in the respective regions) are generally the most expensive events (Peterson et al., 2008; Handmer et al., 2012; Welker and Faust, 2013). In Australia, hailstorms outweigh other hazards, but tropical cyclones are the second most expensive type of natural disasters (Crompton and McAneney, 2008).

These economic impacts (e.g. Pielke et al., 2008; Barredo, 2010; Della-Marta et al., 2010; Donat et al., 2011a; Welker and Faust, 2013) but also their significant influence on other meteorological parameters (Paredes et al., 2006; Raible et al., 2007; Raible, 2007; Hawcroft et al., 2012; Pfahl and Wernli, 2012) or differing components of the climate system (e.g. Yao et al., 2008; Papritz et al., 2014) is the motivation for various scientific studies. Many of these studies follow a Lagrangian approach, i.e. individual systems are identified and tracked. In this respect, no general distinction exists between process-oriented studies (e.g. Dacre et al., 2012; Daloz et al., 2012; Fink et al., 2012; Martinez-Alvarado et al., 2012), analyses focused on climatological characteristics (e.g. Sinclair, 1994; Serreze, 1995; Emanuel, 2000; Manganello et al., 2012), low-frequency variability (e.g. Luksch et al., 2005; Wang et al., 2013), or potential climate change (e.g. Lionello et al., 2002; Webster et al., 2005; Nissen et al., 2014b; Grieger et al., 2014) regarding the respective systems.

Urbain Le Verrier in 1854/55 analyzed a single storm based on comparably sparse and irregularly distributed observational data. For such a topic, a manual and, hence, subjective identification and tracking of the targeted feature may be appropriate. The actual state of the art for other analyses than case studies is to utilize automated and objective schemes, operating efficiently and comprehensibly on large data sets, nowadays available from gridded reanalyses or model simulations. Chapter 2 of this thesis

presents a brief overview (Sec. 2.2) of existing methods in this respect and describes a rather new approach – the *WiTRACK*-algorithm first introduced by Leckebusch et al. (2008a). Differing to most of such established procedures targeted at specific physical characteristics of different storm types, the WiTRACK-scheme is narrowly focused on the potential impacts of wind storm events. Subsequent to a profound description of the actual state of the algorithm (Sec. 2.3), an application of the scheme targeted at meso- to synoptic-scale wind storms is presented in Sec. 2.4. While the current thesis in general focuses on extra-tropical wind storms over the Northern Hemisphere (NH), this application proves WiTRACK being appropriate for the Southern Hemisphere (SH) and wind storms related to tropical cyclones as well. Spatial distributions of the identified storm events as well as seasonal and inter-annual variability and climatological properties for different regions and storm types are examined. Apart from these climatological analyses – to the author's knowledge never before addressed globally in a consistent way – this Chapter 2 constitutes a profound documentation of the method used for the subsequent parts of this thesis.

## 1.2. Wind storm predictions

In early 1855 Le Verrier was convinced that the "Storm of Balaklava" and similar events could be forecast one day ahead by comprehending observational data drawn into synoptic weather maps. Retrospectively, this opinion has to be judged as very ambitious given the available data and state of meteorological knowledge at this time (Landsberg, 1954). Certainly, the scientific understanding of atmospheric processes has evolved dramatically since the times of Le Verrier and computational capacities allow operational forecasts – as deterministic singular model simulations or probabilistic by means of ensemble integrations – for the upcoming days, months, and years. However, the skill inherent to predictions regarding these time scales is very different. There is no doubt that today's short- and medium-range weather forecasts with respect to tropical (see e.g. Kurihara et al., 1995; Goerss, 2000; Rappaport et al., 2009; Tsai et al., 2011) and extra-tropical cyclones (see e.g. Buizza and Hollingsworth, 2002; Jung et al., 2004, 2005; Froude et al., 2007a,b; Froude, 2009, 2010, 2011; Colle and Charles, 2011) provide reliable information used e.g. for specific warnings. Even for more challenging situations of tropical cyclones transitioning to the extra-tropics (e.g. Buckingham et al., 2010) or smaller-scale

events such as Polar Lows (see e.g. Kristiansen et al., 2011) or Medicanes (e.g. Pantillon et al., 2013) clear indications of prediction skill regarding this time scale exist. When it comes to seasonal or even longer time scales, no skillful predictions of individual storms can be made. Instead, spatial and temporal aggregation is performed, e.g. considering the number of events (frequency) to be expected for a certain region. The physical mechanism underlying such long-term predictions generally is the concept already described by Bjerknes (1964), that is the atmosphere driving the inert ocean on short time scales and an oceanic feedback to the atmosphere on longer time scales. Early studies with respect to seasonal forecasts of tropical (e.g. Gray et al., 1994; Chan et al., 1998) or extra-tropical (DeGaetano et al., 2002) storm frequency were based on statistical means, deriving estimations of cyclone activity from various large-scale predictors. It should be noted in this context, that the latter study of DeGaetano et al. (2002) partly employs predictor variables observed coincidently to the predictands, and hence, demonstrates potential predictability only. Today, seasonal predictions of tropical cyclone frequency are also directly diagnosed from dynamical model's output (see e.g. Vitart et al., 2012) and consensus establishes that these predictions contain valuable information (see e.g. Vitart et al., 2007; LaRow et al., 2008; Vecchi et al., 2014). Multi-annual to decadal prediction is still in an experimental stage, fostered by the *Coupled Model Intercomparison Project* in its fifth phase (CMIP5). Nevertheless, some studies regarding such forecasts of tropical cyclone activity already exist – derived directly from model output (Smith et al., 2010; Mori et al., 2013) or utilizing statistical relationships to large-scale precursors (Vecchi et al., 2013; Caron et al., 2014). They yield encouraging results, though reliability of their skill is still under debate. In the extra-tropics, the relation of atmospheric forcing and oceanic feedback generally seems to be notably less favorable in this respect (see e.g. Frankignoul, 1985; Kushnir et al., 2002). However, two studies of Renggli (2011) and Renggli et al. (2011) show that actual seasonal prediction systems do exhibit some skill regarding wind storm frequency over the Northeast Atlantic and Europe for the upcoming (winter) months. The author of this thesis (in collaboration with several co-authors) published the only study available regarding multi-annual to decadal predictions of (NH winter) extra-tropical cyclone frequencies (Kruschke et al., 2014). Building up on the encouraging results presented therein, Chapter 3 of this thesis (to a great extent identical to the first submitted version of the paper by Kruschke et al., 2015) presents further analyses in this respect. Compared to Kruschke et al. (2014) it follows a more impact-oriented approach by applying the wind storm identi-

fication scheme, presented in Ch. 2, utilizes a more elaborated method to account for potential model drifts (Sec. 3.4), and considers more hindcast experiments (and different combinations; see Sec. 3.2.1) yielding more robust estimates of decadal prediction skill.

## 1.3. Winter wind storm regionalization

The promotion of Le Verrier's studies by the French Minister of War was highly motivated by the losses – the French suffered from the "Storm of Balaklava", in particular that of the battleship *Henri IV* (Landsberg, 1954; Lindgrén and Neumann, 1980). Similar catastrophes happen today, too. The European winter storm "Kyrill" in January 2007 made the container ship *MSC Napoli* flounder in the English Channel. More than 110 of its ∼2,300 containers went overboard (Munich RE Group, 2008b). Overall, by striking most parts of Western and Central Europe, "Kyrill" was responsible for approx. 5.8 bnUS$ of insured losses and 10 bnUS$ of total economic losses (Munich RE Group, 2014a). Damage risk for a specific object, such as the *Henri IV* or *MSC Napoli*, but also total economic loss associated with a specific storm depends on a combination of several parameters. A basic factor is exposure: the most extreme event will not produce any harm if no value (economic or other, such as life) is present in the hazardous area. Hence, exposure is a socio-economic variable determined by the spatial distribution of people, property, and infrastructure. A second factor is the meteorological event itself, its overall properties and structure. This includes parameters such as the track of the storm, its size and intensity as well as the respective evolution over lifetime of the event. Apparently, if the "Storm of Balaklava" had weakened on its way over Europe, it would not have had such devastating effects on the allied forces at Crimea. And if "Kyrill" had taken a track a few hundred kilometers north, the *MSC Napoli* probably would have gotten in less severe trouble. Apart from such synoptic scale characteristics, the meso-scale structure and dynamics within the particular event are also highly relevant. In-depth analyses of "Kyrill" (Fink et al., 2009; Gatzen et al., 2011) showed that its cold front was associated with strong convective activity embedded, leading to significant differences regarding wind speeds and gusts on spatial scales of less than 100 km (visible as "streaky" patterns of maximum gusts in dynamical regionalization results presented in Fig. 4.7). Additionally, surface winds are heavily influenced by geophysical conditions, namely orography (Smith, 1979, 1985) and surface roughness (Wieringa, 1986, 1993).

The above-mentioned fact that winter wind storms are the major natural hazard in terms of their economic impacts is the reason for the (re-)insurance industry being committed to research on these phenomena. Facing that even rather small differences in storm intensity can be associated with large variations for damage (Schwierz et al., 2010) in combination with mostly very heterogeneously distributed values requires wind fields of high spatial resolution for insurance loss models (Haylock, 2011). For a limited domain such high resolution wind data may be derived from interpolating observations (e.g. Etienne and Beniston, 2012). On the other hand, many of these stakeholders act internationally and synoptic scale European winter storms often affect multiple countries. Hence, its is not only high resolution but also continental scale coverage that is demanded for follow-on applications by these end-users. The natural solution for such demands is the usage of *Limited Area Models* from numerical weather prediction or *Regional Climate Models* (RCM). However, given the computational costs of such simulations this is feasible for a limited number of storm episodes only. The alternative in this respect is statistically achieving the desired fields. Chapter 4 presents a method developed in this context. Building up on a set of dynamically downscaled storm situations, statistical relationships between coarse scale predictors and high resolution surface gusts are detected and subsequently used for statistical regionalization. Sec. 4.4 documents a systematic analysis regarding the optimal statistical model, finally yielding results better than those of the recently published study of Haas and Pinto (2012) and of the same quality as dynamical downscaling when compared to observations. The developed statistical downscaling exhibits high efficiency in providing high resolution gust fields for the whole European continent, outperforming dynamical simulations by far in this respect. Its flexible design allows application to a wide range of coarse scale model simulations. Besides medium-range weather forecasts, usable for tailored warnings if conducted in real-time or climatological analyses analog to coarse data counterparts such as Osinski (2014); Osinski et al. (2015), the downscaling procedure is also applicable to decadal climate predictions as considered in Ch. 3 or centennial climate projections allowing related analyses on high spatial resolution.

# 2. Objective identification and tracking of wind storms

This Chapter contains a comprehensive overview with respect to WiTRACK, the automatic and objective scheme for identifying and tracking potentially damaging wind storm events. Besides constituting an essential part of this thesis and comprehensively introducing the method used for the following Chapters, Sec. 2.3 is meant to fill the gap of a yet missing profound documentation of WiTRACK. Special emphasis in this respect is put on Sec. 2.3.3, containing the developments and improvements since the introduction of WiTRACK by Leckebusch et al. (2008a) and Renggli (2011). As only part of these developments are achievements of the author of this thesis, Sec. 2.3.3 contains references to the originator(s) of the respective improvements. Parts of this Chapter will be used as a basis for a manuscript currently in preparation.

## 2.1. Introduction

The objectives pursued by this Chapter are twofold:

First and basically, it is meant to provide a comprehensive overview with respect to WiTRACK, the automatic and objective scheme for identifying and tracking potentially damaging wind storm events. As such, this Chapter of the current thesis is the first profound documentation of this algorithm and the underlying concepts. This is especially true for the developments made after the introduction of the scheme by Leckebusch et al. (2008a) and Renggli (2011). Most of the developments rated as "substantial" in Sec. 2.3.3 are achievements of this thesis' author. Hence, this Chapter summarizes the WiTRACK-related methodological progress achieved in the course of the author's doctorate proposition. In this context, the current Chapter constitutes a comprehensive description of the event identification method used for the following Chapters 3 and 4.

Second, several scientific questions are addressed by this Chapter, some of them being of methodological nature, others purely focused on meteorological and climatological aspects.

Regarding the methodological part, two main issues are treated: (i) Did the developments described in Sec. 2.3.3 improve WiTRACK regarding its purpose, that is, identifying potentially damaging wind storms? (ii) Is WiTRACK and the used configuration a suitable tool for identifying wind storms in yet untested regions, that is, others than the Northern Hemisphere (NH) extra-tropics?

Regarding the climatological part, also two major topics are covered: (a) A global climatology of potentially damaging meso- to synoptic-scale wind storms is presented, including analyses regarding seasonal and inter-annual variability as well as potential trends over the recent past, covered by the *ERA-Interim* data. (b) Potentially damaging wind storms from different regions and seasons – assumed to represent different storm types in the sense of their meteorological nature – are analyzed with respect to various characteristics, all of them influencing the overall potential threat associated with the respective events.

This Chapter is structured as followed: Sec. 2.2 contains a general introduction into the topic of automatic feature tracking, as established and applied in the field of meteorology and climatology. Sec. 2.3 is dedicated to the respective method, developed at

the *Institute of Meteorology* at *Freie Universität Berlin*, targeted at the identification and tracking of damage-prone wind fields. In this context, special emphasis is put on the advances achieved as part of the current thesis. An application of this algorithm is presented in Sec. 2.4, meant to answer the climatological and methodological questions, raised above. This Chapter is completed by a summary presented in Sec. 2.5.

## 2.2. Established procedures for automatic meteorological feature tracking

Generally, a meteorological feature or event is characterized by certain properties, that are clearly different from the mean state and/or the environmental background. Some meteorological parameters affected by the specific event exhibit "extreme" behavior in terms of absolute values or with respect to its spatial/temporal derivative. Examples of phenomena to be defined as such features and their characteristics are: clouds (increased optical depth, high humidity, potentially precipitation), thunderstorms (deep convection and, hence, clouds, probable heavy precipitation, lightning), fronts (distinct gradients of temperature and humidity, associated with several weather phenomena), cyclones (low surface core pressure, local extreme in vorticity, often high wind speeds, extra-tropical systems associated with fronts). Hence, such features are more or less obvious indicators of meteorological processes and often related to more or less severe impacts on human society but also on ecosystems and other components of the climate system. Therefore, analyses focusing on meteorological features/events (i) foster the understanding of underlying processes (e.g. Gray and Dacre, 2006; Fink et al., 2012, both analyzing mechanisms forcing the development of extra-tropical cyclones), (ii) are an essential part in evaluating complex weather/climate models regarding their performance in representing such features and underlying processes (e.g. Daloz et al., 2012; Zappa et al., 2013, analyzing GCMs for their ability to simulate tropical and extra-tropical cyclones, respectively), which in turn helps explaining and resolving model deficiencies, and (iii) allow tailored forecasts focused on relevant events (e.g. Vitart et al., 2012, describing operational *ECMWF* forecasts of tropical cyclone paths and intensity) and event-specific forecast verification (e.g. Ebert and McBride, 2000, verifying predictions of specific precipitation events).

In the recent past, primarily two fields within meteorology and climatology made use of objective feature identification and tracking methods. On the one hand, there is operational weather forecasting applying such methods mainly for three purposes, that is (i) data assimilation of atmospheric motion vectors, derived from clouds tracked over sequences of satellite images (see e.g. Velden et al., 2005), (ii) "nowcasting" focused on convective cells and thunderstorms tracked by methods such as the *TITAN*-algorithm (Dixon and Wiener, 1993, until today being part of the *NCAR*'s automatic nowcast system, described by Mueller et al., 2003) or the algorithms developed at the *DLR* (Zinner et al., 2008; Kober and Tafferner, 2009) and subsequently continued into the near future (forecasts regarding next 6–12 hours) primarily by statistical methods, and (iii) medium-range predictions (several days ahead) of tropical cyclones tracked in numerical weather prediction simulations (e.g. Vitart et al., 2012). On the other hand, there is climatological analysis with respect to tropical (e.g. Bengtsson et al., 2007; Manganello et al., 2012) and extra-tropical (e.g. Neu et al., 2013; Ulbrich et al., 2013) cyclonic features and fronts (e.g. Berry et al., 2011; Simmonds et al., 2012; Papritz et al., 2014) as well as polar meso-cyclones (e.g. Blechschmidt, 2008; Zappa et al., 2014).

Although the phenomena considered are partly very different, the methodological approaches of many tracking schemes are quite similar. A comprehensive review of these approaches is beyond the scope of this Section, but the main steps and commonly found solutions are presented by few examples focused on (extra-tropical) cyclonic features. A thorough overview and comparison of 15 widely-used methods in this respect is given by Neu et al. (2013). Basically, the major challenges are the objective identification of targeted features and subsequently the spatio-temporal tracking of the specific events, both steps possibly to be designed at any desired complexity.

Regarding the identification stage, two basic issues are to be addressed. First, a useful parameter or combination of parameters characterizing the considered feature has to be chosen, additionally requiring to account for potential shortcomings of available datasets in that respect. Focusing on extra-tropical cyclones, most of the methods participating in the *IMILAST*-effort (Neu et al., 2013) consider sea-level pressure (e.g. Murray and Simmonds, 1991; Serreze, 1995; Blender et al., 1997; Lionello et al., 2002) and/or vorticity (e.g. Murray and Simmonds, 1991; Sinclair, 1994; Serreze, 1995; Inatsu, 2009). The latter is also analyzed by the probably most widely used approach of Hodges (1994). Alternatively, geopotential height at some pressure level is a suitable parameter (see e.g.

Kew et al., 2010). Related to the choice of a suitable meteorological field is the property, the chosen parameter is examined for. The basic question here is whether the targeted phenomenon is to be identified as a point-like feature, e.g. as a local extremum (e.g. Blender et al., 1997; Hanley and Caballero, 2012, both considering SLP-minima), or a feature of certain spatial extent (e.g. Hodges, 1994; Kew et al., 2010).

Once the feature identification is finished, the second major challenge is to find suitable combinations of the unique features forming reasonable events from the meteorological perspective. The basic difference between the methods targeted at (extra-tropical) cyclones is whether they analyze potential connections independent of each other or apply some kind of cost-function (to be minimized) integrating a set of (all) possible connections simultaneously (e.g. Murray and Simmonds, 1991; Hodges, 1994; Sinclair, 1994; Hewson and Titley, 2010). Both methods, independent connecting and cost-function approaches, require some criteria to decide whether a connection is probable or not. The most simple criterion in this respect is nearest-neighbor-assignment in combination with some threshold for maximum allowed distance (e.g. Serreze, 1995; Blender et al., 1997; Inatsu, 2009). Other methods conduct some "prediction" of the subsequent position based on the prior track (e.g. Wernli and Schwierz, 2006; Hanley and Caballero, 2012) and/or account for climatological behavior (e.g. Sinclair, 1994) and/or steering winds (e.g. Hewson and Titley, 2010).

As emphasized by Neu et al. (2013), there is no single best way to objectively identify and track cyclones or meteorological features in general. The choice of an appropriate method always depends on the targeted feature and the framework of the pursued analysis including technical constraints such as data availability or computational limitations.

## 2.3. The WiTRACK-algorithm

### 2.3.1. Motivation of the WiTRACK-scheme

WiTRACK – the objective identification and tracking algorithm, described in the following – was firstly introduced by Leckebusch et al. (2008a). That study focuses on the characterization of winter storms over Europe and the North Atlantic in terms of their (potential) socio-economic impact. Following the results of Klawa and Ulbrich (2003,

and personal communication to several reinsurance and primary insurance companies), such impact is mainly driven by wind damage, while other loss sources (e.g. hail and rain) are negligible for these events. Hence, Leckebusch et al. (2008a) concentrate on surface wind speeds and define an objective storm severity measure, the *Storm Severity Index* (SSI)

$$SSI = \frac{1}{A_0} \sum_t^T \sum_k^K \left( \max \left( 0, \frac{v_{k,t}}{v_{98,k}} - 1 \right) \right)^3 \cdot A_k \qquad (2.1)$$

This is done by calculating the cubic relative exceedance of wind speed ($v_{k,t}$) for a given time $t$ and location (i.e. grid box $k$) over the local climatological $98^{th}$ percentile of the same parameter ($v_{98,k}$). This key element of the SSI definition is identical to the storm loss model presented by Klawa and Ulbrich (2003). According to this definition, the local climatological $98^{th}$ percentile constitutes a threshold for (potential) wind damage, which is in line with the study of Palutikof and Skellern (1991). This model implicitly accounts for adaptation of infrastructure and building standards to local wind climate. Estimating potential wind damage for a certain region and period can be done by summing over all grid boxes $k$ belonging to this region and all time steps $t$ for the considered period. This approach constitutes the *Area Storm Severity Index* (ASSI), defined by Leckebusch et al. (2008a).

An alternative to this Eulerian perspective is the identification and tracking of wind storm events. Subsequently, an *Event Storm Severity Index* (ESSI) can be calculated by considering only those grid boxes, identified to be affected by the given event and summing over all time steps during the life cycle of the identified storm event. The important difference with this approach is, that not necessarily the same grid boxes are considered for different time steps. This approach was also introduced by Leckebusch et al. (2008a), however, a more profound description of the underlying event identification and tracking scheme – the initial version of WiTRACK – is given by Renggli (2011). The name *WiTRACK* accounts for the core purpose of the algorithm: tracking (extreme) wind fields.

## 2.3.2. Previous work: Principles of the (initial) WiTRACK-algorithm

In summary, the core of the initial WiTRACK-algorithm – after reading the input data; before calculating and writing the output – consists of the following steps:

1. Identification of contiguous areas, exceeding the given threshold

2. Check for minimum size of identified clusters

3. Tracking of events over consecutive time steps

4. Check for minimum life time of tracked events

The remainder of this Section is dedicated to a more detailed description of these steps in order to set the scene for Sec. 2.3.3, containing changes and improvements regarding WiTRACK, that have been developed since the studies of Leckebusch et al. (2008a) and Renggli (2011). A schematic overview of WiTRACK is shown in Fig. 2.1, though presenting the actual WiTRACK-functionality not matching all details of the initial WiTRACK-algorithm, described in this Section.

### Identification

Following the above-stated line of argument, an objective impact-oriented identification of a severe wind storm can be done by searching for wind speeds exceeding the local climatological $98^{th}$ percentile. Thus, the first step of the WiTRACK-algorithm incorporates this analysis of identifying contiguous fields (clusters) of wind speeds exceeding this threshold.

### Size check

The second step of WiTRACK consists of analyzing whether the identified clusters are larger than a pre-defined minimum size. Reasons for this step are (i) the elimination of artifacts from noisy data and (ii) the ability to focus on certain events of characteristic size. The latter reason is of special importance given the fact that WiTRACK was developed primarily to identify wind storms related to extra-tropical cyclones, which are larger than all other types of damaging wind storm events (tropical cyclones, meso-

cyclones). The initial WiTRACK-algorithm as presented by Leckebusch et al. (2008a) and Renggli (2011) required the user to define this minimum size as a number of grid boxes (at the equator). The latitude-dependency of grid box areas was accounted for by the cosine function. With respect to the SSI calculation Leckebusch et al. (2008a) and Renggli (2011) weighted the individual grid boxes equivalently ($A_k = cos(\phi)$ in Eq. 2.1) and scaled with the area of one grid box at the equator which is $A_0 = cos(0°) = 1$. This kind of scaling is the reason why Leckebusch et al. (2008a) present no reference area $A_0$. The disadvantage of this approach is that SSI-values calculated for datasets of differing resolution are not comparable, as it does not include any reference to the real area covered by a grid box of the respective dataset.

**Tracking**

The most complex part of WiTRACK is actually tracking, that is the connection of clusters to corresponding features in subsequent time steps. The tracking routine of WiTRACK can be subdivided into four parts: position calculation, forward assignment (connecting), backward assignment (treatment of mergers), and the treatment of boundary problems.

**Position calculation:**  A point-like position or cluster center is calculated for each cluster as the weighted average longitude and latitude of all grid boxes belonging to the cluster. The grid box weighting consists of the respective contribution to the ESSI, that is the grid box area multiplied by the cube of the relative exceedance of the 98th percentile. It should be pointed out, that this cluster center is not necessarily lying within the entity of the cluster. For example, the identified wind fields of extra-tropical cyclones usually follow their fronts and hence, are often shaped like bows or hooks. In these cases the cluster position is sometimes placed to the concave side of the cluster.

**Forward assignment (Connecting):**  WiTRACK was developed for gridded datasets with a temporal resolution of six hours or less. It was found, that for such comparably high temporal resolution a simple nearest-neighbor-approach yields satisfying results. Hence, a cluster of a given time step is basically connected to the closest cluster of the subsequent time step. Two additional requirements have to be fulfilled (according to

the initial WiTRACK-algorithm): (i) a pre-defined maximum distance – set to 720 km for 6-hourly data (equivalent to translation velocity of 120 km/h) and European winter storms as targeted events – must not be exceeded; (ii) the shape of the clusters must be similar, where similarity refers to an overlap of at least 30% of the cluster grid boxes after shifting along the difference vector between the two cluster centers. The latter requirement is somehow problematic, as WiTRACK does not account for the latitude-dependency of grid box areas during this routine.

**Backward assignment (Treatment of mergers):** After the forward assignment is completed, WiTRACK has to deal with potential mergers, that is if two (or more) clusters of the previous time step were connected to the same cluster in the actual time step. The chosen solution for the initial WiTRACK-scheme was to continue the track, which was existing longer until this time step and to delete the other connection. If two (or more) of these merging tracks are of equal (temporal) length, the track incorporating the preceding cluster with the largest overlap is continued, again a problematic procedure given the neglecting of latitude-dependent grid box-areas.

**Treatment of boundary problems:** WiTRACK offers the possibility to consider a certain region only when identifying and tracking wind storms. In fact, all studies using WiTRACK that are published so far make use of this possibility. For such analyses of regional domains the cluster identification is easier (see. *Substantial modifications* in Sec. 2.3.3 regarding column-periodical and global fields), the tracking procedure is affected by boundary problems. Features already existing outside of the analyzed domain may travel inside and events identified and tracked within the domain may travel outside. For such regional analyses, a sponge-like boundary zone has to be defined. Once a track enters this boundary zone, the according cluster center is stored as the last position (event lysis) for the respective track, no further clusters in subsequent time steps are added. Please note, that this procedure implicitly means that for a pre-defined minimum lifetime (see next paragraph) of more than one time step, new tracks can be established (event genesis) only within the inner domain.

**Lifetime check**

For similar reasons as the check for a minimum size, that is (i) the elimination of artifacts from noisy data and (ii) the ability to focus on certain events (of characteristic lifetime), the last step of WiTRACK consists of analyzing whether the wind storm events could be tracked for a pre-defined minimum lifetime (number of time steps). For the core purpose of WiTRACK, the identification and tracking of European winter storms, this threshold was set to four 6-hourly time steps, equivalent to a minimum lifetime of 18 hours.

Besides the already cited studies of Leckebusch et al. (2008a) and Renggli (2011), the so far described initial WiTRACK-algorithm was used by Nissen et al. (2010), Renggli et al. (2011), Pardowitz et al. (2014), Pardowitz (2014), and Renggli et al. (2014).

## 2.3.3. New developments and progress of the WiTRACK-algorithm

This Section is meant to present a comprehensive overview regarding all WiTRACK-developments since its introduction by Leckebusch et al. (2008a) and Renggli (2011). These developments can be sub-divided into three categories: technical developments, changes to the scheme for reasons of flexibilization, and substantial modifications.

**Technical developments**

**Code revisions for enhancing performance (achievement of Philip Lorenz):** The whole WiTRACK source code was revised in order to enhance efficiency and performance of the algorithm. Simple changes like re-arrangement of loops and optimization of matrix operations lead to substantial benefits of approx. 90% shorter computing time and 50% less memory usage.

**Modularization of WiTRACK source code (achievement of Daniel Befort):** The WiTRACK source code is written in *FORTRAN90*. Some data pre-processing is done by a *shell*-script that finally addresses the compiled WiTRACK binary file. The source code of the initial WiTRACK scheme was written in one single file of approx. 1150 lines of code. In the current version (Revision 211; see below) of WiTRACK – now approx.

2250 lines of source code – the individual subroutines are separated into module files. This effort makes the code much easier to review and less conflict-prone for parallel modifications of different developers.

**Code repository and revision control (idea and set up by Philip Lorenz):** After WiTRACK was introduced by Leckebusch et al. (2008a), so far developed only by one programmer (Dominik Renggli), the algorithm became much more spread in terms of application but also for developing. Temporarily, eight people were working parallel on the code. This growth required some effort for managing parallel developments, fostering traceability of the overall progress, and ensuring reproducibility of WiTRACK-related results. The natural solution was to set up a software repository – managed via *Apache Subversion* (SVN) – for the WiTRACK-code and related scripts. Each time changes to one of the files belonging to WiTRACK are checked in to the repository, a new revision number (basically a counter) is assigned to the respective version. To demonstrate the traceability: The change from revision 126 (r126) to r127 (initiated by Tim Kruschke and checked in 3 March 2012) was that an additional header line is printed into the basic output files (see 2.3.4) containing information about the underlying revision of WiTRACK. This development guarantees the reproducibility results – since each (old) revision can be restored from the repository – and facilitates clarifying causes of potentially differing results, produced by different WiTRACK revisions.

**Cluster file output in netcdf- or grib-format (mainly achievement of Daniel Befort):** WiTRACK offers the optional output of grid files containing information about the actual wind field clusters belonging to the tracked wind storm events (see 2.3.4). These output grid files of the initial WiTRACK scheme were unformatted binary files, complemented by related description (text) files. Since r57 WiTRACK offers the possibility to write out *netcdf*- or *grib*-files instead, the latter produced from a *netcdf*-file by means of the *Climate Data Operators* (`cdo -f grb copy`). Usage of these more standardized file formats facilitates further processing with standard software tools.

**Mask and multiplicator file(s) for additional SSI-related indices (joint effort of Tim Kruschke, Katrin Nissen, and Jens Grieger):** A modification of the initial WiTRACK scheme, made by Kruschke (2008), contained the incorporation of a mask file in order

to calculate additional SSI-related indices. Kruschke (2008) used a land-sea-mask for an ESSI only from land grid boxes ($ESSI_l$). Besides such a dichotomous mask, Kruschke (2008) also used a gridded file of population densities as a multiplicator file to derive an event based *Storm Impact Index* (SII) following the storm-loss model developed by Klawa and Ulbrich (2003). The SII is calculated analogue to the SSI (see Eq. 2.1) but additionally multiplying the storm intensity at each individual grid box by the respective population density ($P_k$). These developments were later revived and implemented into WiTRACK (r146) by Katrin Nissen and Jens Grieger. Their developments were again taken up by Tim Kruschke and modified in the sense that a single multiplicator file (in any *cdo*-compatible format, e.g. *grib* or *netcdf*) can be specified for WiTRACK, containing any desired number of multiplicator or mask fields along the third dimension of the data matrix (r151). WiTRACK automatically detects the number of multiplicator fields, calculates the resulting indices and adjusts the output accordingly. The results of the WiTRACK-application presented in Sec. 2.4 contain "Region"-ESSIs – i.e. ESSIs calculated by considering only grid boxes in certain domains – for all 26 regions defined by the IPCC (2012).

**Flexibilization**

**Configuration files for event-specific WiTRACK-parameters (achievement of Daniel Befort):** All event-specific WiTRACK-parameters (core tracking-parameters) formerly widely spread over batch-script and WiTRACK source code were concentrated in configuration files. This purely technical change makes WiTRACK much easier to use, especially when changing tracking parameters, and hence, its application more flexible. Tab. 2.1 contains a list of all variables and tracking parameters to be set in the configuration file according to r211, which is the actual revision, including all developments described in this Sec. 2.3.3. These configuration files (with the file suffix `.txt`) are sourced in the batch script, hence, *bash*-syntax can be used here, e.g. by using variables already specified.

Only few variables have to be set in the batch-script (called `batch_wtrack_v2.sh`, including extensive explanatory comments). The most important is to specify the "`model`", which must be identical to the name of the respective configuration file without the file suffix `.txt`. Apart from that, only technical things may be adapted. Modifi-

cations regarding temporary and output directories can be done here and the switch `writegridfile` sets whether cluster files are written and which format is used (see subsection *Technical* and Sec. 2.3.4). Additionally, the variables `maxclust`, `maxevents`, and `maxlength` specify the maximum number of clusters per tracking period, the maximum number of events per tracking period, and the maximum length (number of time steps) for the individual event. These parameters are primarily of technical meaning, as they are of high importance for the memory requirements of WiTRACK. However, they depend on the targeted events and according tracking parameters. If a very low threshold is used for the event (exceeding the threshold) definition or the minimum cluster size is set comparably low, more clusters will be identified, potentially defying `maxclust`. The same is true for `maxevents`, additionally and heavily depending on the minimum lifetime of the tracks. The temporal length of the tracks and potential conflicts with `maxlength` are a main characteristic of the considered events and the data analyzed for this purpose. A violation of one these parameters will stop the tracking (for the actual tracking period) and produce a respective warning message.

**Real grid box area for all purposes (achievement of Tim Kruschke):** As mentioned in the description of the size check routine in Sec. 2.3.2, the initial WiTRACK-algorithm simply accounted for the latitude-dependent area of the analyzed grid boxes by the cosine function. This is completely sufficient for all tracking-related purposes, as these require an appropriate relative area weighting only. However, this approach suffers from a lack of comparability with respect to SSIs calculated from data sets of different spatial resolution. To overcome this issue, the determination of real grid box areas and subsequent usage for the size check and tracking routines as well as the SSI-calculations was implemented (first in r55, consequently since r123). In order to keep the SSI a dimensionless quantity and to scale the calculated numbers to values similar to those published by Leckebusch et al. (2008a) and Renggli (2011), $A_0$ in Eq. 2.1 was set to 12,363.7 km$^2$, approximately the equivalent of a 1°x1°-grid-box at the equator. A value of km$^2$ for the minimum cluster size (`minarea`) is approximately the equivalent of the thresholds used by Leckebusch et al. (2008a) and Renggli (2011) for the objective identification (and tracking) of mid-latitude winter wind storms.

**Flexibilization of additional parameters and principles of WiTRACK (achievement of Tim Kruschke):** An important step in developing the initial WiTRACK scheme were sensitivity studies, performed by Renggli (2011), in order to find the optimum parameters for the core purpose of WiTRACK, tracking extra-tropical winter storms. One element of this thesis is to test the applicability and transferability of WiTRACK to other types of wind storm events. For this purpose, additional flexibilization became necessary. Hence, formerly fixed parameters were made flexible, i.e. configurable. To be specific, while the initial WiTRACK scheme used a fixed cubic power for grid box weighting during the tracking routine (i.e. when calculating the cluster positions) and for deriving an integral intensity index (the SSI), these powers (`power_track` and `power_SSI`) can be freely configured in the actual revision of WiTRACK. This enables the user to put more or less pronounced weighting on extreme threshold exceedances. It is also possible to neglect the exceedance magnitude by setting the respective power to 0. In that case, only the area of the individual grid box is used for weighting. In principle, even inverted weighting is possible (be specifying negative powers), though generally there should be no reason to do so.

Equally for reasons of transferability, logical switches were implemented to define if relative or absolute exceedances are used for grid box weighting during tracking and calculation of an intensity index (`relexceed_track` and `relexceed_SSI`). These are also set in the configuration file. In this context, the cluster identification was slightly changed in order to be able to deal with relative exceedance weighting for a potential threshold of 0. Such a threshold is a natural choice when applying WiTRACK to other parameters, trying to track any anomalies. The cluster identification of the actual WiTRACK scheme (since r182) analyzes whether the individual grid box values are larger than the given threshold. The difference with the older revisions is, that these were also including grid boxes exactly matching to the threshold. In practice, this change hardly makes a difference for the WiTRACK results.

While the above-mentioned technical developments, the creation of WiTRACK configuration files, and the incorporation of real grid box areas are mainly for reasons of efficiency and better structure with respect to further developments as well as more straight-forward analyses, this flexibilization of additional tracking parameters and the substantial WiTRACK modifications – outlined further below – affect the basic principles of WiTRACK-functionality. Fig. 2.1 illustrates the actual (r211) WiTRACK-

algorithm, containing routines unchanged since the work of Leckebusch et al. (2008a), such as the size and lifetime check, and others slightly (Identification) or fundamentally changed (Tracking) since the initial state of WiTRACK.
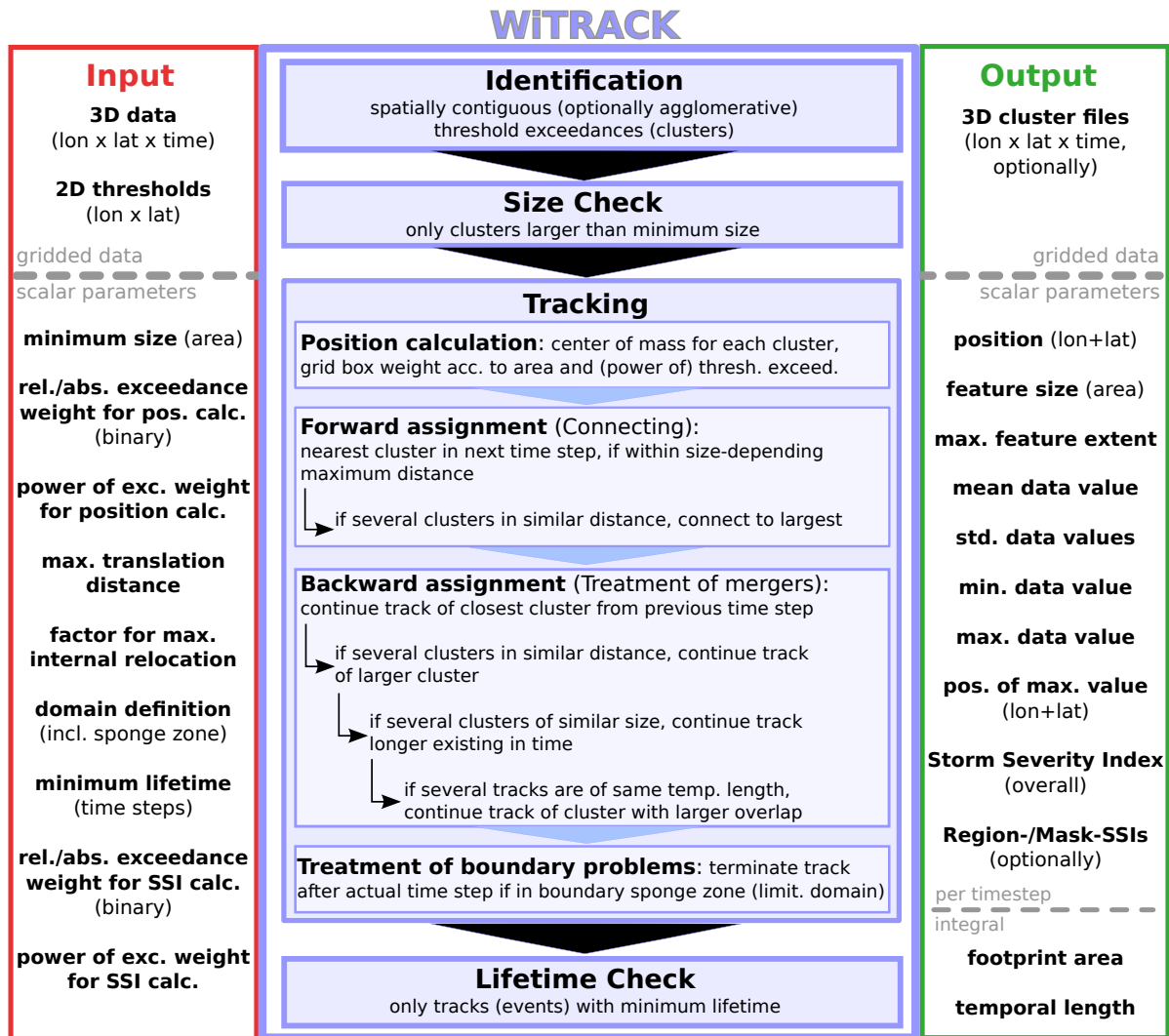
## WiTRACK

**Input**

**3D data**
(lon x lat x time)

**2D thresholds**
(lon x lat)

gridded data
scalar parameters

**minimum size** (area)

**rel./abs. exceedance weight for pos. calc.**
(binary)

**power of exc. weight for position calc.**

**max. translation distance**

**factor for max. internal relocation**

**domain definition**
(incl. sponge zone)

**minimum lifetime**
(time steps)

**rel./abs. exceedance weight for SSI calc.**
(binary)

**power of exc. weight for SSI calc.**

**Identification**
spatially contiguous (optionally agglomerative)
threshold exceedances (clusters)

**Size Check**
only clusters larger than minimum size

**Tracking**

**Position calculation**: center of mass for each cluster, grid box weight acc. to area and (power of) thresh. exceed.

**Forward assignment** (Connecting):
nearest cluster in next time step, if within size-depending maximum distance

if several clusters in similar distance, connect to largest

**Backward assignment** (Treatment of mergers):
continue track of closest cluster from previous time step

if several clusters in similar distance, continue track of larger cluster

if several clusters of similar size, continue track longer existing in time

if several tracks are of same temp. length, continue track of cluster with larger overlap

**Treatment of boundary problems**: terminate track after actual time step if in boundary sponge zone (limit. domain)

**Lifetime Check**
only tracks (events) with minimum lifetime

**Output**

**3D cluster files**
(lon x lat x time,
optionally)

gridded data
scalar parameters

**position** (lon+lat)

**feature size** (area)

**max. feature extent**

**mean data value**

**std. data values**

**min. data value**

**max. data value**

**pos. of max. value**
(lon+lat)

**Storm Severity Index**
(overall)

**Region-/Mask-SSIs**
(optionally)

per timestep
integral

**footprint area**

**temporal length**

Figure 2.1.: Scheme of the WiTRACK-functionality as actual for revision r211, including necessary input and produced output

**Substantial modifications of the scheme**

**Envelope for cluster identification (achievement of Tobias Pardowitz):**   The initial WiTRACK scheme was developed and tested for standard GCM output and reanalysis fields with a spatial resolution of approx. 2° or even coarser. Surface wind data with increased spatial resolution typically exhibits larger variance, especially over land areas. Straightforward application of WiTRACK to output fields of a regional climate model (RCM) lead to disrupted wind clusters, as quite often large areas of wind speeds exceeding the local 98$^{th}$ percentile were intercepted by comparably small areas of slightly lower wind speeds. Facing this issue, the option of defining a secondary threshold was implemented in order to re-conflate these features in the sense of an agglomerative clustering approach. If this option is used, discrete areas exceeding the primary threshold (e.g. the 98$^{th}$ percentile) are treated as belonging together if surrounded by a contiguous area exceeding the secondary threshold (e.g. the 95$^{th}$ percentile). All event characteristics, including the cluster positions, are still based only on those grid boxes exceeding the primary threshold.

**Automatic detection and treatment of poles and column-periodicity (achievement of Tim Kruschke):**   As described in Sec. 2.3.2, the initial WiTRACK scheme was designed for applications on limited domains. Though slightly more challenging regarding the tracking – requiring a special treatment for events traveling in and out of the domain as already described – such applications are comparably easy in terms of cluster identification and position calculation. However, applying WiTRACK to geodata like meteorological fields basically demands for a solution to deal with poles and column-periodicity. These two possibilities are special for cluster identification and position calculation: (i) a pole means that all elements of the first or last matrix row (if a rows are oriented along latitudes) are next to each other as they all border the pole; (ii) column-periodicity means that the left and right margin of a matrix are identical, i.e. the first and the last column are next to each other, which happens to be the case for global fields with columns oriented along longitudes. Both possibilities are now implemented in WiTRACK (development started already in Kruschke, 2010, in final form since r123) with automatic checking of these possibilities and according consequences for cluster identification and position calculation.

**Cluster identification, excluding values matching threshold (achievement of Tim Kruschke):** A very basic modification considering the identification routine was implemented in revision r182. Since then, only those grid boxes really exceeding the given threshold are identified to potentially belong to a feature of interest. So far, grid boxes exactly matching the threshold were also considered. This modification exhibits hardly any effect in practice regarding the WiTRACK results but is essential for the straightforward application of WiTRACK for other variables in a more general context. Such "foreign" WiTRACK-applications are not presented in this thesis but were complementary conducted (by the author of this thesis) for peripheral studies.

All substantial modifications described so far, do affect the feature identification of WiTRACK. The developments to be described below fundamentally affect the tracking routine.

**Storm-size dependent maximum distance between cluster positions in subsequent time steps (achievement of Tim Kruschke):** The cluster positions are calculated twice in WiTRACK: (i) in the tracking routine, and (ii) when calculating the event properties for the WiTRACK output. Unfortunately, the initial WiTRACK-scheme contained an inconsistency between these two calculations, by not weighting the grid boxes with the storm intensity, i.e. the cubic power of relative threshold exceedances, for the tracking. However, as part of the tracking routine, the check for the allowed maximum distance between cluster positions in subsequent time steps is performed. This inconsistency lead to a (small) number of tracks constructed though violating the maximum distance criterion for individual time steps according to the output positions.

As the basic idea of WiTRACK was to perform this weighting for the position calculation, the calculation in the tracking routine was adapted. This adaptation in turn made clear, that the so far used maximum distance of 720 km for extra-tropical winter storms (given 6-hourly data) is not appropriate for all events when cluster positions are calculated including the intensity weighting. Very large winter storms, such as "Vincinette", "Vivian", or "Kyrill" exhibit distances of more than 1000 km (in 6 hours) between some of their track positions. The reason is that both, the translation of the synoptic system, and the internal relocation of the "center of mass" of the storm affect the total distance of the cluster positions. A simple solution (not used in any published study)

was to choose larger values for the maximum allowed distance. However, this is not appropriate for comparably small events.

Hence, a more thorough approach was implemented, accounting for the two above-mentioned factors controlling the maximum allowed total distance (`maxdist`)

$$\texttt{maxdist} = \texttt{max\_translatdist} + \texttt{max\_internrelocfac} \cdot max \left( max\_extent_{t=i,i+1} \right) \quad (2.2)$$

The maximum translation distance of the feature (`max_translatdist`, to be specified as part of the WiTRACK configuration) is complemented by a storm size-dependent addition. For this addition, the maximum extent of the clusters ($max\_extent_t$; since r125 part of WiTRACK output – labeled `RADIUS` – for each time step of the tracked events) considered in time step $t = i$ and $t = i+1$ is internally calculated as the distance between the cluster position ("center of mass") and the most distant grid box, identified to belong the respective cluster. The larger of these two maximum cluster extents is multiplied by a factor `max_internrelocfac` – first introduced in r84 with a fixed value of 0.5, since r156 to be freely specified when configuring WiTRACK. While developing this approach, extensive sensitivity studies have been conducted, especially for the major European winter storm events that are considered in Ch. 4.

For the objective identification and tracking of mid-latitude winter wind storms a maximum translation distance of 600 km (in 6 hours) in combination with a value of 0.5 for `max_internrelocfac` was found to be the best solution. A side effect of this combination is that identifying and tracking mid-latitude winter storms with this parameter combination and the already mentioned minimum size of 150,000 km$^2$ implicitly works with a total allowed distance (`maxdist`) of approx. 720 km for the smallest and most compact (circular) clusters possible. Hence, the actual WiTRACK scheme can not miss any events identified by previous revisions because of an effectively shorter distance allowed between two time steps.

**Renunciation of pure nearest-neighbor-method for tracking (forward and backward assignment; achievement of Tim Kruschke):** The basic principle of WiTRACK in tracking, that is the connection of identified wind fields (clusters) from one time step to another, is a nearest-neighbor-approach. The complexity necessary for a successful

tracking algorithm is determined by the events targeted (basically their translation velocity) and the available temporal resolution. Nearest-neighbor-tracking is probably the most simple approach, performing the better the higher the temporal resolution of the data. The study of Renggli (2011) and the results of Nissen et al. (2014a), Renggli et al. (2011), and Renggli et al. (2014) strongly indicate that this approach is absolutely sufficient for the majority of mid-latitudinal storm events given data in a temporal resolution of at least 6 hours. Facing the fact that other storm types, such as tropical cyclones, Medicanes, or Polar Lows typically do not exhibit higher translation velocities than extra-tropical cyclones, the nearest-neighbor-tracking should be appropriate for these event types, too. However, the performance of WiTRACK is naturally not perfect. Especially, occasionally existing rapid storms sequences (see e.g. Mailier et al., 2006; Vitolo et al., 2009; Pinto et al., 2013) as well as mergers and splitting of a storm field pose challenges for each tracking algorithm. Mergers and splittings of storm fields – identified as exceedances of the $98^{th}$ percentile – happen quite regularly. An example for the former is an occlusion of two fronts, each producing strong surface winds, identified as storm event by WiTRACK. It is impossible to design an objective method, deriving an event track which is in line with (subjective) expert judgment for each of these cases. Similarly a rapid sequence of identified clusters may result into a logical – from the computational perspective – but wrong – from a meteorological perspective – track of a quasi-stationary pseudo-event, tracked by a nearest-neighbor-based algorithm. That happens, if the successor event travels to a position within one time step, closer to the old position of the predecessor event than its new position. In some cases such problems lead to unsatisfactory results of WiTRACK for the respective event(s). One example was the succession of storms in late December 1999, including winter storms "Lothar" and "Martin", which is described as one of the case studies presented in Sec. 2.4.2.

To improve the WiTRACK performance regarding this kind of events, a few simple modifications were introduced to the tracking routine, essentially meaning a renunciation of a pure nearest-neighbor-approach. Instead additional parameters are taken into account in an iterative decision process, also illustrated in Fig. 2.1. As WiTRACK is meant to be an impact-oriented tracking algorithm and one essential parameter for potential storm impact is the size of the system, the actual WiTRACK algorithm additionally considers the storm size during the forward assignment routine (since r125). The current solution is to connect to the closest cluster in the subsequent time step, if no other clusters

are in similar distance. Similarity in this respect is defined as a difference regarding the distance of no more than half of the user-specified maximum translation distance (`max_translatdist`). If there is more than one cluster within such similar distance, the larger cluster is chosen for connecting.

Equally the backward assignment of the tracking routine (treatment of mergers) offered possibilities for optimization. Although based on the nearest-neighbor-principle for forward assignment, the initial WiTRACK-algorithm did not account for the distance of the tracks to be potentially continued. As described in Sec. 2.3.2, only the temporal length of the tracks under consideration was the basis for decision. Hence, the backward assignment was changed (first in r93) to account for the distance, but – similarly to the forward assignment – additionally considering other parameters for very close decisions, that is if more than one track potentially to be continued is found in a similar distance for the previous time step (see Fig. 2.1). Here, similarity was first declared to be a difference of less than 10% regarding these distances (in r93), since r125 it is defined compliant with the forward assignment to be a difference of less than half the user-specified maximum translation distance. For such cases of merging tracks positioned in similar distance for the previous time step, again the size of the storm events under consideration is regarded, though now not being the last authority to decide which track to continue. Instead, for storm events of very similar size (difference of less than 10% ) in the previous time step, the decision (which track to continue) is passed to the original criteria, that is the temporal length followed by the (problematic) overlap criterion (see Sec. 2.3.2).

Several studies already exist, using "new" features of WiTRACK described in this Section. Namely the thesis of Osinski (2014, additionally being the basis for analyses presented in Ch. 4) and the related paper of Osinski et al. (2015) used r105 of WiTRACK. Regarding the described substantial developments, that revision incorporates the envelope for cluster identification (Osinski, 2014, and Osinski et al., 2015, used the climatological 95$^{\text{th}}$ percentile as secondary threshold), the storm-size dependent maximum distance between cluster positions, a (not final) version regarding the automatic detection and treatment of poles and column-periodicity, as well as parts of the developments described in "Renunciation of pure nearest-neighbor-method for tracking". Regarding

the latter, r105 was purely nearest-neighbor-based regarding the forward assignment and primarily distance-oriented regarding the backward assignment, switching to temporal length as secondary criterion for mergers of tracks from similar distance, which was defined in that revision to be a difference of less than 10% . Two studies exist, using revisions of WiTRACK containing all substantial developments made so far. One is the paper of Befort et al. (2014, based on r145), making use of the agglomerative clustering during identification with the 95th percentile as secondary threshold. The other is the study of Kruschke et al. (2015, based on r167; also to be found as Ch. 3 of this thesis), not using the agglomerative clustering approach.

### 2.3.4. Output

The basic output of WiTRACK are ASCII-text files, so-called track tables, containing scalar parameters derived for the events identified, tracked and passing the lifetime check. These files (in the current WiTRACK version) start with one header line giving the already-mentioned revision information, followed by one line containing headers for each column of the tables that build the actual WiTRACK-output and hence the remainder of these ASCII-text files. For each event one table is listed in the output files. Each of these tables consists of a header line for the respective event, giving some integral information, and one line per time step of the event with respective informations.

Optionally, gridded "cluster files" can be output by WiTRACK. These files have the same dimension (spatial and temporal) as the input data (after potential domain and season selection). They contain missing values for all grid boxes not identified to be affected by a storm event. Those grid boxes affected (for the respective time step) contain the event identifier – an integer number assigned to each event and also output in the track tables – though without the heading digits representing a running index specified in the *shell*-script starting WiTRACK within a loop over several seasons (or years).

## 2.4. Application of the WiTRACK-scheme

### 2.4.1. Data, configuration, post-processing, and technical results

WiTRACK is designed to be flexibly configured for the specific needs, that is targeted at certain events and their unique characteristics. Hence, past studies used configurations of WiTRACK tailored to the respective storm types considered. Most of the existing WiTRACK-based studies analyze wind storms related to extra-tropical cyclones (Leckebusch et al., 2008a; Kruschke, 2008, 2010; Renggli, 2011; Renggli et al., 2011; Nissen et al., 2014a; Kruschke et al., 2015; Befort et al., 2014; Osinski, 2014; Osinski et al., 2015; Pardowitz, 2014; Pardowitz et al., 2014) and use very similar settings. However, some studies differ with respect to individual parameters. Nissen et al. (2010, 2014b), analyzing wind storms related to Mediterranean cyclones, used a comparatively low threshold for the minimum size: approx. 45,000 km$^2$ instead of 150,000 km$^2$. Analyses on the applicability of WiTRACK for identifying Polar Lows were done by using an even smaller threshold of approx. 12,000 km$^2$ combined with a very short minimum lifetime of only 6 hours, and Sakuth (2011, with the author of this thesis setting up and conducting the WiTRACK-based event identification for this study) successfully conducted a similar exercise with respect to wind storms resulting from tropical cyclones, choosing a minimum size of approx. 36,000 km$^2$ and a maximum translation distance of 500 km (instead of the 600 km usually taken for extra-tropical systems) within 6 hours.

For this Chapter and the underlying question of applicability and transferability of WiTRACK to different storm types, a different approach is chosen. WiTRACK is applied globally to all-year data of the *ERA-Interim*-reanalysis (Dee et al., 2011) using one uniform configuration. This uniform configuration is a synthesis of the mentioned previous studies, compiled to be a suitable setting for a variety of wind storm events, especially related to tropical as well as extra-tropical cyclones. Tab. 2.1 contains a short overview of all parameters used. Based on 6-hourly instantaneous 10m-wind speeds, the empirical 98th percentile derived from the reference period 1979–2013 (all time steps) is used as identification threshold. A minimum size of 36,000 km$^2$ is required, in line with the threshold found to be suitable for tropical storms. Regarding the maximum translation distance, the more permissive threshold of 600 km for extra-tropical systems was used. Finally, a minimum lifetime of 36 hours is requested. This is a comparably restric-

Table 2.1.: WiTRACK-configuration used for global all-year *ERA-Interim* tracking, further analyzed in Sec. 2.4.2 and 2.4.3

| | |
|---|---|
| period covered: | 03/1979–08/2014, split up in 13-month slices starting each March (last slice 18 months 03/2013–08/2014), overlap post-processed to avoid doubling of events |
| domain: | global |
| boundary sponge zone: | none |
| considered variable: | 6-hourly instantaneous 10m-wind speed |
| threshold: | empirical 98$^{\text{th}}$ percentile of 10m-winds, calculated from reference period 1979–2013 (all time steps) |
| secondary thresh.: | none |
| minimum size: | 36,000 km$^2$ |
| grid box weighting during tracking: | area multiplied by third power of relative threshold exceedances (SSI-contribution, see Eq. 2.1) |
| maximum translation distance: | 600 km |
| factor for max. internal relocation: | 0.5 |
| minimum lifetime: | 7 time steps $\sim$ 36 hours |
| grid box weighting for SSI-calculation: | standard SSI-contribution (see Eq. 2.1, equivalent to tracking) |

tive value, however, the storm events of interest here – extreme, potentially damaging systems of both, tropical and mid-latitudinal origin – are expected to fulfill this requirement. The tracking was started in March 1979 and terminated in August 2014. The reason for this period is to consider all *ERA-Interim*-data available at the time of this study but avoid incomplete seasons (the winter 1978/1979 would be incomplete without December 1978, which is not covered by *ERA-Interim*). The full period of more than 35 years was split up for computational reasons (memory usage). To avoid discontinuities (partial or missing events at slice edges) 13-month slices were tracked, including one-month overlaps, followed by post-processing procedures to remove (partially) doubled events.

Regarding the effects of the new tracking routines, some technical information with respect to this 35-year application of WiTRACK can be summarized:

- total number of $\sim$ 1,965,000 clusters (larger than minimum size) identified
- forward assignment:

  $\sim$ 65.5% connected (to cluster in next time step) because of distance criterion

  $\sim$ 3.0% connected because of size criterion

  for $\sim$ 31.5% no cluster within maximum allowed distance
- backward assignment (treatment of mergers): $\sim$ 214,500 mergers in total

  $\sim$ 59.6% decisions (which to continue) because of distance criterion

  $\sim$ 37.8% decisions because of size criterion

  $\sim$ 1.7% decisions because of temporal length

  $\sim$ 0.9% decisions because of (still problematic) overlap criterion
- total number of $\sim$ 829,000 tracked storm events
- total number of 59,559 output storms, longer than specified minimum lifetime

The synthesis of these numbers (derived before overlap-post-processing, hence, only of technical meaning) is that WiTRACK primarily remained a nearest-neighbor-based tracking algorithm. On average, $\sim$ 140 clusters (larger than the minimum size, set to 36,000 km$^2$) are identified per time step and – if possible – connected to another cluster in the subsequent time step. With a maximum translation distance of 600 km (which is also important for the definition of distance similarity during tracking; see above), for only $\sim$ 4.4% of the cases, a connection within the maximum allowed distance is possible, it is more than one cluster in similar distance, causing the size criterion to be consulted during forward assignment. For about one sixth of the connections made, these links compete with others, that is several clusters are connected to the same feature in the next time step, tracks are merging. For the majority of these cases, a decision which track to continue is also based on the distance, the closer track is continued. The secondary criterion of cluster size is consulted in less than 40% and the criteria used in the initial WiTRACK-algorithm, that is the temporal length of the tracks and the overlap, come into play for less than 2% respectively 1% of such incidents.

The number of tracked events is slightly less than half the number of clusters, which is equivalent to an average temporal track length of only $\sim 2.36$ time steps. As shown in Sec. 2.4.3, generally the number of WiTRACK-derived tracks exponentially decays for longer lifetimes. This is reflected in the total number of output storms, that is events tracked longer than the specified minimum lifetime of 36 hours, which is only $\sim 7.2\%$ of the total number of identified storm events.

## 2.4.2. Case studies

Before analyzing the full 35-year climatology of globally identified storm events in Sec. 2.4.3, two selected case studies are presented to prove the chosen WiTRACK-configuration successful in identifying and tracking different types of events. As the WiTRACK-scheme has been demonstrated to be very useful for NH extra-tropical systems and European winter storms in particular, the first example of winter storms "Lothar" and "Martin" is chosen to explain the impact of some of the new features (see Sec. 2.3.3) implemented in WiTRACK. The sequence of these storms is subject of a number of studies regarding their dynamical aspects (e.g. Wernli et al., 2002), their predictability (e.g. Buizza and Hollingsworth, 2002), or their economic impacts (e.g. Munich RE Group, 2001). Afterwards, the example of Hurricane "Sandy" is shown to prove the general applicability of WiTRACK for tropical storms. This event heavily affected the US East Coast in late October 2012, also stimulating various studies tackling the different aspects (e.g. Galarneau et al., 2013; Hall and Sobel, 2013; Magnusson et al., 2014; Munich RE Group, 2013). Here, none of these event specific questions will be addressed. They are only shown as case studies illustrating the performance of WiTRACK.

### European winter storms "Lothar" and "Martin" (12/1999)

Fig. 2.2 presents snapshots of the European winter storms "Lothar" and "Martin in December 1999. The first plot represents the state of Christmas Eve at 12 UTC, the first time step winter storm "Lothar" has been identified from *ERA-Interim* by WiTRACK (using the configuration described in Sec. 2.4.1) over the North Atlantic, south of Newfoundland. Further east, several wind fields and related tracks are visible for this time step and the following day. This is a classic example of a quasi-stationary large scale cyclone, with several smaller scale cyclones and disturbances traveling along the are of
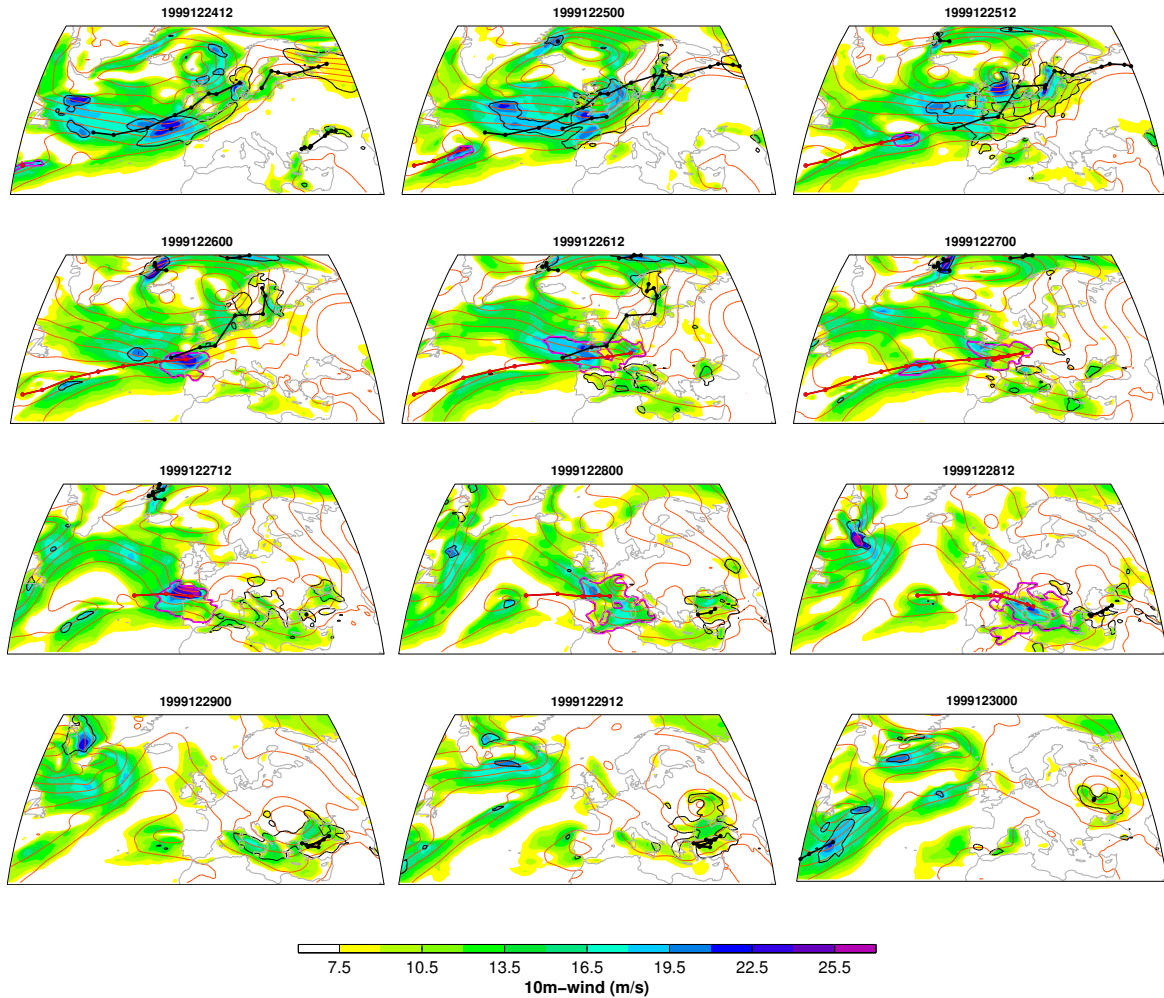
Figure 2.2.: 12-hourly snapshots of European winter storms "Lothar" and "Martin", as identified and tracked from *ERA-Interim* by WiTRACK configured according to Tab. 2.1: 10m-wind speeds (shaded), exceedances of local 98th percentile (black contours, magenta if associated to "Lothar" or "Martin"), and tracks of identified storm events (solid black with position markers, red if associated to "Lothar" or "Martin"); SLP-contours (orange) for additional information

maximum pressure gradients. "Lothar" is one of these disturbances, quickly traveling across the North Atlantic, reaching the French coast on 26 December 00UTC, perfectly tracked by WiTRACK so far. Crossing France, the connection from 00 UTC to 06UTC (not shown, but position marked over central France in all following plots) is possible only by renunciating waiving the nearest-neighbor-scheme in its pure implementation,

as done for the initial WiTRACK-scheme. The closest feature at 06 UTC would have been the follow-up wind field, clearly visible in the plot for 00UTC. Hence, the initial WiTRACK-scheme would have lost "Lothar" here, unfortunately right at the time "Lothar" gains its maximum intensity (06 UTC and 12 UTC exhibiting the highest SSI-values) traveling further across France to Southern Germany (track position at 12 UTC). However, the result of the actual WiTRACK-scheme is not perfect, either. Because of a similar distance to the real "Lothar"-wind-field and its follower, but a larger size of the latter at 18UTC the WiTRACK-yielded track jumps back to the western coast of France. In summary, the rather tolerant definition of similar distance (half the allowed translation distance) and alternatively analyzing cluster size was beneficial for tracking "Lothar" from 00UTC to 06UTC, but somewhat less than optimal for tracking from 12UTC to 18UTC on 26 December. The resulting WiTRACK-yielded "Lothar"-track is terminated at 27 December 1999, 00UTC, over central France.

This is actually the time step, winter storm "Martin" is first identified by WiTRACK, located north of the Azores. Approaching the Bay of Biscay and traveling over Southern France into the Mediterranean, "Martin" affected large parts of France and Spain during December 27th and 28th. Its track is compiled straightforward by WiTRACK until a position over Southern Italy on 28 December 18UTC. At 00UTC December 29th "Martin's" wind field unifies with the feature more or less stationarily tracked over Turkey, resulting into a cluster position close to the Bosporus. As this position is much closer to the previous position of this quasi-stationary event than "Martin's" last position, the track of the latter is terminated.

This short episode is a perfect example for the challenges, such an objective tracking algorithm is facing when applied to extra-tropical wind fields. Nevertheless, the new developments implemented with respect to the tracking routines of WiTRACK and the its configuration used for this study yield separate tracks for the European winter storms "Lothar" and "Martin" that do include all essential (in terms of intensity and potential impacts) wind fields to be associated with the respective events. The ESSI-values, calculated for the two events as tracked in this example are $\sim 27$ for "Lothar" and $\sim 44$ for Martin, respectively.
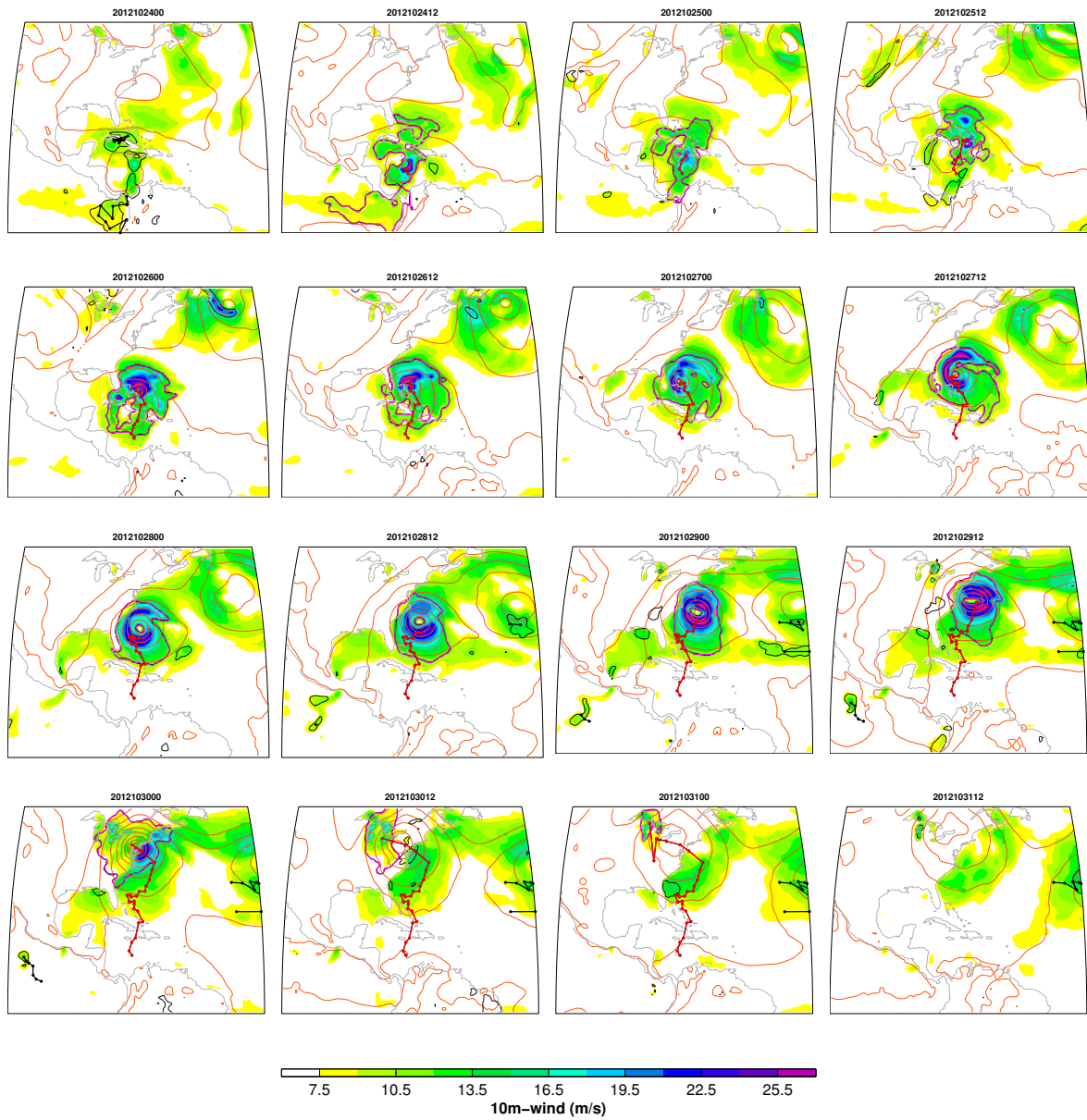
Figure 2.3.: 12-hourly snapshots of Hurricane "Sandy", as identified and tracked from *ERA-Interim* by WiTRACK configured according to Tab. 2.1: 10m-wind speeds (shaded), exceedances of local 98th percentile (black contours, magenta if associated to "Sandy"), and tracks of identified storm events (solid black with position markers, red if associated to "Sandy"); SLP-contours (orange) for additional information

46

**Hurricane "Sandy" (10/2012)**

The second example to be shown in this Section is the WiTRACK-result for Hurricane "Sandy" (see Fig. 2.3). Here, like for many tropical storms (not shown), the challenge for WiTRACK is a proper tracking during the pre-mature phase and decay, respectively. Indications of these challenges are the plots from 24 October 2012 00UTC until October, 25th 12UTC. Several wind wields exist in the vicinity of the cyclone, repeatedly unifying and splitting from time step to time step, while the system traveled from the Caribbean over Cuba to the Sargasso Sea. Overall, WiTRACK does a pretty good job in tracking "Sandy" even for this early phase (comparing to best track data, as provided by Blake et al., 2013, ; also the basis of all further statements regarding development in reality), after first identifying it on 24 October 06UTC, six hours after "Sandy" reached Hurricane intensity. After a two-day weakening phase over the southwestern Sargasso Sea, which is less obvious from *ERA-Interim*-data, from 27 October 00UTC "Sandy" re-intensified, developing a well-organized, almost circular symmetric structure with respect to surface wind-speeds, visible in *ERA-Interim*, too. This mature phase of the Hurricane poses no problem for WiTRACK, given its distinct wind pattern, clearly deviating from environmental conditions, and relatively slow translation velocity, as typical for tropical cyclones. Hence, WiTRACK yields a track in very good agreement to the best track (not shown) until "Sandy" made landfall on 30 October 00UTC. After that, the wind field of "Sandy" is disrupted into several clusters, with WiTRACK favoring the western part because of the larger size (following the WiTRACK-identification as exceedances of the local 98th percentile). At October, 31st 00UTC the WiTRACK-yielded track is terminated, as no wind fields larger than the required minimum size could be identified thereafter, revealing an ESSI of $\sim 309$ for "Sandy".

## 2.4.3. Climatologies

**Global results: Inter-annual variability and seasonality of spatial distributions**

This Section is dedicated to a comprehensive analysis of the results derived from the WiTRACK-application, configured as described above. This means a statistical analysis of the WiTRACK-yielded tracks and related parameters in their totality. This results into an examination of climatologies (and variability) of the identified wind storm events,

which will be divided into several samples, representing different regions, seasons, and (potentially) storm types. Wherever possible, these analyses – based on the WiTRACK-results for the *ERA-Interim*-reanalysis – are related to findings of other studies. Regarding tropical regions, this is done primarily against the recently published study of Schreck et al. (2014) which analyzed two different "best track" archives with respect to tropical cyclones globally. For extra-tropical cyclones and wind storms no such "best track" data is available, primarily because the structure and appearance of extra-tropical cyclones is more variable than those of tropical systems. Hence, there is no unique way to define and identify extra-tropical cyclones. Instead a number of identification procedures exists, each based on specific focus aspects. The *IMILAST* community conducted an inter-comparison experiment for 15 commonly used detection procedures, which predestinates their results (Neu et al., 2013) to be compared to the findings of this study. However, to the author's knowledge, no study is available performing an comparative analysis of extra-tropical cyclone properties for several regions spread over both hemispheres at the same time. Some studies generally compare northern-hemispheric and southern-hemispheric cyclones (e.g. Hodges et al., 2011; Neu et al., 2013), others contrast properties of cyclones for different regions within one hemisphere (e.g. Gulev et al., 2001; Lim and Simmonds, 2007). Thus, various studies will be regarded, putting the WiTRACK-results into the context of current scientific knowledge.

Matching such results of other studies and the current thesis indeed requires to keep in mind, that most of the other studies consider cyclones as meteorological features in their entirety, while this thesis and WiTRACK focus on the potentially damage-prone parts exhibiting extreme wind speeds. For mature tropical cyclones with approximately radial symmetric structure this differentiation will probably lead to less severe differences than for extra-tropical systems with pronounced frontal structures, often associated with the WiTRACK-identified clusters.

As already mentioned in Sec. 2.4.1 a total number of 59,559 storm events was identified and tracked by WiTRACK, based on global data of *ERA-Interim* from March 1979 through August 2014. Post-processing the overlap of the discrete 13-month-slices tracked yields a remainder of 55,210 storms. This implies an average of 1554.4 storms per year. The temporal distribution regarding individual years (Fig. 2.4) reveals a comparably low degree of inter-annual variability ($\sigma = 91.7$ storms per year $\approx 5.8\%$). Interestingly, the four years exhibiting the most storm events are the last four (complete) years analyzed,
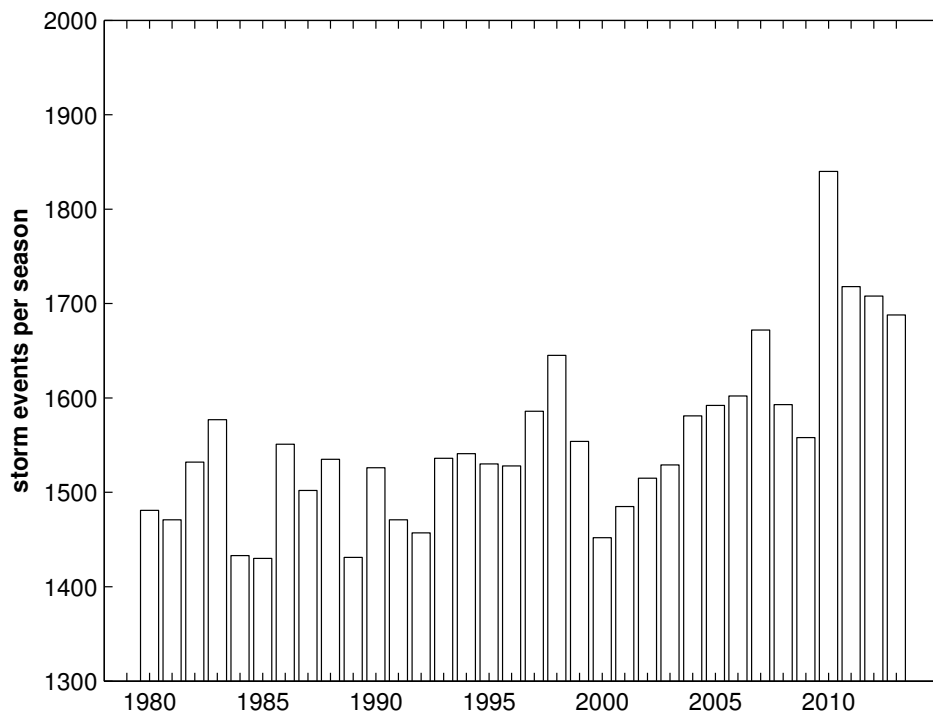
Figure 2.4.: Number of storm events, identified and tracked **globally** by WiTRACK per calendar year (without incomplete years 1979 and 2014)

that is 2010–2013, all of them beyond two standard deviations above the mean. The reason for this phenomenon is unclear, but a physical forcing is assumed, as such a sequence of extreme events is very unlikely to happen by coincidence ($\alpha < 0.001\%$ for standard Gaussian distribution assuming serial independence).

Analyzing the spatial distribution, differentiated between seasons (Fig. 2.5) by means of wind storm track densities (number of events per season within radius of 500 km) shows the well-known patterns of the extra-tropical stormtracks. The highest climatological track density (also to be considered as a frequency/occurrence rate for a given location) can be detected over the central North Pacific in boreal winter (DJF; $\sim 11.5$ storms per DJF within 500 km), followed by the central North Atlantic with $\sim 10.5$ storms per DJF and 500 km-radius. The North Pacific stormtrack is oriented zonally, while the North Atlantic stormtrack exhibits a distinct southwest-northeast tilt.

The Southern Hemisphere (SH) stormtrack is marked by less concentrated wind storm tracks. The maximum track density is found over the Southern Indian Ocean in austral winter (JJA) with $\sim 9.5$ storms per JJA within a 500 km-radius. However, this maximum

(a) MAM

(b) JJA

(c) SON

(d) DJF

0.5  1.5  2.5  3.5  4.5  5.5  6.5  7.5  8.5  9.5  10.5  11.5
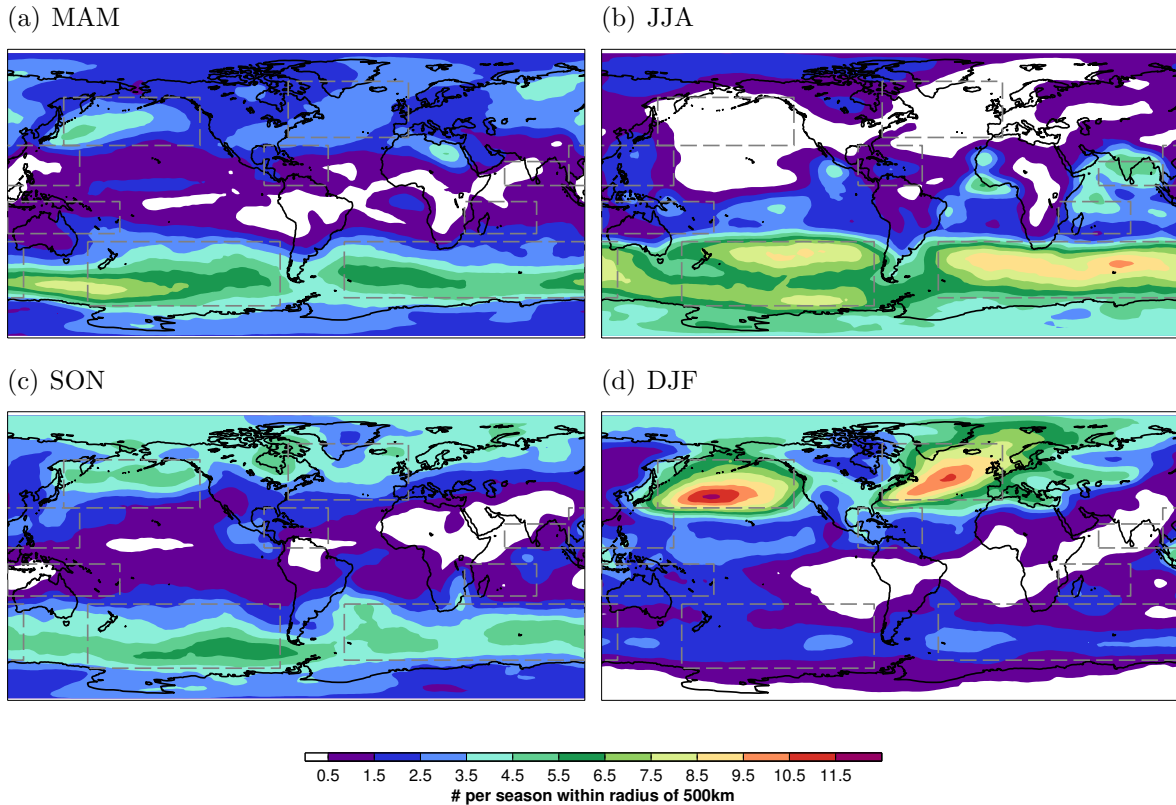**# per season within radius of 500km**

Figure 2.5.: Seasonal wind storm track densities (number of tracks per season within radius of 500 km)

is elongated with values contiguously above 8.5 storms per JJA from the South-West Atlantic to the South-East Indian Ocean. Generally, the SH stormtrack exhibits a spiral-like shape, beginning east of Australia, spinning to the East, shortly intermitted by South America and the Andes Mountains, and subsequently continued over the South Atlantic and Indian Ocean, passing Australia now to the South, and extended to the Amundsen Sea.

Overall these patterns are in very good agreement with results of many extra-tropical cyclone identification algorithms, as presented e.g. by Neu et al. (2013). In particular, the maximum track density over the North Pacific in boreal winter being more pronounced than the North Atlantic signal is a feature also shown e.g. by Hodges et al. (2011, the underlying algorithm not part of the *IMILAST* inter-comparison). This study of Hodges et al. (2011) but also Lim and Simmonds (2007) do provide similarly agreeing results with respect to the austral winter stormtrack, showing a maximum SH track density

over the Southern Indian Ocean and generally higher track densities over the Atlantic and Indian Ocean than over the Southern Pacific.

Several secondary maxima are visible in the tropics, rather in JJA than in DJF. For the NH, many of these features are in line with well-known centers of tropical cyclone activity (see e.g. Schreck et al., 2014). For the SH, such relations are less obvious. Partly, this is due to the global perspective, with the distinct extra-tropical maxima masking less pronounced signals in the tropics. Additionally, yet no WiTRACK-based studies have been conducted for wind storms related to SH tropical cyclones. Maybe, the uniform WiTRACK-configuration chosen for this Chapter is not appropriate for these events. This question will be taken up again later, when considering seasonal variability and properties of the identified wind storm events in the Southern Hemisphere tropics.

Regarding the transition seasons (MAM and SON), it is notable that track densities along the extra-tropical stormtracks for both hemispheres are higher for the respective autumn than for spring. Over the NH, the track density maxima for SON are located further north and downstream than those for MAM.

Complementing these findings regarding wind storm track densities are analyses with respect to genesis (Fig. 2.6) and lysis (Fig. 2.7) of these events. The highest climatological wind storm genesis density globally (>3 storms per season and 500 km-radius) is found around the southern tip of Greenland during boreal winter. A similarly high storm genesis density is located at the Gulf Stream and along the climatological winter sea-ice edge of the North-West Atlantic between Iceland and Spitsbergen. A comparison of the extra-tropical North Atlantic and North Pacific shows, that the former exhibits higher geneses densities over a larger area than the latter (here, the maximum genesis density is only $\sim 2.5$ storms per DJF and 500 km-radius), although the North Pacific is characterized by higher track densities. The conclusion of this apparent contradiction is that North Atlantic wind storms on average cover smaller areas, that means traveling shorter distances because of shorter lifetimes and/or slower translation velocities. This issue will be taken up again as well, when comparing storm properties later.

Comparing these results to the already cited study of Hodges et al. (2011) yields some differences: also based on *ERA-Interim*, this study and the underlying cyclone identification algorithm detect areas over continental North America and Asia (lee of Rockies and Tibetan Plateau) as primary cyclogenesis locations and the Gulf Stream and Kuroshio

(a) MAM

(b) JJA

(c) SON

(d) DJF



0.25  0.5  0.75  1  1.25  1.5  1.75  2  2.25  2.5  2.75  3
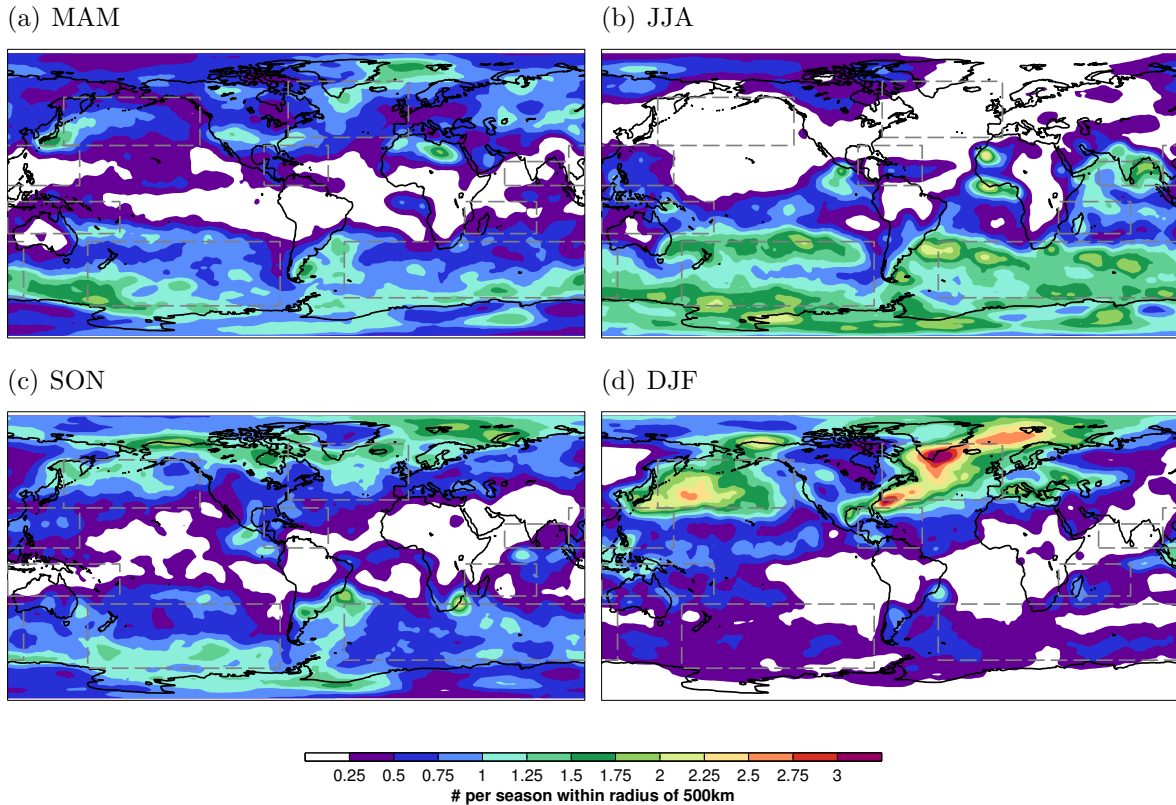**# per season within radius of 500km**

Figure 2.6.: Seasonal wind storm geneses densities (number of track geneses per season within radius of 500 km)

regions as secondary maxima. Besides the essential point, that cyclones (considered by Hodges et al., 2011) are typically not related to intense wind storms (considered by WiTRACK and this thesis) at the time of their genesis but to a later stage, it should be pointed out that other cyclone detection algorithms also yield climatological genesis regions further downstream (mentioned e.g. by Kruschke et al., 2014).

Wind storm geneses in the SH (in JJA) again are more spread than their NH counterparts. A zonal band of local maximum genesis density is located along -40°N, with highest values (between 2 and 2.25 storm geneses per JJA and 500 km-radius) in the lee of the Andes Mountains off the coast of Uruguay. Additionally, many wind storms emerge along the coast of Antarctica, especially at the coast of Marie Byrd Land towards the Ross Sea. Both features are in very good agreement with the cyclone-focused studies of Hoskins and Hodges (2005), Mendes et al. (2010), and Hodges et al. (2011).

In the tropics, typical genesis-regions for NH tropical cyclones (again see e.g. Schreck et al., 2014) stand out in JJA: (i) the Bay of Bengal and the Arabian Sea, (ii) most prominently two hot-spots over the Gulf of Guinea and Western Sahara, both related to the African Easterly Jet, generating tropical waves, and (iii) the tropical North-West Pacific south of Acapulco. Regarding the tropical North Atlantic, a weaker local maximum can be detected over the central to western part of the basin, towards the Caribbean.

The transition seasons offer few additional interesting features. One exception is the eastern coast of South-Africa, a local maximum of wind storm genesis from MAM through SON, but exhibiting the highest intensity with 2.25 storms per season within a 500 km-radius in SON.

(a) MAM

(b) JJA

(c) SON

(d) DJF



0.25  0.5  0.75  1  1.25  1.5  1.75  2  2.25  2.5  2.75  3
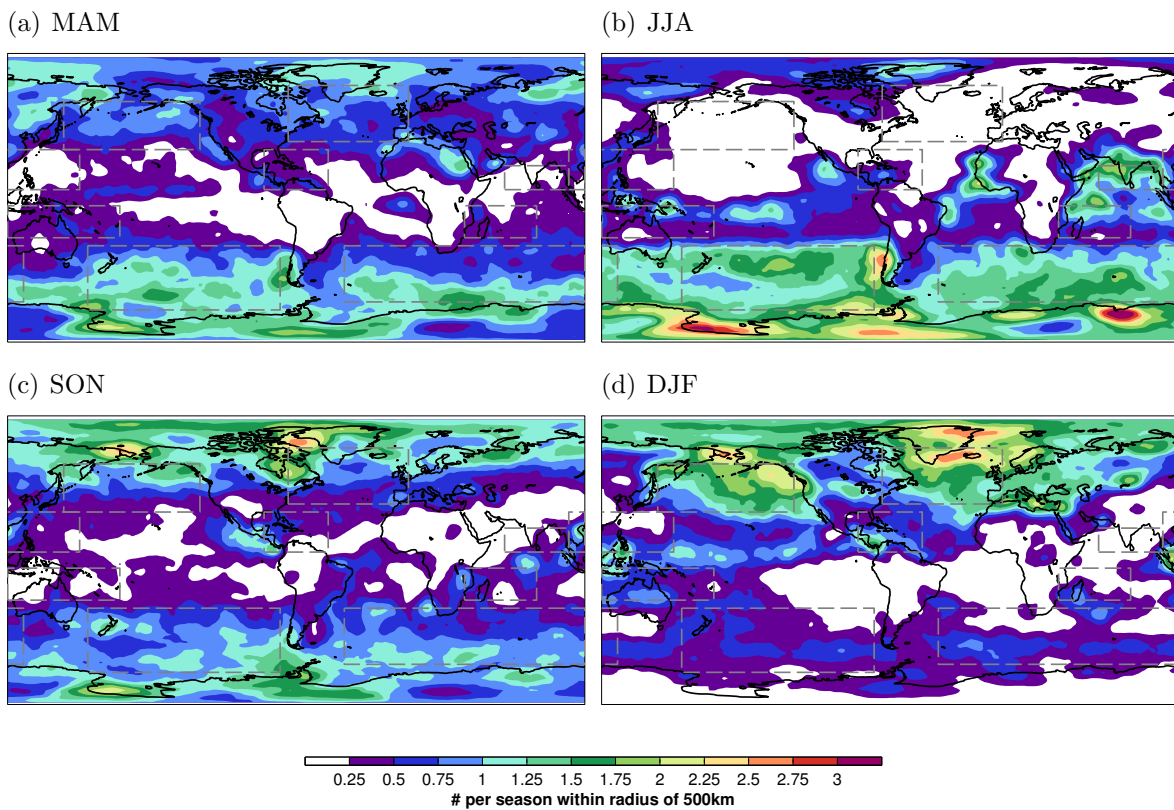**# per season within radius of 500km**

Figure 2.7.: Seasonal wind storm lyses densities (number of track lyses per season within radius of 500 km)

Analyzing the spatial distribution of climatological wind storm lysis density (Fig. 2.7) yields the SH during JJA exhibiting global maxima. Both, the Amery Ice Shelf and the

Ross Ice Shelf show mean lyses densities of >3 storms per JJA within a radius of 500 km. Two other local maxima (both $\sim$ 2.5 storm lyses per JJA and 500km-radius) are located over West Antarctica and off the Chilean coast. The latter feature is related to the local genesis maximum in the lee of the Andes Mountains, with re-emerging storm events dissipated upstream in the luv region. In the tropics, climatological lysis maxima generally can be detected slightly downstream the genesis maxima.

Spatial maps of climatologically typical cyclone or storm lyses regions are less often presented than geneses or track densities. One of the few exceptions is the study of Hoskins and Hodges (2005) for the SH. Based on cyclones objectively detected from relative vorticity in 850 hPa, they show the luv of the Andes Mountains, an area west of the Antarctic Peninsula, and another region north of Antarctic Wilkes Land to be primary cyclolysis regions. This is closely matching Fig. 2.7(d) with respect to the maximum west of South America, but less obvious for the other regions.

In the NH, highest lyses densities in DJF (>2.5 storm lyses per season and 500km-radius) are located between the southern tip of Greenland and Iceland, as well as at the northwestern edge of Greenland. The Bering Strait and the Gulf of Alaska are "the graveyards" (Mesquita et al., 2010) of North Pacific winter wind storms. Overall, these lysis-related results agree reasonably well with the findings of Hodges (1996) regarding climatological cyclolysis.

**Differences between regions and storm types**

Based on the spatial analyses just shown, several regions of interest are defined for further examination of the wind storms identified therein. All of these regions are marked by gray dashed lines in Figs. 2.5, 2.6, and 2.7. Four regions are settled in the extra-tropics (two in each hemisphere), five in the tropics (three in NH, two in SH). The exact definitions of these regions and their abbreviations used in the following, are listed in Tab. 2.2. In order to analyze the specific characteristics of wind storms in these regions, representative samples were compiled by selecting storm events, identified for at least five time steps, that is 24 hours, in the respective region.

**Seasonal and inter-annual variability in the extra-tropics:**   Regarding the climatological seasonal cycle of extra-tropical NH wind storms, the expected result is shown

Table 2.2.: Region definitions for analyses regarding specific wind storm properties

| Abbrev. | Description | Spatial definition |
|---|---|---|
| NA | extra-tropical North Atlantic | 75°W–0°E; 35°N–70°N |
| NP | extra-tropical North Pacific | 145°E–130°W; 30°N–60°N |
| SP | extra-tropical South Pacific | 160°E–80°W; 30°S–70°S |
| SAI | extra-trop. South Atlantic & Indian Ocean | 40°W–120°E; 30°S–65°S |
| tropNA | tropical North Atlantic | 90°W–50°W; 5°N–30°N |
| tropNWP | tropical North-West Pacific | 100°E–155°E; 5°N–30°N |
| tropNI | tropical North Indian Ocean | 60°E–100°E; 5°N–20°N |
| tropSPI | tropical South Pacific & Indian Ocean | 110°E–180°E; 5°S–25°S |
| tropSI | tropical South-West Indian Ocean | 35°E–80°E; 5°S–25°S |

(Fig. 2.8 (a) and (c)), featuring the majority of storm events in the extended boreal winter season. Out of a total number of 3026 (3278) wind storms identified for at least 24 hours in the NA-region (NP-region), 2619 (2892) are to be found in ONDJFM, which is equivalent to 86.5% (88.2%). In the summer months June, July, and August less than one storm per month is detected climatologically for each of the two regions, while January exhibits an approx. 20-fold average value for both regions. This is generally in line with other studies, based on cyclone-detection algorithms (e.g. Hodges et al., 2011). However, this extremity of the seasonal cycle is not detected for the total sample of cyclones, regardless their intensity. This is a specific characteristic of very intense systems, as identified by WiTRACK (also presented by Renggli et al., 2011; Renggli, 2011, for the North Atlantic region and NDJFMA), but potentially evident for subsets of very intense cyclones, too. Based on this finding, ONDJFM is defined as the core season for Northern Hemisphere winter wind storms. Both basins exhibit a very similar shape of a slightly asymmetric seasonal cycle with almost as many storm in March as in February and a distinct drop in April. Examining the number of storm events during the core season (Fig. 2.8 (b) and (d)) reveals that the relative inter-annual variability is also comparable for the NA- ($\sigma_n/\overline{n}$ =21.0%) and the NP-region (19.6%). The absolute numbers are of limited meaning, as they are strongly dependent on the size of the defined regions. No obvious indication of any trend over the analyzed period is detectable, which confirms the study of Tilinina et al. (2013) but disagrees to the negative trend found by (Neu et al., 2013), both considering NH deep cyclones. However, Tilinina et al. (2013) detect the existence of pronounced multi-annual to decadal variability for intense cyclones over both, the North Atlantic and the North Pacific basin. Such low-frequency variability is
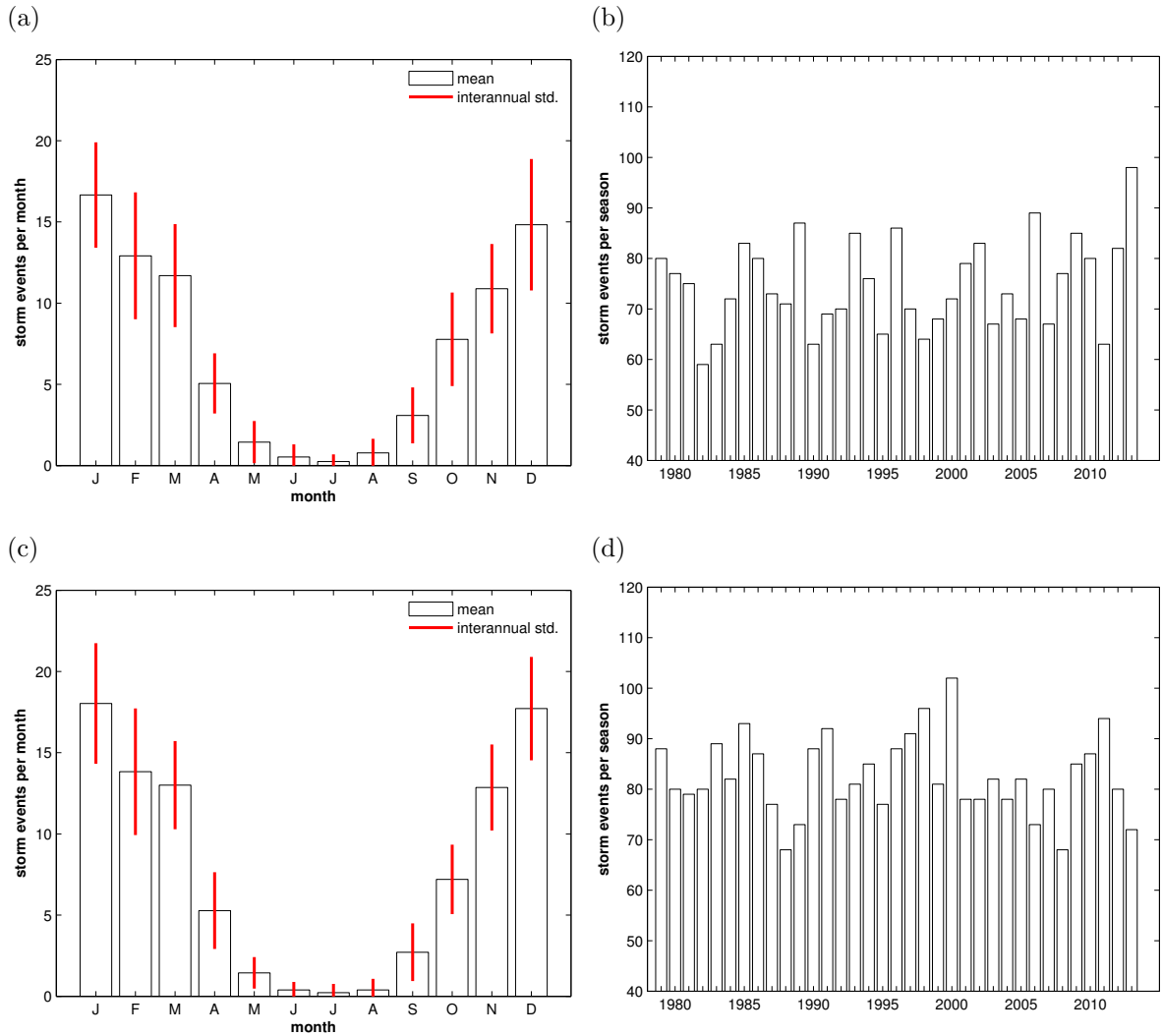
Figure 2.8.: Number of storm events identified and tracked (for at least 24h) within (a,b) **NA-region** and (c,d) **NP-region**: (a,c) climatological mean (and standard deviation) per month; (b,d) total number per ONDJFM-season (winters labeled according to starting year, i.e. 1980 for winter 1980/1981)

not evident from the here-presented WiTRACK-results.

Consistent analyses for the SH extra-tropics yield analog findings. The climatological seasonal cycle regarding the number of wind storms in the SP- and SAI-region (Fig. 2.9 (a) and (c)) exhibits that most storms are detected during the extended austral winter season (AMJJAS). However, the contrast with the summer months is not as striking as in the NH. Based on a total number of 6475 (6997) wind storms identified for at least
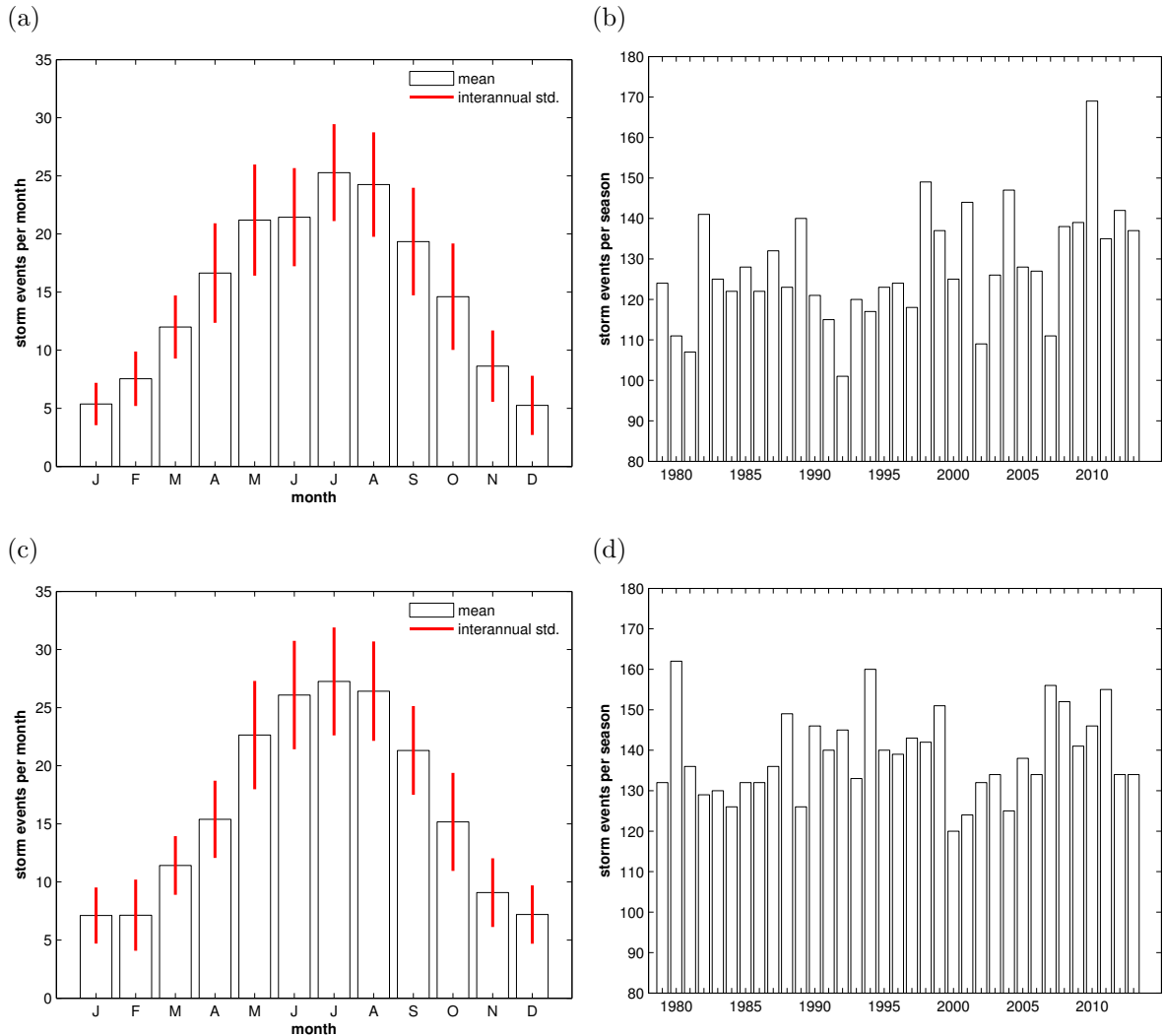
Figure 2.9.: Number of storm events identified and tracked (for at least 24h) within (a,b) **SP-region** and (c,d) **SAI-region**: (a,c) climatological mean (and standard deviation) per month; (b,d) total number per AMJJAS-season

24 hours within the SP-region (SAI-region), 4594 (4986) of these events were identified during AMJJAS, meaning a fraction of 70.9% (71.3%). The minimum monthly means of December, January, and February for both regions are still about 20–25% of the climatological peak monthly number, found in July. Generally, this is again in agreement with cyclone-identifying studies, such as Hodges et al. (2011), though the described subtle difference to the NH regarding a less extreme seasonal cycle is not known from this or similar studies. Inter-annual variability of the number of storm events during

the core season AMJJAS is comparable for the two Southern Hemisphere extra-tropical regions ($\sigma_n/\overline{n}$ =10.8% for SP, 7.6% for SAI). Apparently, relative inter-annual variability is less pronounced compared to the NH extra-tropics. However, an influence of region size (higher degree of intra-regional variability for larger regions) is assumed in that respect, though not quantitatively analyzed. Again, no trend is indicated by these results, which is in line with the results of Neu et al. (2013), regarding deep extra-tropical cyclones.
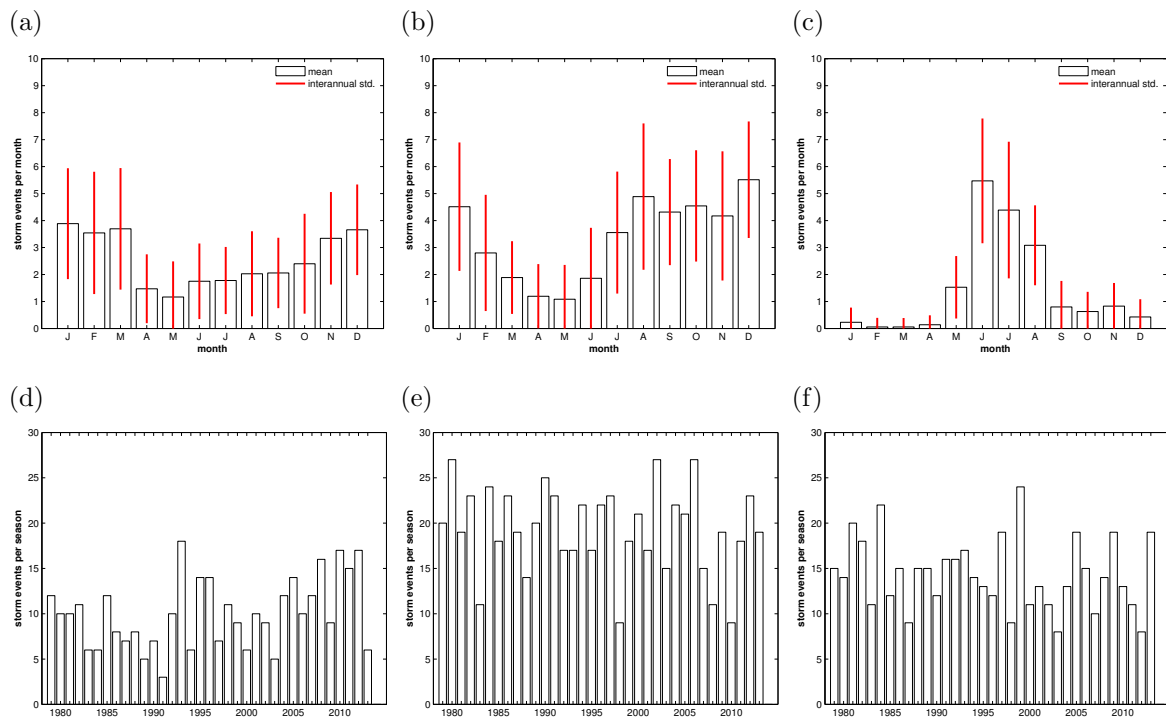


Figure 2.10.: Number of storm events identified and tracked (for at least 24h) within (a,d) **tropNA-region**, (b,e) **tropNWP-region**, and (c,f) **tropNI-region**: (a,b,c) climatological mean (and standard deviation) per month; (d,e,f) total number per JJASO-season

**Seasonal and inter-annual variability in the tropics:** As expected, the climatological seasonal cycle of NH tropical wind storms looks very different (Fig. 2.10 (a), (b), and (c)). Most unambiguous are the results with respect to the tropNI-region with almost no storms from January through April, but significantly more events during boreal summer abating towards autumn. Having defined those regions targeted at wind storms related to tropical cyclones, the climatological seasonal cycle of the tropNWP-region and especially

the tropNA-region exhibit disappointing results, derived by the described purely spatial selection procedure. Facing analyses of actual tropical cyclone seasonality (e.g. Schreck et al., 2014), a clear maximum in late summer to autumn was expected. On the other hand, both regions, tropNWP and tropNA, are comparably close to the extra-tropical stormtracks (see e.g. Fig. 2.5), in fact flanking preferred areas of winter wind storm genesis (see Fig. 2.6(d)). Hence, it has to be discerned that the seasonal cycle of these regions is influenced by these winter storm events, exhibiting the maximum number of storm events in January for the tropNA-region and December for the tropNWP-region, respectively. However, for the latter a secondary maximum of similar magnitude is detected in August. Probably, and also indicated by Figs. 2.5(b) and (c) and 2.6(b) and (c), the summer wind storms are to be found in the core of the tropNWP-region, while in autumn genesis and track density is progressively shifted to the northern edge of the defined region. Figs. 2.5(b) and (c) and 2.6(b) and (c) in principle show the same development for the tropNA-region but the climatological seasonal cycle calculating event numbers for discrete months (Fig. 2.10(a)) does not resolve a secondary maximum during summer or early autumn. Nevertheless, we hypothesize that the WiTRACK-identified summer and early autumn wind storms in all three tropical NH regions are related to tropical cyclones and define a common season for these events and the three regions, being JJASO. With respect to tropNI, this hypothesis is hard to justify when taking into account monthly counts of "best track" data from the Northern Indian Ocean (Schreck et al., 2014), featuring a bimodal seasonal cycle with maxima from April to June as well as October through December. However, in that case the WiTRACK results will be consequently followed, further analyzing Northern Indian Ocean summer/autumn wind storms.

Compiling subsamples from these months only yields 356, 681, and 516 storms for the tropNA-, tropNWP, and tropNI-region, respectively. These are equivalent to 32.7%, 47.8%, and 81.6% of the total number of storm events identified in these regions. Analyzing these events regarding their inter-annual distributions shows a comparably high degree of inter-annual variability, largest for the tropNA-region ($\sigma_n / \overline{n}$ =38.4%) and similar for the tropNWP- (24.7%) and tropNI-region (27.4%). For tropNA and tropNWP, these numbers are in line with Schreck et al. (2014), counting $\sim$ 12 and $\sim$ 25 tropical cyclones (with at least category 1 intensity according to Saffir-Simpson-scale) for the respective regions (though exact region definitions not identical), whereas the numbers for

the tropNI-region are much larger than expected from an average of $\sim 4$ actual tropical cyclones per year (Schreck et al., 2014).
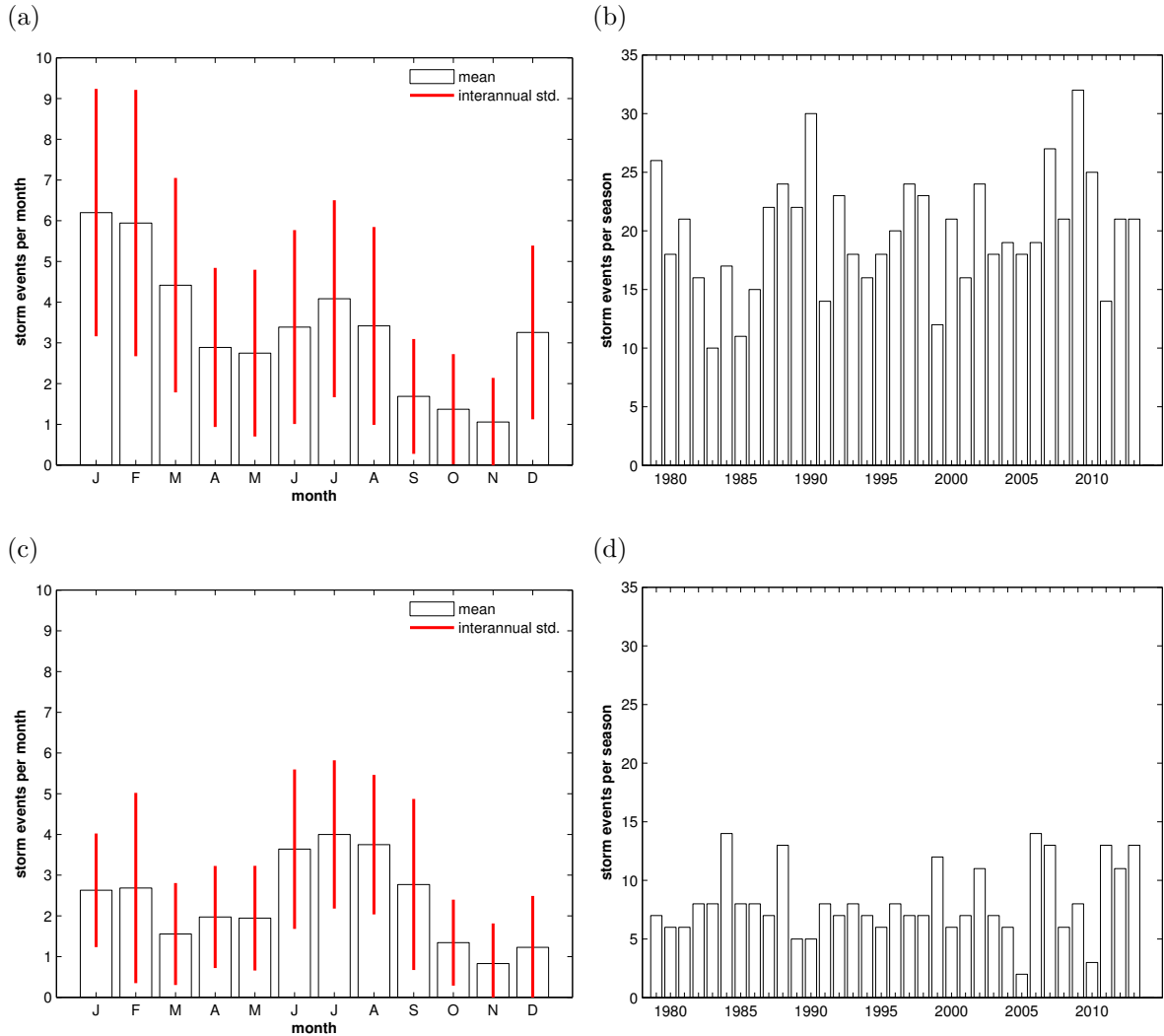
(a)

(b)

(c)

(d)

Figure 2.11.: Number of storm events identified and tracked (for at least 24h) within (a,b) **tropSPI-region** and (c,d) **tropSI-region**: (a,c) climatological mean (and standard deviation) per month; (b,d) total number per DJFM-season (labeled according to starting year, i.e. 1980 for 1980/1981)

The two regions settled in the SH tropics exhibit clearly bimodal climatological seasonal cycles regarding the number of wind storms identified therein (Fig. 2.11(a) and (c)). The primary peak of the tropSPI-region is found in austral summer with a secondary maximum in winter, while the tropSI-region shows converse behavior. Again, it is assumed

that these distributions originate from interfering counts of wind storms related to tropical and extra-tropical cyclones. In agreement with other studies (Dare and Davidson, 2004; Ramsay et al., 2012; Schreck et al., 2014) DJFM was defined to be the core season of wind storms potentially related to tropical cyclones for both Southern Hemisphere regions and further analyses.

In total, 1437 and 1009 wind storms were identified in the tropSPI- and tropSI-region, respectively. Confining to DJFM yields 696 (48.4%) and 285 (28.2%) remaining events. Once again, no trend is obvious from the annual numbers (Fig. 2.11(b) and (d)) and inter-annual variability is quite large ($\sigma_n/\overline{n}$=30.8% and 41.7%, respectively), though again potentially related to region size and sampling error.

**Properties of different storm types:**  Based on the very simple selection criteria described, that is a spatial confinement, selecting only storms identified for at least 24 hours within pre-defined regions, and a restriction to certain region-specific seasons, samples of (potentially) different wind storm types were compiled. These samples are now analyzed with respect their distributions regarding different physical characteristics. The statistical basis (the number of selected storm events) is very different for the considered regions:

- **NA**: 2619 North Atlantic wind storms in ONDJFM

- **NP**: 2892 North Pacific wind storms in ONDJFM

- **SP**: 4594 South Pacific wind storms in AMJJAS

- **SAI**: 4986 South Atlantic & Indian Ocean wind storms in AMJJAS

- **tropNA**: 356 North-West Atlantic wind storms in JJASO

- **tropNWP**: 681 North-West Pacific wind storms in JJASO

- **tropNI**: 516 North Indian Ocean wind storms in JJASO

- **tropSPI**: 696 South-West Pacific & Indian Ocean wind storms in DJFM

- **tropSI**: 285 South-West Indian Ocean wind storms in DJFM

It should be noted, that the full (tracked) life cycle of each storm event is considered for the respective region, even if only a (sufficiently long enough) part of its track was identified within the region. The hypothesis is that the majority of the wind storms

selected for the "trop"-regions is related to tropical depressions and cyclones, while those for the other regions are related to mid-latitudinal (winter) cyclones. The associated basic expectation is significantly different characteristics for tropical and extra-tropical wind storm samples.
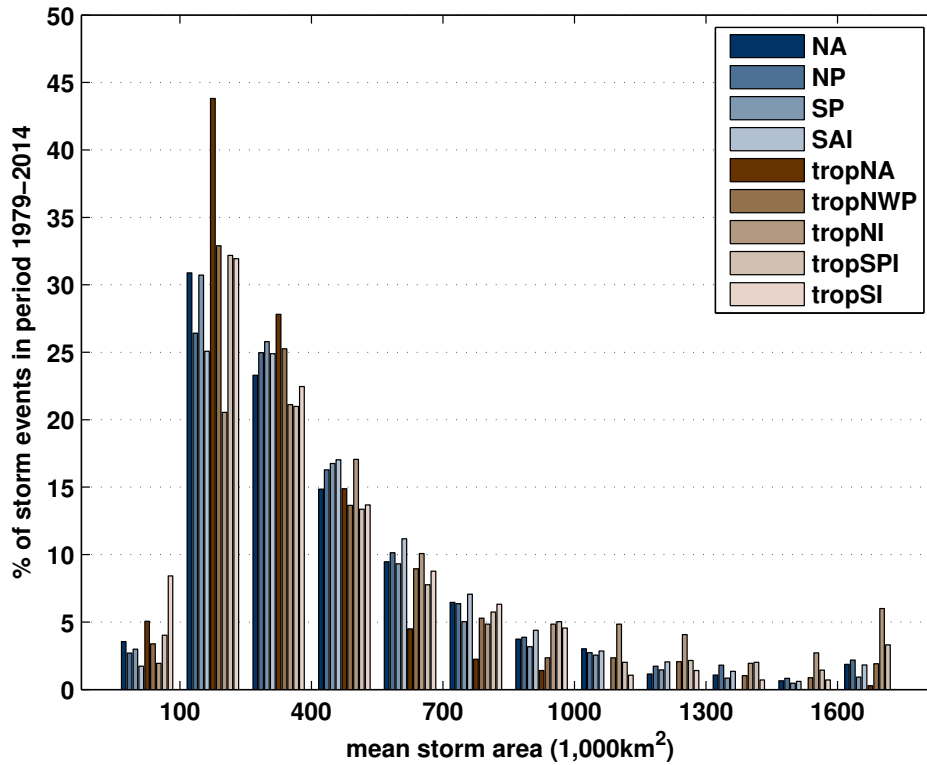


Figure 2.12.: Relative distribution (histogram) of mean (averaged over lifetime) storm size (area) for samples of different storm types, compiled by geographic and seasonal selection

The first storm property compared for the different regions (and seasons) is the mean wind storm area (averaged over lifetime of each event; Fig. 2.12). All wind storm samples exhibit modal values between 100,000 km$^2$ and 400,000 km$^2$. Assuming circular shape, these values are equivalent to a diameter of approx. 350–700 km. It should be noted, that this assumption is not realistic for most extra-tropical storms but made to foster the reader's feeling for the dimensions expressed in Fig. 2.12. Comparing the histograms over all classes, most wind storm samples exhibit rather similar distributions. Only two samples show very obvious differences: (i) the tropNA-storms-distribution is significantly shifted towards smaller mean storm areas with on the one hand about 50% of the events

featuring a mean storm area of less than 250,000 km$^2$, a comparably small fraction of events in intermediate classes (550,000–1,000,000 km$^2$) and on the other hand only one single event ($\sim$ 0.3%) identified with a mean size of more than 1,000,000 km$^2$; (ii) the tropNI-distribution reveals opposite results with only about 22.5% of the events below a mean area of 250,000 km$^2$ and 19.6% above 1,000,000 km$^2$, constituting a shift to larger areas even when compared to the extra-tropical wind storms. While the tropNA-result is compliant with the hypothesis of those events being related to tropical cyclones (and their typical size), this seems not to be the case for tropNI. A recent study by Knaff et al. (2014) analyzing tropical cyclones sizes from satellite data showed that such systems over the Northwest Pacific are significantly larger than in other basins. This result cannot be confirmed by the findings of this study. However, the fact that area distributions of most tropical wind storm samples are similar to extra-tropical wind storms indicates that the used WiTRACK-configuration might not be optimal for the former. This issue will be addressed in Sec. 2.5. Several studies analyze the size of extra-tropical cyclones in different regions: Lim and Simmonds (2007) show that the radius of cyclones in the high-latitude SAI-region is typically larger than over the SP-region, which is supported by Fig. 2.12 with the SAI-storms exhibiting larger fractions than SP for all classes beyond a mean area of 400,000 km$^2$, although the majority of these differences are only subtle. Rudeva and Gulev (2007) and Schneidereit et al. (2010) state that the mean (effective) radius of (deep) North Pacific extra-tropical cyclones is (slightly) larger than over the North Atlantic. Fig. 2.12 matches this finding, too. However, it is important in that respect to note the radius of a cyclone is not necessarily related to the (mean) area of the wind storm field, produced by the respective system. The cyclone radius does not include any direct information about cyclone intensity or shape of the pressure field, such as gradients or curvature, while such characteristics are those more strictly related to wind speeds. Hence, these accordances are not naturally to expect.

Regarding the (tracked) lifetime of the different wind storm samples (Fig. 2.13), more systematic differences are visible, especially when comparing tropical and extra-tropical storms. Regarding short lifetimes of less than 72h, the tropical storms – except for tropNI – show a remarkably smaller fraction, while they exhibit much more events (relatively) with lifetimes of more than 120h (5 days; again not valid for tropNI). Some differences can be detected within the groups of tropical and extra-tropical storms, respectively. North-West Atlantic tropical wind storms (tropNA) slightly tend to be
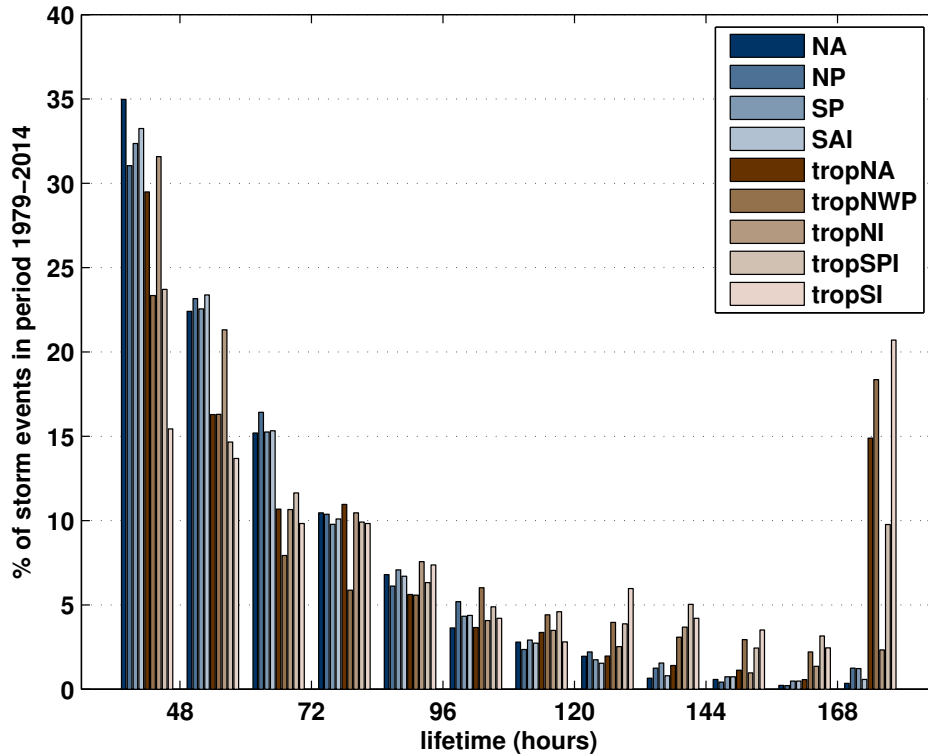
Figure 2.13.: Relative distribution (histogram) of lifetime for samples of different storm types, compiled by geographic and seasonal selection

shorter in time (56.5% shorter than 72h, 19.9% longer than 120h) than their tropical counterparts in other regions considered (neglecting tropNI). The largest fraction of events longer than 120h is associated to tropSI-storms (36.8%) and tropNWP-storms (30.5%). Rough estimations based on the plots presented by Webster et al. (2005) yield Northwest Pacific tropical cyclones to be the longest ($\sim$ 14 days) and systems over the Northern Indian Ocean to be the shortest ($\sim$ 8 days), while all other basins exhibit lifetimes of approx. 11 days. Dare and Davidson (2004) analyzed tropical cyclones north of Australia (approximately matching the tropSPI-region), finding an average lifetime of $\sim$ 7.5 days. Hence, the results presented here do not confirm these findings, but cannot be seen as serious counterproofs either. More detailed analyses targeted at such a question would be necessary, once WiTRACK and the used configuration is found to be appropriate for analyzing tropical storms. For the extra-tropical wind storms, such differences are not as easy to detect as for the tropics. The extra-tropical North Atlantic (NA) exhibits the largest fraction of short wind storms (72.6%<72h) and the smallest

fraction of very long lifetimes (3.7%>120h), while the lifetime distributions of wind storms over the Pacific basin are slightly shifted towards longer lifetimes (70.6%<72h and 5.3%>120h for NP; 70.2%<72h and 5.7%>120h for SP). However, these differences are marginally significant and hence, not sufficient to explain the apparent contradiction of wide-spread higher storm geneses densities but lower track densities over the North Atlantic compared to the North Pacific during boreal winter (Figs. 2.5(d) and 2.6(d)). The overall shape of the lifetime distribution with respect to extra-tropical systems is in line with Tilinina et al. (2013) and Neu et al. (2013), the latter study not able to prove any significant differences between Northern and SH, either. Gulev et al. (2001) analyzed several sub-regions of the North Pacific and North Atlantic regarding the lifetime of extra-tropical cyclones. While finding remarkable differences between these sub-regions, no obvious distinctions for the basins as a whole are visible from their study, either. Mendes et al. (2010) compared the lifetime of extra-tropical cyclones within the South Atlantic basin and the totality of SH systems, finding Atlantic cyclones tending to be shorter. However, the domain definition in their study is hardly compatible with the mutual treatment of South Atlantic and Indian Ocean systems, performed for this thesis. The tropNI-storms possess a lifetime distribution – as for mean wind storm area – which is more similar to extra-tropical wind storms than to tropical events, as diagnosed from the here-studied samples.

The third wind storm property to be compared is mean translation velocity. To derive this parameter, distances between track positions were calculated, divided by the 6-hourly time step, and finally averaged over lifetime for each event. Hence, it should be noted that a high translation velocity is not necessarily related to a long overall distance traveled. Besides the lifetime of a storm event affecting this overall distance, in principle there is also the possibility of "wobbling" events exhibiting comparably high translation velocities although quasi-stationary over lifetime. Analyzing these translation velocities (Fig. 2.14) yields the most consistent distinction between tropical and extra-tropical wind storms but also within these groups, yet. Very clearly do tropical wind storms (neglecting tropNI, which is again more similar to extra-tropical storms) exhibit slower translation velocities than extra-tropical systems. 90.4%, 95.3%, 92.4%, and 97.5% of the tropNA-, tropNWP-, tropSPI-, and tropSI-storms, respectively, travel by mean velocities of less than 50 km/h. For all extra-tropical storm types, the modal value can be detected beyond this threshold. Wind storms in the tropSI-region are considerably
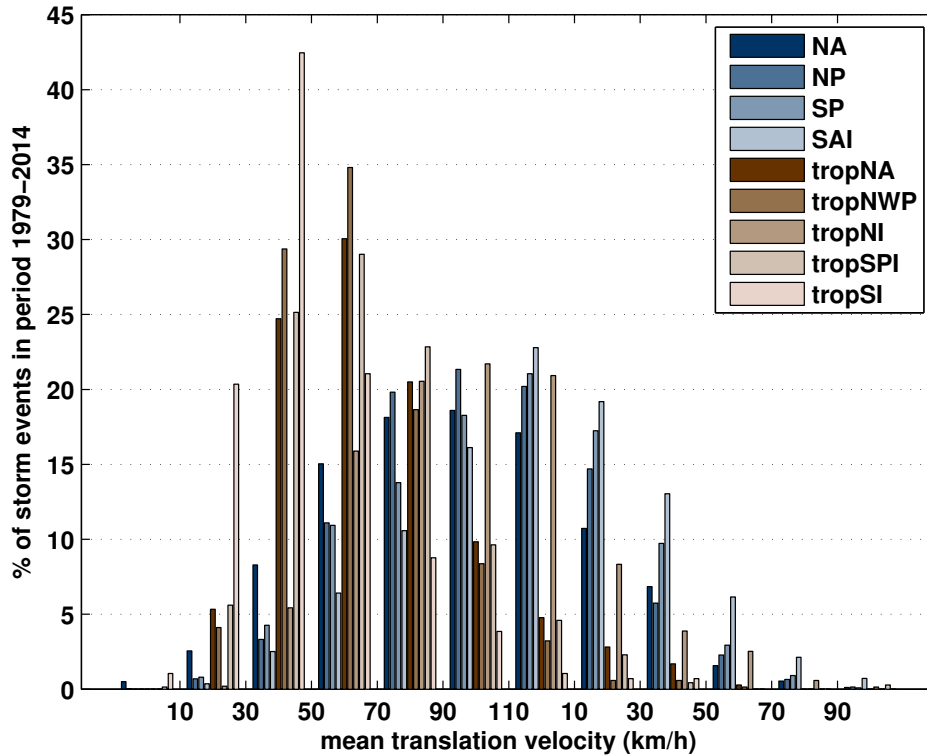
Figure 2.14.: Relative distribution (histogram) of mean (averaged over lifetime) trans-
lation velocity (distance between subsequent track positions divided by 6-hour time
steps) for samples of different storm types, compiled by geographic and seasonal se-
lection

slower than all other storm types. Analyzing tropical cyclones around Australia only,
Dare and Davidson (2004) find a mean translation velocity of 16 km/h which is about
half of the mean value found here for the tropSPI-region. Nevertheless, these numbers
are assessed to be in reasonable agreement, bearing in mind that the translation of
WiTRACK-derived track positions (if configured with `power_track`$\neq$0, as done for this
study) additionally includes a component accounting for the storm-internal relocation
of the center of mass (weighing storm intensity). To the author's knowledge, no study
exists, comparing translation velocities of tropical cyclones in different basins. Hence,
no reference exists in that respect. This is different when considering extra-tropical cy-
clones, which should be well comparable to WiTRACK-identified wind storms regarding
this parameter, though also affected by the internal relocation of the storms' center in-
cluded in the WiTRACK-results. Fig. 2.14 shows that SH extra-tropical wind storms

generally travel faster than their NH counterparts, and velocities in the South Atlantic to Indian Ocean are even faster than those over the South Pacific. Tilinina et al. (2013) found a modal value of 30–40 km/h for NH extra-tropical winter cyclones, which agrees reasonably well with the findings presented here. No obvious differences between NH and SH extra-tropical cyclones were identified by Neu et al. (2013), but they present relatively large uncertainty due to the underlying identification method with respect to this parameter. Lim and Simmonds (2007) show that SAI-cyclones generally travel faster than SP-cyclones, which is consistent with the results of the current study, but also with Hoskins and Hodges (2005). Additionally, it is noteworthy that NA-storms are generally slower than NP-storms according to WiTRACK. This is very likely to be the origin of the North Pacific exhibiting higher track densities but lower geneses densities than the North Atlantic in DJF (Figs. 2.5(d) and 2.6(d)).
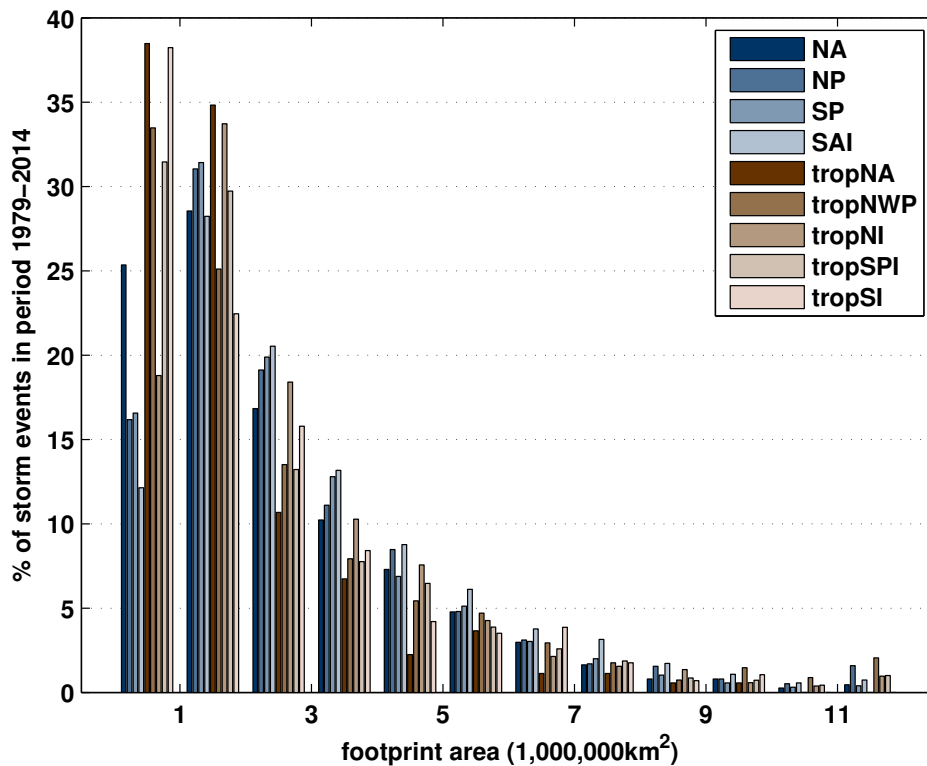


Figure 2.15.: Relative distribution (histogram) of footprints (total area affected for at least one time step) for samples of different storm types, compiled by geographic and seasonal selection

The comparison of wind storm properties is to be completed by two characteristics that integrate several primary features. The first of these integral parameters is the storms' footprint. The footprint is the total area affected by the storm for at least one time step, but neglecting storm duration at a given location. Thus, the footprint integrates storm size (area), translation velocity, and lifetime. A stationary system is associated with a smaller footprint, independent of its lifetime, than a traveling system of comparable size. Hardly any consistent differences between the various storm types can be detected from classes of very large footprints ($>7,000,000$ km$^2$). Largest differences are seen with respect to the lowest class (footprints$<1,000,000$ km$^2$): tropical wind storms exhibit small footprints much more often than extra-tropical storms and NH Atlantic events generally possess smaller footprints than their counterparts from other basins (NA compared to other extra-tropical regions, tropNA to tropical regions). Again, tropNI does not fit the impression given by the other tropical regions. Overall, these results regarding the storm footprints are in very good agreement with the findings related to translation velocity. Based on these results, translation velocity in general seems to be the primary factor for the footprint of a storm. No analog studies regarding footprint comparisons exist according to the author's knowledge. Mendes et al. (2010) analyzes the traveled distance of SH, and in particular South Atlantic extra-tropical cyclones while Dare and Davidson (2004) does the same for Australian tropical cyclones. However, without considering the size of these systems, no conclusions regarding the footprints are possible. Additionally, as already mentioned the region definition of Mendes et al. (2010) is not compatible with those used for the current study, and Dare and Davidson (2004) performed no comparisons to tropical cyclones of other areas.

Finally, wind storm intensity is compared for the different regions. A number of studies dealing with such comparisons exist. Regarding extra-tropical cyclones, Neu et al. (2013) show that SH cyclones tend to be deeper (in terms of surface or sea-level pressure) than their counterparts over the NH. According to Hodges et al. (2011), this is also true when considering 925hPa-winds or relative vorticity in 850 hPa as intensity measure. SP-cyclones on average are slightly more intense than SAI-cyclones when using the Laplacian of sea-level pressure as intensity measure but vice-versa if cyclone depth is used (Lim and Simmonds, 2007), and NA-cyclones are deeper than NP-cyclones in boreal winter (Schneidereit et al., 2010). For NH tropical cyclones, among others, Emanuel (2000) as well as Schenkel and Hart (2012) showed that systems over the Northwest Pacific are
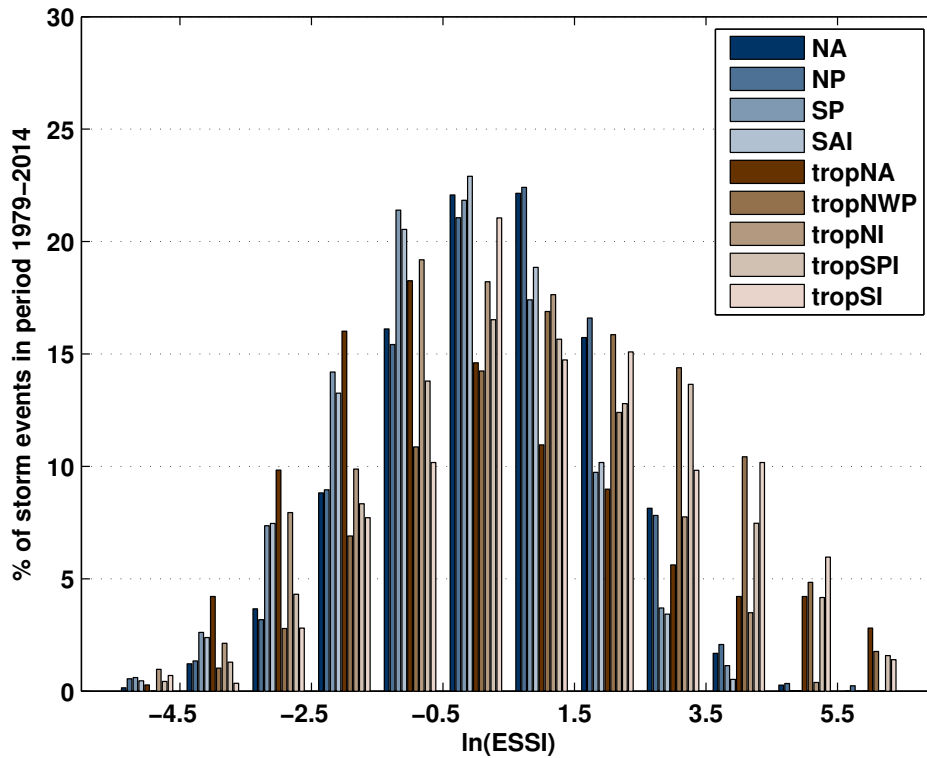
Figure 2.16.: Relative distribution (histogram) of storm intensity, measured by the logarithmic *Storm Severity Index* (SSI), for samples of different storm types, compiled by geographic and seasonal selection

more intense (in terms of wind speed) than over the North Atlantic. In this respect, Ramsay et al. (2012) found no significant differences between different clusters of tropical cyclones, spatially approximately matching the tropSPI- and tropSI-regions. The global inter-basin comparison of "best track" data performed by Schreck et al. (2014) yields no significant differences regarding the intensity distributions, except for the Northern Indian Ocean exhibiting much weaker systems than all the other basins. For the current study, the *Event Storm Severity Index* (ESSI), as developed by Leckebusch et al. (2008a) and already presented in Sec. 2.3.1 is compared for the wind storms identified in the nine defined regions and respective seasons. In particular, it is not the ESSI itself presented, but its natural logarithm, providing histograms easier to interpret (Fig. 2.16). From this analysis, it is obvious that the tropical wind storms (again except for tropNI) exhibit substantially larger fractions of very high ESSI-values. Calculating percentages of events beyond a ln(ESSI) of 2.5 (equivalent to an absolute ESSI of ∼ 12.2) yields 16.8%, 31.4%,

26.9%, and 27.4% for tropNA, tropNWP, tropSPI, and tropSI, respectively, while the extra-tropical regions present fractions of only 10.1% (NA), 10.5% (NP), 4.9% (SP), and 4.0% (SAI). This result was expected though not self-evident given the intensity scaling based on local climatology inherent to the (E)SSI. The tropical North Indian Ocean once again features numbers similar to the extra-tropics with 11.6% of its events possessing such high ESSIs. These percentages indicate two other facts to derive from the ESSI-distributions: (i) Wind storms over the tropical North Atlantic show a wider range regarding their ESSIs than the other tropical regions considered, with a comparably low fraction beyond ln(ESSI)=2.5, but a relatively fat tail of the distribution, presenting the largest fraction (2.8%) of all analyzed regions in the highest class of ln(ESSI)>5.5 ($\sim 245$). At the same time tropNA features the highest fraction of all considered regions at the lower end of the ESSI-distribution (14.3% below an ln(ESSI) of -2.5, equivalent to an absolute value of $\sim 0.082$). (ii) In terms of the ESSI, NH extra-tropical wind storms clearly tend to be more intense than SH storms, while no clear differences are visible between the basins within the respective hemisphere. Of course, these results are not directly compatible with the above-mentioned other studies, as the (E)SSI involves the intensity scaling based on local climatology. Instead, they complement existing results – based on absolute physical quantities – regarding potential damage risk, assuming that building standards are adapted to a local climatological threshold (such as the 98th percentile).

## 2.5. Summary and conclusions

As stated in Sec. 2.1, the current Chapter aims at several aspects: On the one hand side, it constitutes a comprehensive description of the WiTRACK-algorithm, which did not exist so far. As such it is meant to be a detailed documentation of the basic method used for the following Chapters 3 and 4. Apart from that, various scientific questions were addressed, some of them focused on the methodological aspect others purely climatological.

## 2.5.1. Methodological aspects regarding WiTRACK

From the methodological point of view, two main questions are addressed by this study:

Facing the various developments made to the WiTRACK-algorithm and described in Sec. 2.3.3, the first and natural question is, whether these changes basically improved WiTRACK. Regarding the technical advances and the flexibilization efforts, the advantages of the actual WiTRACK-algorithm are self-evident. The same is true for the automatic detection of poles and column-periodicity as well as the option of using an envelope-like secondary threshold to perform an agglomerative clustering for the identification procedure. As the latter feature is optional only, it constitutes an extension of identification possibilities and, hence, an additional (substantial) level of flexibilization. The basic difference with the last two substantial developments is that these constitute profound changes of the tracking procedure. The benefits of such changes have to be critically evaluated. The best solution in this respect would be an extensive sensitivity study including statistical analyses quantifying the potential improvement. However, such skill measures require a reference data set. Such reference data is not available for the core target of WiTRACK, that is extra-tropical wind storms. Not even for the more commonly addressed mid-latitudinal cyclones any kind of "best track" data exists. Choosing one specific cyclone track data set produced by one of the well-established objective identification methods is no solution, either. As impressively shown by Neu et al. (2013), these methods provide a high degree of compliance as far as statistics of large samples are considered (e.g. inter-annual variability, trends, spatial distribution, . . . ), but once individual tracks are compared, they disagree significantly for many events. For tropical cyclones, such "best tracks" as presented by Knapp et al. (2010) exist, constituting a natural reference. However, an additional fact has to be kept in mind: Indeed, the majority of WiTRACK-identified wind storms can be related to cyclonic events (e.g. noted by Nissen et al., 2014b; Kruschke et al., 2014, for the Mediterranean and the NH, respectively), but the WiTRACK-derived track positions based on weighted averages of wind storm intensity cannot be expected to generally match such "best track" positions that are indicating the core of the respective system. Hence, the distance between track positions would be of limited use when quantifying the WiTRACK-skill. In the light of these constraints, the author of this thesis refrained from such a quantitative evaluation of the WiTRACK-improvements. Instead, a qualitative though subjective assessment of the benefits achieved was performed for the totality of 56 European winter wind storms,

being the core subject of the analyses presented in Ch. 4. As already stated in Sec. 2.3.3, the presented changes made on the way to the actual revision of WiTRACK comprises the best solutions found for this set of 56 wind storm events. The underlying premise in this respect was to evolve the WiTRACK-scheme balancing progress regarding the tracking results and consistency with previous studies as far as possible. The case studies shown in Sec. 2.4.2 indicate that the actual WiTRACK-scheme yields satisfactory tracking results, even for challenging situations such as rapid succession of multiple wind storms. Certainly, the algorithm and its tracking routine in particular can be further optimized. According to the author's point of view, the most promising approach in that respect would be to incorporate some approximation of steering winds – also used by Hewson and Titley (2010) and the actual algorithm of Murray and Simmonds (1991) – in order to achieve an expected track position for the next time step which can be matched with actually existing wind storm positions. It should be possible to derive such an approximation from the surface wind components. Indeed, upper level circulation parameters would be more appropriate in this respect, but the usage of surface wind components would retain the basic WiTRACK-characteristic of being an impact-oriented, simple, and flexible tool, based on surface winds only. However, in tropical regions upper level steering is less "deterministic" for the track of an individual system. Hence, such tracking optimizations often coincide with a certain specialization of the scheme. Regarding the short-mentioned peripheral activities of applying WiTRACK to completely different variables, not necessarily containing directional components, such a development would be completely inappropriate.

The second methodological question raised in Sec. 2.1 is, whether WiTRACK and the used configuration is a suitable tool for identifying wind storms in other regions than the NH extra-tropics. This question is closely related to the results of the climatological analysis summarized in the following paragraph. Overall, this study strongly indicates that WiTRACK is equally appropriate for the SH extra-tropics and, in general, also very useful for analyzing tropical wind storms. However, for the latter the conclusion of this study is that the uniform configuration globally used here is convenient but not optimal. The spatial patterns of track-, geneses-, and lyses-densities and the climatological intra-annual variability analyzed for several tropical target regions, presented in Sec. 2.4.3, indicate that other wind storm types besides the targeted events related to tropical cyclones are identified in these regions. Especially when facing the results regarding

storm characteristics over the tropical Northern Indian Ocean, this conclusion seems to be very likely. Detailed analyses would be required to assess what kind of events these "unexpected" wind storms are. A new hypothesis to be tested in future work would be the assumption that the current application of WiTRACK identified many wind storms related to synoptic scale monsoon low-pressure systems. These systems are considerably larger, but associated with less distinct pressure gradients and wind speeds than typical tropical cyclones. These properties would be in line with the storm characteristics for the tropNI-region found here and – as shown in the recent study of Hurley and Boos (2014) – these systems feature a seasonal cycle which is also in very good agreement to the current study for the tropNI-region (Fig. 2.10(c)). When WiTRACK is used specifically for wind storms related to tropical cyclones, the conclusion of this study is that a configuration more tailored to the unique characteristics of these systems would be beneficial. Based on the results presented here, two parameters are rated as most important in this respect. Very basically, a higher threshold than the climatological 98$^{\text{th}}$ percentile seems to be appropriate in these regions. The case study of Hurricane "Sandy", presented in Fig. 2.3, shows that the actual storm (subjectively) exhibits a very clear, approximately circular structure while WiTRACK identifies large areas peripheral to this structure also exceeding the 98$^{\text{th}}$ percentile used as threshold here. Fig. 2.12 reveals that the size distributions of wind storms over tropical regions are very similar to those from extra-tropical regions. Assuming that the majority of the identified events is related to tropical and extra-tropical cyclones, respectively, this result is not in line with basic meteorological knowledge about these systems. A higher threshold certainly will lead to smaller storm clusters identified. However, a more detailed analysis regarding an appropriate threshold for the tropical regions would be necessary this respect. Either reanalysis data (essentially model results including respective biases) should be scanned for typical wind speeds of well represented tropical cyclones or observations could be analyzed to translate traditional absolute thresholds (18 m/s for a tropical storm, 33 m/s for hurricane force) to quantiles for the affected regions in order to derive a useful quantile for WiTRACK-applications targeted on tropical cyclones. To the author's knowledge, no study exists so far yielding a threshold for wind damage in tropical regions (see also respective parts of the following paragraph). The second parameter which clearly could be adapted for WiTRACK-studies focusing on tropical cyclones is the allowed translation distance. As shown in Fig. 2.14, wind storms in the tropical regions exhibit clearly slower translation than extra-tropical systems. For the study

of Sakuth (2011), a maximum translation distance of 500 km (for 6-hourly data) was used for WiTRACK. Meanwhile implemented developments, in particular the storm-size dependent maximum total distance (see Eq. 2.2 in Sec. 2.3.3) allows significantly smaller values for this parameter. Further tests – though not finalized yet – indicate that a value equivalent to 40 km/h is appropriate for this purpose. However, such an adaptation alone is not sufficient to eliminate all other kinds of wind storms such as the potential monsoon lows identified over the Northern Indian Ocean.

## 2.5.2. Climatological analysis of potentially damaging wind storms

According to the author's knowledge, the first consistent global climatology of potentially damaging meso- to synoptic-scale wind storms is presented in Sec. 2.4.3. Based on a uniform configuration of the WiTRACK-algorithm the full *ERA-Interim*-reanalysis (using 10m-winds, neglecting Jan./Feb. 1979 to consider complete seasons only) currently available is scanned for severe wind storm events, globally applying the concept of Klawa and Ulbrich (2003) and Leckebusch et al. (2008a). According to this, such events can be identified as extensive areas of wind speeds beyond the local climatological 98th percentile. It should be noted, that this concept was developed and tested primarily for the mid-latitudes and Europe in particular. Hence, the transfer to the global domain must be regarded as experimental. Studies relating wind speeds to (potential) damage for tropical cyclones are so far based on absolute wind speeds (e.g. Bell et al., 2000; Emanuel, 2005) or include generally fixed thresholds (no reference to local climatology, e.g. Kantha, 2006; Powell and Reinhold, 2007) originating from the traditional *Saffir-Simpson Hurricane Scale*. In view of this fact, the experiment can be considered as successful. The spatial analysis of the results achieved by this WiTRACK-application (track-, geneses-, and lyses-densities presented in Fig. 2.5, 2.6, and 2.7, respectively) reveal a global pattern of damage-prone wind storms that is in very good agreement with basic scientific knowledge assembled from multiple studies, each of those focused on limited areas and/or specific storm types, such as Schreck et al. (2014) dealing with tropical storms or Neu et al. (2013) analyzing extra-tropical cyclones. Seasonal variability of these patterns and a slightly more detailed analysis in this respect for nine target regions (two extra-tropical NH and SH each, three in the NH tropics, two in the SH tropics, presented in Figs. 2.8–2.11) yield further confidence in WiTRACK and the

used uniform configuration. Overall, they also confirm the expectations derived from the more limited/focused studies and related general knowledge. The new scientific aspect provided by the current study is that the uniform objective identification approach facilitates comparative analyses of properties of very different wind storm events (completely different regions and seasons) in a consistent way. Characteristics such as storm area, lifetime, translation velocity, storm footprint, and storm intensity are compared (Figs. 2.12–2.16) based on exactly the same objective identification and tracking procedure (and data basis).

These comparative analyses yield the following central results: Wind storms over the tropical Northern Indian Ocean show characteristics very different from all other tropical regions studied. As already stated in Sec. 2.5.1, the consequent new hypothesis – supported by a comparison to the study of Hurley and Boos (2014) – is that the majority of the events identified here are monsoon low-pressure systems. The profound test of this hypothesis is a potential subject to future work. However, for the current study the tropNI-results are excluded from comparisons to tropical storms in the other regions. Wind storms over the tropical North Atlantic are considerably smaller (distribution shifted towards smaller areas) than their counterparts over all other studied tropical basins and extra-tropical events. However, these size distributions are very sensitive to the chosen identification threshold (98$^{\text{th}}$ percentile) which is probably too low for an appropriate identification of wind storms related to tropical cyclones. Regarding extra-tropical systems, it is shown that wind storms over the NP are slightly larger than over the NA and those over SAI tend to be larger than over SP which is in line with studies on extra-tropical cyclone radii for the respective regions (Rudeva and Gulev, 2007; Schneidereit et al., 2010; Lim and Simmonds, 2007). Apart from the expected finding that tropical wind storms exhibit longer lifetimes than extra-tropical storms, no remarkable result is found in this respect. Analyzing translation velocities, it is shown that SH wind storms are generally faster than NH systems. The intra-hemispheric comparison yields the NA (SP) storms to be slower than their NP (SAI) counterparts which is again confirmed by different cyclone-focused studies (e.g. Hoskins and Hodges, 2005; Lim and Simmonds, 2007). This difference in translation velocity is also found to be the primary origin of the apparent contradiction when comparing wind storm geneses and track densities for NA and NP. The former features considerably higher geneses densities but lower track densities than the latter. No additional information can be ex-

tracted from the comparison of wind storm footprints but qualitative assessment yields the analyzed translation velocities to be the primary driver of footprint area. Finally, all regions and respective storms selected are compared in terms of their ESSI. According to these results, tropical wind storms tend to be more (potentially) damaging than extra-tropical events, with the tropNA-storms exhibiting the largest spread. Extra-tropical wind storms over the NH tend to show considerably higher ESSI-values than over the SH. All regions feature high inter-annual variability regarding the number of wind storms (within defined seasons), hampering serious indications of any long-term trends or low-frequency variability. This is different when considering the totality of all wind storms identified globally (Fig. 2.4) which implies a positive trend over the period covered by the *ERA-Interim* reanalysis. However, facing the rather short time span (and limited spatial coverage) of observational records, such analyses are better studied by means of long-term integrations of general circulation models. The current Chapter of this thesis definitely fosters confidence in the underlying concept and the flexibility of the WiTRACK-algorithm to be utilized for studies in that respect.

# 3. Probabilistic evaluation of decadal predictions of NH winter storms

This chapter is to a great extent identical to a manuscript, submitted to *Meteorol. Z.*. Only Sec. 3.2.2 is adapted to match the introduction of the WiTRACK-scheme, already provided in Ch. 2, including a summarizing table regarding the WiTRACK-configuration. A revised version of this manuscript is now published (Kruschke et al., 2015). The initiative to write this paper and the conduct of all analyses is to be associated with the author of this thesis. Dr. Wolfgang A. Müller and Dr. Holger Pohlmann receive credit for performing the *MPI-ESM* model simulations subject to all analyses presented here. Additionally, they gave helpful advice on early drafts of the manuscript. Christopher Kadow implemented the WiTRACK-scheme on a server with access to all model data and performed the trackings according to the guidelines set up by the author of this thesis. Prof. Dr. Uwe Ulbrich is responsible for the impetus of considering the study of Kharin et al. (2012), which was the starting point for the drift correction used in this study. Additionally, he contributed by internally reviewing a preliminary version of the manuscript. The same is true for PD Dr. Gregor C. Leckebusch and Dr. Henning W. Rust, the latter most deeply involved in discussing analysis results, especially with respect to the drift correction.

## 3.1. Introduction

Beyond century time scale climate change, climate evolution exhibits substantial variability driven by natural processes. At decadal timescales these variations coincide with typical planning horizons of political and economic stakeholders. Thus, climate predictions for the next decade(s) would be of great socio-economic value (Solomon et al., 2011) if proven to be of significant skill with respect to relevant parameters. Winter storms are responsible for approx. 25.2bn€ overall losses in Europe during 1980–2006 (Munich RE Group, 2008a, in values of 2006), making them the most expensive type of natural catastrophe in this area. These numbers highlight the potential relevance of decadal predictions in this respect, further emphasized by the fact that the frequency of such events exhibits a high degree of inter-annual to multi-decadal variability (see e.g. Donat et al., 2011b; Nissen et al., 2014a; Welker and Martius, 2014). The German initiative *Mittelfristige Klimaprognosen* (MiKlip) is dedicated to the development of a model system to produce skillful predictions for up to a decade ahead. The present study evaluates basically five MiKlip-experiments conducted so far, focusing on the predictive skill regarding the frequency of winter storms over the Northern Hemisphere.

Following some pioneering studies regarding potentials and limitations of decadal climate predictions (see Meehl et al., 2009, for a thorough review), the *Coupled Model Intercomparison Project* in its fifth phase (CMIP5) included a new framework for initialized decadal predictions. This was done to set the scene for a coordinated assessment of current earth system models´ ability to produce reliable climate predictions. One of the hindcast experiments (*baseline0*, see Sec. 3.2.1), analyzed in this study, actually contributes to CMIP5, as these decadal hindcasts are identical to the CMIP5 decadal experiments. The basic idea behind decadal predictions is to initialize the models with an observed climate state instead of an arbitrary state as it is the case for standard transient runs, such as future climate projections. It is expected that these simulations subsequently follow to some degree the observed climate evolution, as they should include the unforced component of climate variability (Taylor et al., 2012). Several studies exist, searching for optimal strategies in this respect (e.g. Matei et al., 2012; Smith et al., 2013; Hazeleger et al., 2013; Counillon et al., 2014; Polkova et al., 2014). Basically their results are promising, however, further improvements are necessary enabling the model to catch the right phase (and amplitude) of actual climate evolution.

Another issue is the existence of model biases; if these are constant over time, they are rather unproblematic and can be overcome by standard evaluation procedures based on (climatological) bias adjustments or analyzing anomalies. However, biases represent some challenge if not constant over time. While a bias of a single prediction growing over lead time might be just a bad forecast, more systematic issues may exist, potentially masking any predicted climate signal. In that context, it is essential to understand the reasons of such systematically changing biases. One possible reason is the existence of externally forced long-term changes of the model differing from those evident in the chosen observational data. Another reason is that the initialization (especially full-field initialization) sets the model close to the observational state, which is not compatible with its own preferred (systematic error) state (Meehl et al., 2014). In that case the model will exhibit systematic drifts back towards its equilibrium over simulation time. For coupled models, including an ocean component with its inertial character, these (not necessarily monotonous) drifts might be of significant influence on the model results for several years or even decades, depending on the order of the initial shock. Such systematic but non-stationary biases require appropriate bias-correction methods (see e.g. Kharin et al., 2012; Hawkins et al., 2014) applied to decadal climate predictions before they can be properly analyzed regarding their predictive signals and the associated skill of these predictions.

In recent years, more and more studies have become available, dedicated to skill assessment of existing individual forecast systems (e.g. Müller et al., 2012, 2014; Boer et al., 2013; Goddard et al., 2013) or multi-model ensembles (e.g. van Oldenborgh et al., 2012; Doblas-Reyes et al., 2013; Meehl and Teng, 2014). Most of these studies consider fields of primary meteorological parameters, such as mean surface air temperature and precipitation. Other studies focus on specific meteorological phenomena or indices, e.g. Garcia-Serrano and Doblas-Reyes (2012), assessing decadal prediction´s skill regarding large scale mean temperatures, or Scaife et al. (2014), analyzing predictions of the *Quasi-Biennial Oscillation* (QBO). So far, very few studies deal with the predictive skill with respect to the frequency of certain meteorological events (Eade et al., 2012; Hanlon et al., 2013; Kruschke et al., 2014). As the common understanding of the climate system emphasizes the role of the ocean for decadal climate variability, many studies concentrate on this subsystem, analyzing (global) fields of sea surface temperatures (SST) and upper ocean heat content (e.g. Matei et al., 2012) or the oceanic variability of specific

regions and oceanic phenomena such as the North Atlantic and the Atlantic Meridional Overturning Circulation (AMOC; e.g. Kröger et al., 2012; Pohlmann et al., 2013b) or the Atlantic Subpolar Gyre (Yeager et al., 2012; Robson et al., 2012, 2014).

This paper extends the study of Kruschke et al. (2014) which showed that the first two development stages of the MiKlip decadal prediction system exhibit promising levels of skill with respect to the frequency of (intense) Northern Hemisphere extra-tropical cyclones. The present study (i) is more impact-oriented, considering winter storms instead of cyclones (see Sec. 3.2.2), (ii) is more elaborated in terms of adjusting for potential model drifts over forecast lead time (see Sec. 3.4), and (iii) considers additional hindcast experiments (a total number of five instead of two) as well as different combinations of the original hindcast experiments (see Sec. 3.2.1) in order to derive more robust estimates of predictive skill.

The following questions are addressed:

1. Do decadal predictions of winter storm frequency provide significantly more information for the next few years than the climatological forecast (i.e. using the climatology as a forecast)?

2. Does initialization from actual climate states (as realized so far) provide any additional value compared to uninitialized simulations (including responses to external forcing only) and linear approximations of long-term change?

3. Is any of the so-far-realized initialization strategies clearly superior to the others with respect to predictive skill regarding winter storm frequencies?

4. What is the effect of the chosen parametric drift-correction approach on estimations of predictive skill compared to the standard non-parametric procedure, used in many other studies?

Sec. 3.2 describes all used data, that is the different decadal hindcasts, but also the uninitialized simulations and reanalyses. Additionally, Sec. 3.2 contains the method to identify winter storm events, the calculation of storm frequencies, as well as the chosen probabilistic verification metric. Sec. 3.3 is dedicated to analyzing systematic differences, such as biases and deviating long-term trends, between MPI-ESM-LR and reanalysis. Subsequently, an appropriate approach to statistically adjust the model for these systematic deviations is shown in Sec. 3.4. The results of probabilistic hindcast verification can be found in Sec. 3.5, while Sec. 3.6 summarizes the paper and its conclusions.

## 3.2. Data and methods

### 3.2.1. Data

All model simulations, analyzed in this study, are conducted using the *Max-Planck Institute Earth System Model* in a low-resolution configuration (MPI-ESM-LR, see Giorgetta et al., 2013). The atmospheric component of MPI-ESM-LR is ECHAM6 (Stevens et al., 2013) in T63L47-resolution while the ocean is represented by MPIOM (Jungclaus et al., 2013) with a nominal horizontal resolution of 1.5° and 40 levels.

Table 3.1.: Overview of analyzed hindcast experiments

| Hindcast experiments | Initialization atmosphere | Initialization ocean | Ensemble members |
|---|---|---|---|
| *baseline0* | none | anomalies from NCEP/NCAR-forced ocean run | 3 (10) |
| *baseline1* | full-fields from ERA40/ERA-Int. | anomalies from ORA-S4 | 10 |
| *ORAff* | " " | full fields from ORA-S4 | 10 |
| *GECCOano* | " " | anomalies from GECCO2 | 3 |
| *GECCOff* | " " | full fields from GECCO2 | 3 |

Basically, five sets of decadal hindcasts are analyzed in this study. Each consists of 41 hindcasts, initialized annually at 1$^{st}$ January 1961-2000 and integrated for ten years each. The five hindcast sets differ with respect to the underlying initialization strategies, as summarized in Tab. 3.1. The starting-point within the *MiKlip*-initiative is called the *baseline0*-system, identical to the decadal prediction set up used for the CMIP5-exercise. In order to generate the initial conditions for each *baseline0*-hindcast, an ocean-only-experiment was conducted, forcing MPIOM with atmospheric data from *NCEP/NCAR*-reanalysis (Kalnay et al., 1996). The ocean temperature and salinity anomalies of this experiment are then used for nudging a run of the coupled model. The initial states of the decadal hindcasts are taken from this run. Only the oceanic component (i.e. not the atmosphere) is relaxed in this way to the observed state for *baseline0*. This system was introduced by Müller et al. (2012), also analyzing its performance with respect to decadal forecasts of seasonal mean surface temperatures.

The remaining four hindcast sets were initialized by nudging the coupled model with

fields directly derived from reanalyses. Regarding the atmospheric component, this is done identically for all four sets via full-field-initialization from ERA40 (Uppala et al., 2005, for the initializations 1961-1989) and ERA-Interim (Dee et al., 2011, for initializations since 1990). The difference between these four hindcast sets is to be found in the ocean initialization. Two sets were initialized from the ORA-S4-ocean-reanalyses (Balmaseda et al., 2013) while GECCO2 (Köhl, 2015) was used for the other two. Each of these pairs is split up by using full-field-initialization (i.e. nudging the ocean model to absolute values of the ocean reanalysis) and anomaly-initialization (nudging to values resulting from ocean reanalysis anomalies that were added to the model´s climatology), respectively. The system characterized by anomaly-initialization from ORA-S4 is called *baseline1* in the *MiKlip*-context and was elaborately described by Pohlmann et al. (2013a), studying the decadal prediction skill of this system in comparison to both, *baseline0* and a prediction system with different resolution. For this study, we call the full-field-initialized hindcasts based on ORA-S4 simply *ORAff*. The two GECCO2-initialized systems are accordingly called *GECCOano* and *GECCOff* in this study. All ensembles are generated by lagged-day-initialization.

The five hindcast sets also differ in terms of ensemble size (see Tab. 3.1). Most of the hindcast experiments consist of three ensemble members only. These very small ensemble sizes pose a challenge for probabilistic hindcast verification, though if analyzing parameters exhibiting low signal-to-noise-ratios, deterministic verification (of the ensemble mean) is similarly difficult. This means that verification results will suffer from high uncertainty, which – in the field of decadal predictions – is additionally fueled by the limited number of independent hindcasts, i.e. initializations. To reduce this uncertainty stemming from the small ensemble sizes, but also to more systematically examine the effects of the different initialization strategies, we additionally analyze the skill of ensembles composed by different combinations of the above-mentioned hindcast experiments as listed in Tab. 3.2 (after separately adjusting them according to Sec. 3.4). Two of these multiple system ensembles comprise all available ensemble members produced by anomaly- (*ano-Ens.*) and full-field-initialization (*ff-Ens.*) respectively. Two more multiple system ensembles compound all members initialized from ORA-S4- (*ORA-Ens.*) and GECCO2-ocean-reanalysis (*GECCO-Ens.*), respectively. Additionally, a *Grand Ensemble* is put together by combining all available hindcasts, including *baseline0*.

The benefit of the initialization must be assessed. As mentioned previously, this can

Table 3.2.: Overview of analyzed multiple system ensembles, produced by combining
different hindcast experiments, listed in Tab. 3.1

| Multiple system ensembles | Hindcast experiment combinations | Ensemble members |
|---|---|---|
| *ORA-Ens.* | *baseline1 + ORAff* | 20 |
| *GECCO-Ens.* | *GECCOano + GECCOff* | 6 |
| *ano-Ens.* | *baseline1 + GECCOano* | 13 |
| *ff-Ens.* | *ORAff + GECCOff* | 13 |
| *Grand Ens.* | *baseline0 + baseline1 + ORAff + GECCOano + GECCOff* | 29 (36) |

be done, computing the skill against a forecast representing climatological conditions. Knowing about observed climate change assigned to transient greenhouse gas and aerosol forcing, the larger challenge of the initialized forecasts is to beat the so-called uninitialized simulations. An ensemble of ten such simulations is generated by starting them from randomly chosen states of a long pre-industrial coupled control simulation. According to the years covered by the analyzed hindcasts, we examine only those model years representing the period 1961-2011.

To assess the quality of the hindcasts (and the uninitialized runs) we use the reanalyses of the *European Centre for Medium-Range Weather Forecasts* (ECMWF). Winter storm frequencies are determined per boreal winter half year (ONDJFM, see below) in this study. Thus, ERA40 is used for the winters 1961/62-1989/90 and ERA-Interim for 1990/91-2011/12 which is in correspondence to the above-mentioned atmospheric initialization of four hindcast experiments.

### 3.2.2. Winter storms: identification and frequency calculation

Northern Hemisphere winter storms are identified from reanalysis and model data via the objective WiTRACK-scheme (using revision r167), presented in Ch. 2. Tab. 3.3 presents a summary of the WiTRACK configuration used for the study presented in this chapter, targeted at meso-alpha- to synoptic-scale wind storms related to extra-tropical cyclones.

As for the results shown in Ch. 2, winter storm frequencies are calculated as track densities on a pre-defined 2.5°-grid as the number of tracks crossing a circular region.

Table 3.3.: WiTRACK-configuration used for tracking NH winter wind storms from decadal hindcasts and ERA reanalyses (as observational reference)

| | |
|---|---|
| period covered: | winter (ONDJFM) 1961/62–2010/11 |
| domain: | Northern Hemisphere |
| boundary sponge zone: | between 0°N and 10°N |
| considered variable: | 6-hourly instantaneous 10m-wind speed |
| threshold: | empirical 98$^{th}$ percentile of 10m-winds, calculated from reference period 1961–2000 (all timesteps) ERA-Int. percentile of 1979–2000 was climatologically corrected to match 1961-2000 period by scaling with interpolated ratios of resp. ERA40 percentiles |
| secondary thresh.: | none |
| minimum size: | 150,000km$^2$ |
| grid box weighting during tracking: | area multiplied by third power of relative threshold exceedances (SSI-contribution, see Eq. 2.1) |
| maximum translation distance: | 600km |
| factor for max. internal relocation: | 0.5 |
| minimum lifetime: | 4 timesteps $\sim$ 18 hours |
| grid box weighting for SSI-calculation: | standard SSI-contribution (see Eq. 2.1, equivalent to tracking) |

Allowing for spatially less accurate decadal forecasts, a comparably high degree os spatial aggregation was applied, counting the events within a radius of 1000km (great circle distance) around the respective grid points per boreal winter half year (ONDJFM). To avoid boundary effects and focus on extra-tropical phenomena, only results north of 30°N are used for all further analyses in this study.

To prevent inconsistencies regarding the observational reference, the winter storm frequencies of ERA40 were corrected to match mean and variance of ERA-Interim for the 22 winters existent in both data sets (1979/80-2000/01). Additionally, correlations of winter storm frequencies between the two data sets were calculated, based on these 22 winters. Those grid points exhibiting insignificant ($\alpha > 1\%$) correlations were rated as not reliable and thus excluded from all further analyses (masked in gray for all Figs.).

### 3.2.3. Probabilistic hindcast verification

Winter storm frequencies for individual seasons were mapped onto one of three categories: below normal, normal, or above normal. The categories are separated by the first and second terciles, empirically derived from the 51 winters 1960/1961-2010/2011 (i.e. 17 values in each category). For the model, all ten ensemble members of the uninitialized runs and the respective period (i.e. 510 model winters) were used to derive corresponding thresholds.

While the reanalysis provides one certain category as observational reference (a probability of 1 for this category), the fraction of model ensemble members forecasting one specific category yields the respective forecast probability. We consider cumulative probabilities for the three classes and hence use the *Ranked Probability Score* (RPS) as probabilistic verification measure. We apply an estimator of the RPS, developed by Ferro (2007, see also Ferro et al., 2008) and adapted by Kruschke et al. (2014) to account for ensemble size varying over initializations (as in the case of *baseline0*) in order to eliminate the systematic ensemble-size-dependent bias of the RPS:

$$\text{RPS}_{\tau,M} = \frac{1}{I} \sum_{i=1}^{I} \sum_{k=1}^{K} (F_{\tau,i,k} - O_{t,k})^2 - \frac{M - m_i}{M(m_i - 1)} F_{\tau,i,k}(1 - F_{\tau,i,k}) \tag{3.1}$$

$F_{\tau,i,k}$ is the cumulative forecast probability of class $k$ (with $K = 3$) derived from the forecast ensemble of initialization $i$ (with $I = 41$) for a specific forecast lead time $\tau$. $O_{t(i,\tau),k}$ is the cumulative probability of class $k$ from observations for the time $t(i,\tau)$, corresponding to the time of initialization and forecast lead time. $O_{t(i,\tau),k}$ effectively is the Heaviside step function with $O_{t(i,\tau),k} = 1$ if class $k$ or lower is observed and $O_{t(i,\tau),k} = 0$ otherwise. The second term of the equation constitutes the bias-correction, subtracting the systematic RPS-bias, an ensemble of size $m_i$ suffers from, when compared to an ensemble of size $M$.

The benefit of a forecast compared to a reference forecast is quantified by the *Ranked Probability Skill Score* (RPSS)

$$\text{RPSS}_{\tau} = 1 - \frac{\text{RPS}_{\text{fc},\tau}}{\text{RPS}_{\text{ref},\tau}}. \tag{3.2}$$

Significance ($\alpha < 5\%$) of calculated skill scores is assessed by 1000-fold bootstrapping from the available 41 hindcasts.

## 3.3. Systematic model deviations: climatology and long-term trend

As already mentioned in Sec. 3.1, systematic deviations of the model simulations from observational reference might pose serious challenges for verification. The most problematic issues in this respect are model biases that are not constant over time but a result of modeled externally forced long-term trends differing from observations or drifts of the predictions following a potential initialization shock. It is important to estimate these different components of a model bias and properly separate them. This section is dedicated to the assessment of the model´s climatology and long-term trend in comparison to ECMWF´s reanalyses. Differences regarding these two characteristics should basically be evident and comparable for the uninitialized simulations and the initialized hindcasts. The latter may additionally suffer from drifts but these are to be addressed in Sec. 3.4. Hence, the regression analysis with a linear trend in time ($N = N_0 + N_t \cdot t$) for the annual winter storm frequencies ($N$) presented in this section are based on the uninitialized simulations of MPI-ESM-LR and the ERA-reanalyses only. Note that trend estimates are generally a result from external forcing and potential multi-decadal natural variability. For the ensemble of ten uninitialized simulations, the influence of multi-decadal variability on the trend will be averaged out as these ten transient simulations should represent very different phases of natural low-frequency variabilities.

The offset parameters ($N_0$) derived for the linear trend (Fig. 3.1) are equivalent to climatological winter storm frequencies over the analyzed period (winters 1961/62-2009/10), while the slope parameters ($N_t$, Fig. 3.2) naturally estimate a linear change in this respect. Climatological NH winter storm frequencies are highest over the oceans and considerably lower over the continents (black contours in Fig. 3.1 denote absolute numbers from ERA-reanalyses). Winter storm frequencies over the North Pacific are slightly higher than over the North Atlantic and the overall patterns are in very good agreement with other studies on extra-tropical cyclones and storm track diagnostics (see e.g. Ulbrich et al., 2008, 2009). The differences between model and reanalyses (color shadings)
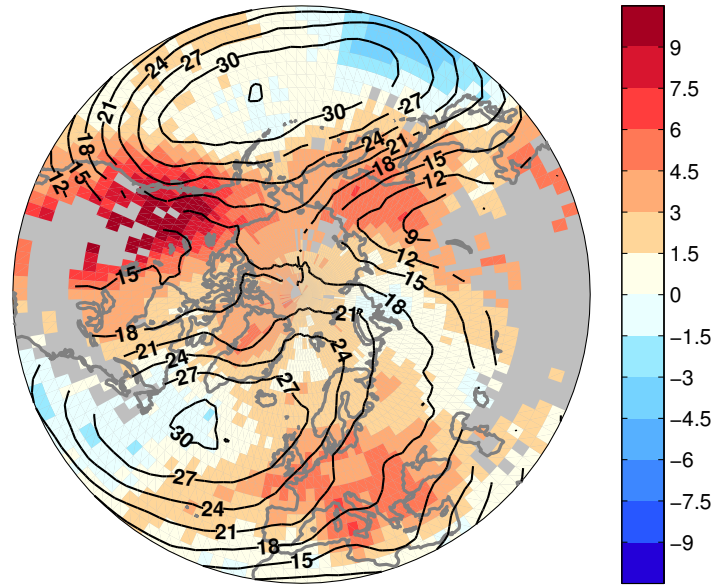
Figure 3.1.: Climatological winter storm frequency (number of tracks per ONDJFM and
    1000km radius) from regression analysis (offset) for ERA-reanalyses (black contours)
    and biases of uninitialized simulations of MPI-ESM-LR (colored), calculated over
    winters 1961/62-2009/10

reveal patterns very similar to those reported by Kruschke et al. (2014, see their Fig. 1b)
for cyclones instead of windstorms. The North Atlantic storm track is too zonal in the
model with a slightly negative bias in the core of the storm track and positive devia-
tions along the southern edge, especially over Europe. Winter storm genesis over the
Mediterranean is overestimated in the model, which is reflected by the positive frequency
bias in this region. The North Pacific storm track, however, is shifted northward in the
model (especially over the Northwest Pacific) and extends too far over North America,
resulting into local winter storm frequencies up to more than 50% higher than those
from reanalyses.

The slope estimates reveal several interesting features regarding linear changes of NH
winter storm frequencies over the considered period. According to ERA-reanalyses
(Fig. 3.2(b)), the North Pacific, as well as the mid-latitudinal North Atlantic and the po-
lar latitudes are dominated by significantly (95%-confidence intervals of slope estimates
not including zero) positive trends. Trends over the North Pacific seem more pronounced
than those over the North Atlantic. The strongest trends in reanalysis data are to be
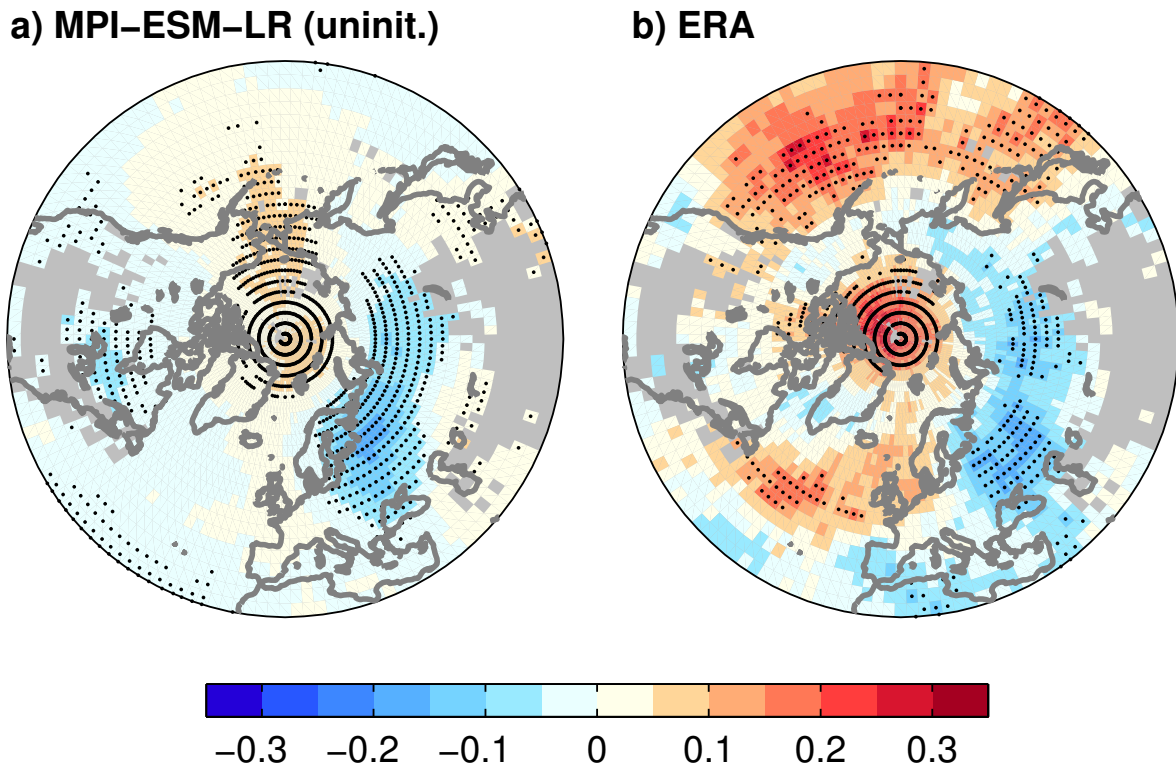
Figure 3.2.: Linear change of winter storm frequency per year from regression analysis (slope) calculated over winters 1961/62-2009/10 for a) uninitialized simulations of MPI-ESM-LR and b) ERA-reanalyses; Trend significance (95%-confidence intervals of slope estimates not including zero) marked as black dots

found in the Arctic. On the other hand, the Eurasian continent exhibits predominantly negative trends.

The uninitialized simulations of MPI-ESM-LR (Fig. 3.2(a)) show considerably weaker trends for most regions. Regarding the sign of the trend, model and reanalysis agree for the significantly positive trend (albeit considerably weaker in the model) over the Arctic. The significantly negative trend over Eastern Europe and Russia is also picked up by the model, even the magnitude is confirmed, but however, this trend pattern is slightly shifted to the North. Additionally, MPI-ESM-LR features significantly negative trends over North America, which cannot be found in reanalyses. For large parts of the North Atlantic and North Pacific, MPI-ESM-LR exhibits no significant trends.

Both features tackled here – a climatological bias as well as deviating long-term trends of the model – are of relevance for decadal predictions and their verification. Climato-

logical bias patterns as depicted in Fig. 3.1, proving systematic misrepresentations like shifts or deformations of relevant variability patterns and phenomena, pose serious issues with respect to the interpretation and verification of predictions regarding the respective phenomenon and its implications. Even if the model would be able to perfectly predict the temporal evolution of a specific feature, e.g. the North Atlantic storm track activity, its deterioration results into related signals at locations differing from observations, hampering straight-forward use of the predictions and their verification in the sense of grid-point-wise comparisons.

Regarding deviating trends, two possible reasons have to be distinguished. Part of the trends obtained for reanalysis data will be due to natural multi-decadal variability with associated periodicities and phases such that they mimic a trend for the available observational period used. As already stated, we cannot expect the uninitialized simulations to exhibit similar trends, while the initialized hindcasts may be able to (at least partially) pick up such behavior. We assume that the model generally shows an appropriate response to external forcing (at least the same sign) with respect to the frequency of winter storms. Hence, we conclude that most of the major trend discrepancies between model (Fig. 3.2(a)) and reanalyses (Fig. 3.2(b)) are a result of such multi-decadal variabilities appearing as linear trends for reanalyses over the analyzed period. However, another part of the observed long-term trends is a result of external forcing. It is obvious that a model not exhibiting the same response to that must be limited in terms of its predictive skill: it misses one component relevant for climate evolution on decadal time scales. Furthermore, if the long term trend differs between model and reanalysis, the model drift after initialization is a function of time. This needs to be accounted for, see Sec. 3.4.

## 3.4. Statistical adjustment of model bias, long-term trend, and hindcast drifts

The International CLIVAR Project Office (2011) recommends subtracting a climatological bias from anomaly-initialized predictions and uninitialized simulations, while for full-field-initialized predictions, the lead-time dependent bias is subtracted to account for probable model drifts over forecast lead time. Kharin et al. (2012), however, showed
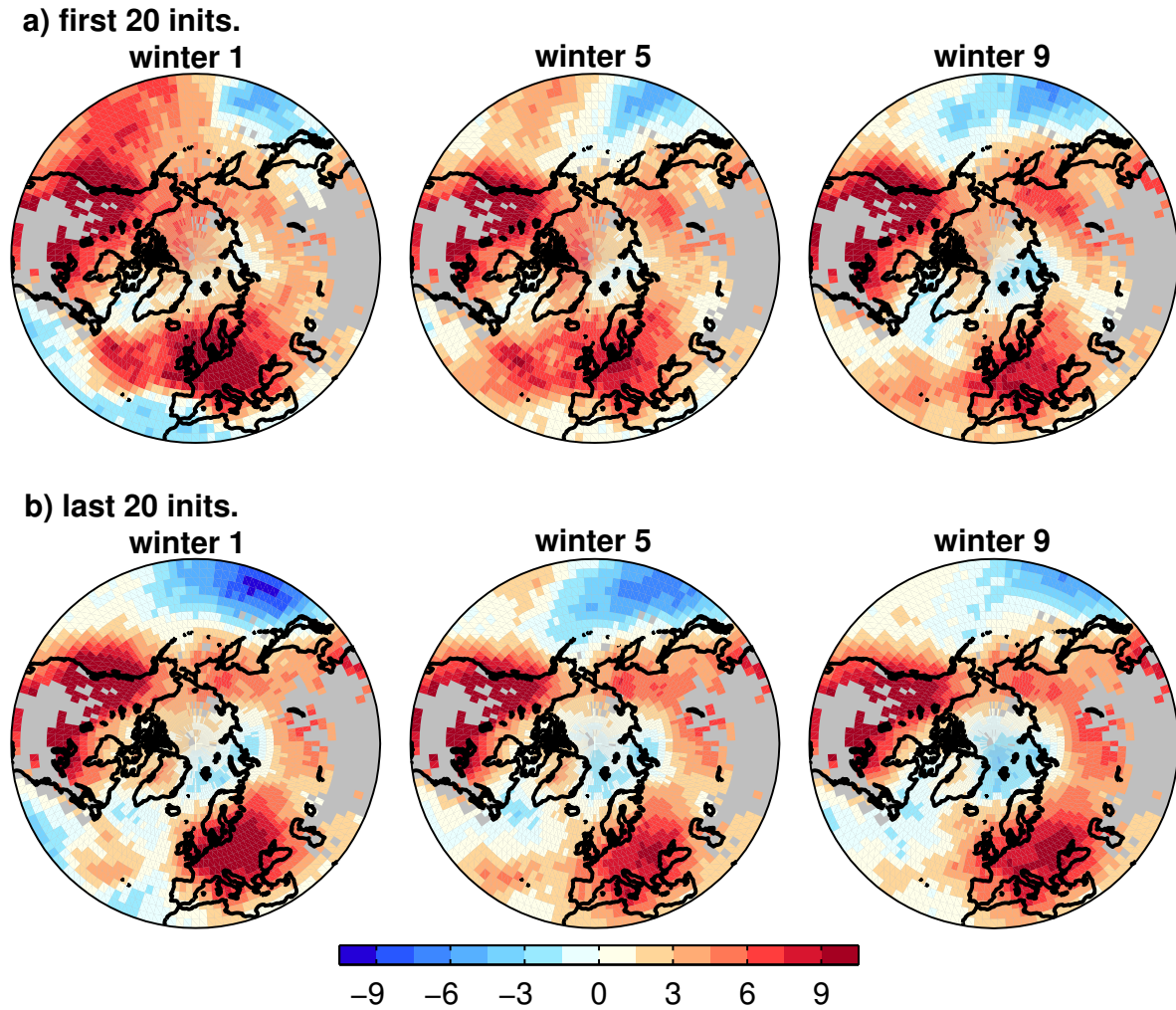
Figure 3.3.: **Lead-time-dependent bias** of *ORAff*-hindcasts with respect to NH winter storm frequency (number of winter storm tracks per ONDJFM within 1000km radius) with ERA-reanalyses as reference: a) calculated from first 20 initializations (1961-1980) only; b) calculated from last 20 initializations (1982-2001) only

that this approach of assuming a model drift being constant for all initialization times is problematic, especially in the presence of long-term climate change signals differing between model and observations. We illustrate this issue with respect to winter storm frequencies using the example of the *ORAff*-hindcasts (all 10 ensemble members) and a lead-time dependent bias calculated from the first 20 (1961-1980) and the last 20 (1982-2001) initializations, respectively. The recommendation of ICPO (2011) would be justified if both results exhibit no significant differences. However, model bias and its

evolution with lead time (results for winter 1, 5 and 9 shown in Fig. 3.3) substantially differ between the earlier and later initialization times, especially for shorter lead times. While the model bias over the Northeast (Northwest) Pacific is positive (negative) during the first hindcast winters of the first 20 initializations and evolves towards neutral (more negative) conditions over lead time, it starts neutral (more negative) and shows no clear (positive) trends over lead time if calculated from the last 20 initializations. Over the North Atlantic, a strongly positive (slightly negative) bias is obvious over the Mid-Latitudes (Subtropics) during the first winters but evolving negatively (positively) over lead time, when calculated from the first 20 initializations. Based on the last 20 initializations, the temporal evolution of the bias over the subtropical North Atlantic is generally similar to that analyzed from the first 20 initializations, while no remarkable drifts are found over the mid-latitudinal North Atlantic with a bias constant over lead time and slightly negative here. A two-sided $t$-test shows significant ($\alpha < 0.05$) differences between early and later initializations regarding winter 1 for most regions of the NH, including both stormtracks (not shown here). The corresponding differences for winter 9 are less pronounced, fewer regions exhibit significant disagreements. In fact, both bias patterns are generally similar to the climatological bias pattern of the uninitialized simulations (see Fig. 3.1), which confirms the general expectation of the initialized hindcasts to drift towards model climatology over lead-time.

We conclude, that we have to account for model drifts that are changing over time. Inspired by the study of Kharin et al. (2012), we chose a parametric approach to assess the hindcast drifts. We take up the suggestion of Gangstø et al. (2013) and use a third order polynomial instead of an exponential function as chosen by Kharin et al. (2012) to account for the drift along lead time $\tau$

$$\widehat{H}_{i,\tau,j} = H_{i,\tau,j} \; - \; a_0 - a_1\tau - a_2\tau^2 - a_3\tau^3. \tag{3.3}$$

To account for the non-stationarity of the model drifts, we allow the polynomial parameters $a_k, k = 0, \ldots, 3$ to change over time $t$, i.e. $a_k = a_k(t)$. The most simple model is a linear trend in time as suggested by Kharin et al. (2012), i,.e. $a_0 = b_0 + b_1 t$. This means, that for a certain initialization $i$ and lead time $\tau$ the corrected hindcast of the $j^{\text{th}}$ ensemble member is given by

$$\widehat{H}_{i,\tau,j} = H_{i,\tau,j} \; - \; (b_0 + b_1 t) - (b_2 + b_3 t)\tau - (b_4 + b_5 t)\tau^2 - (b_6 + b_7 t)\tau^3, \tag{3.4}$$

with $H_{i,\tau,j}$ being its uncorrected equivalent.

The parameters $b_0 .. b_7$ are estimated by the standard least-squares-method from the differences between all available hindcasts (the individual ensemble members) and the reanalysis valid for the respective time $t$, corresponding to the given initialization and lead time. Compared to the previously described estimation of a separate bias for all of the nine hindcast winters, we have here only eight parameters to estimate instead of nine. The model is thus more parsimonious.

Figure 3.4 shows the effect of this parametric drift correction in comparison to the non-parametric standard procedure (ICPO, 2011) exemplarily for the *baseline1*- and *ORAff*-hindcasts and the winter storm frequency, calculated for a grid point over the central North Atlantic (-30°E, 48.75°N). Already visible from the non-parametric estimation of a lead-time-dependent bias (Fig. 3.4(a) and (b)), a drift over lead-time is more obvious for *ORAff* than for the anomaly-initialized *baseline1* (for this grid point); meeting the expectations of more pronounced drifts of full-field-initialized systems. The bias estimations resulting from the parametric approach (Fig. 3.4(c) and (d)) now clarify how bias and drifts are changing over time. Closely related to the trend differences already seen in Fig. 3.2, bias estimations for winter 1 and *baseline1* (*ORAff*) range from above 6 (10) for the earliest initialization to below $-4$ (0) for the latest initialization. Again the *baseline1*-bias shows comparably little change over lead-time, that is no substantial drifts are evident, while the *ORAff*-bias decreases remarkably, though not completely reaching the level of *baseline1*. Interestingly, this means that the bias of the uninitialized simulations for the respective time (not shown here explicitly) is smaller, too. Hence, this result suggests (not tested for statistical significance) that for winter storm frequencies over the North Atlantic, a simulation of approximately nine years is not long enough to get completely rid of the initialization shock. Also note that the range of the bias over different initialization times (Fig. 3.4(c,d), different colors) is larger than its change over lead time.

A correction of the hindcasts $H_{i,\tau,j}$ with this parametric approach (Eq. (3.3)) effectively eliminates a) a climatological bias of the model, b) a long-term trend linear in time, and c) a potential cubic drift with parameters varying linearly in time. Consequently, for a fair comparison of these corrected initialized hindcasts with the results of the uninitialized model runs, we have to adjust the latter as well for a) their climatological bias and b) their deviating long-term trend. The third correction c) for the drift is not
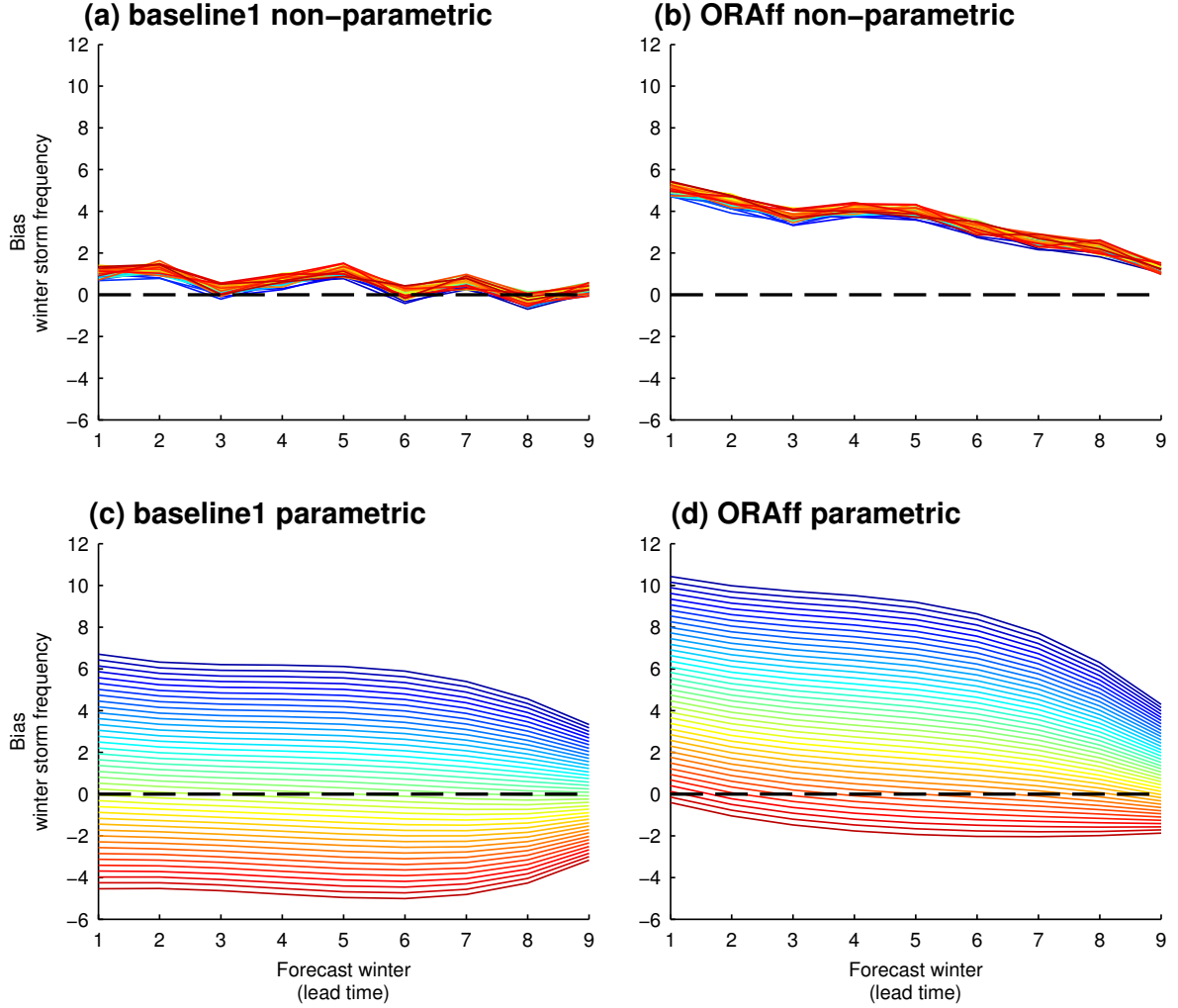
Figure 3.4.: Lead-time-dependent bias (ERA-reanalyses as reference) of the 41 *baseline1*-
(left) and *ORAff*-hindcasts (right) regarding winter storm frequency over the central
North Atlantic (-30°E, 48.75°N), estimated with standard non-parametric approach
in cross-validated manner (leaving out the respective hindcast, the bias is calculated
for; top) and parametric approach applied in this study (bottom); line color denoting
initialization time with blue for early hindcasts (starting 1961) and red for most recent
hindcasts (ending 2001)

necessary as the uninitialized simulations cannot have a drift by definition. Without any
drift related terms, the adjusted "forecast" of an uninitialized simulation $\widehat{U}_{t,j}$ is derived
from

$$\widehat{U}_{t,j} = U_{t,j} - (c_0 + c_1 t)$$

based on the unadjusted "forecast" $U_{t,j}$. The climatological bias parameter $c_0$ is exactly that depicted as color shadings in Fig. 3.1 and the linear trend parameter $c_1$ for the uninitialized simulations is the difference between Fig. 3.2(a) and (b).

We individually apply these corrections to the winter storm frequencies calculated for all grid points (of initialized and uninitialized simulations). Kharin et al. advised against using this approach for localized quantities as local trends derived from observations exhibit large uncertainties. They suggested a simple approach (for surface temperature) by correcting local model trends (quite robust if derived from ensembles) with the ratio of global trends from model and observations, which are more reliable for the latter than their local peculiarity. The current understanding of long-term climate change signals with respect to storm tracks and typical pathways of intense extra-tropical cyclones and winter storms (summarized in Hartmann et al., 2013; Christensen et al., 2013) is more characterized by shifts and local trends than by globally uniform changes. Hence, such an approach is not appropriate in the context of our study. Here, the winter storm frequencies are derived for a high degree of spatial aggregation (counting numbers of tracks within a radius of 1000km around the given grid-point) and thus these values are very different from, e.g., grid-box-wise temperatures. We thus state that the respective values derived from reanalyses are robust enough to be directly used for estimating the parameters of the time dependent cubic drift correction. This is supported by trend patterns calculated for the *Twentieth Century Reanalysis* (20CR, Compo et al., 2011) being in very good agreement with those of the ERA-reanalyses, only slightly weaker with respect to the maxima (not shown).

## 3.5. Decadal prediction skill

The multiple-system ensembles deliver more robust estimates of skill because of their larger ensemble sizes (compared to the individual hindcast experiments). Thus, we address our first two research questions regarding skill over climatological forecasts and the benefit from initialization by means of the *Grand Ensemble*.

To answer the general question, whether decadal predictions contain any valuable information about the winter storm frequency of the upcoming years, we compare the results of the *Grand Ensemble* to climatological forecasts. In the context of probabilistic
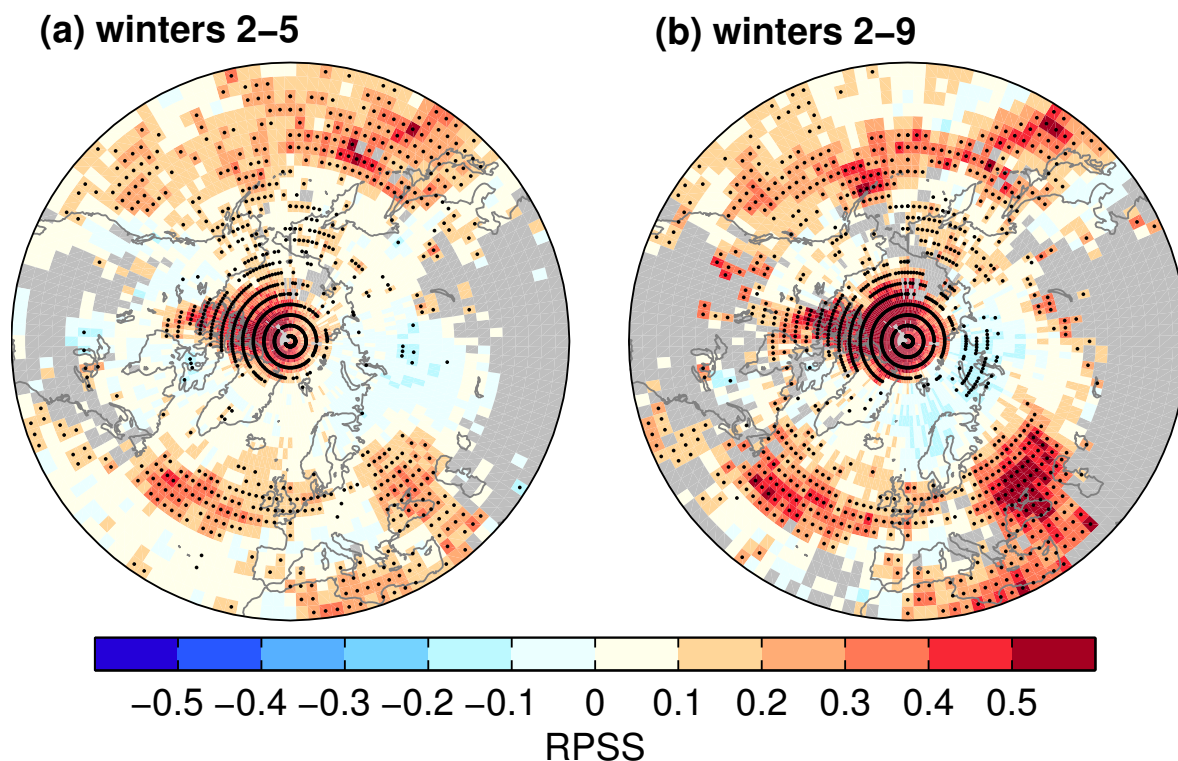
**(a) winters 2–5**  **(b) winters 2–9**



Figure 3.5.: **RPSS of *Grand Ensemble* over climatological forecasts** regarding the average winter storm frequency (number of tracks per ONDJFM in the vicinity of 1000km) for (a) hindcast winters 2–5 and (b) hindcast winters 2–9 based on ERA-reanalyses as observational reference; significant skill scores ($\alpha < 5\%$) as black dots, areas of strong inconsistencies between ERA40 and ERA-Int. are masked out (grey)

prediction of three discrete classes the reference forecasts are climatological category probabilities ($\frac{1}{3}$, $\frac{2}{3}$, and 1) which hold for all hindcasts and lead times. The forecasts of average winter storm frequency for winters 2–5 (Fig. 3.5(a)) exhibit significantly positive skill for the entire Pacific basin, the NH polar latitudes, as well as the mid-latitudinal North Atlantic and a region over the Mediterranean and the Black Sea. For winters 2–9 (Fig. 3.5(b)) skill scores are even higher for most parts of the NH. The only region contiguously exhibiting zero or slightly negative skill scores is the Atlantic sector of the Arctic ocean and adjacent land areas, such as Greenland, Scandinavia and parts of Russia. Over the North Pacific, skill seems to be concentrated in the mid-latitudes. Skill over the central sub-tropical North Pacific is slightly lower for winters 2–9 than for the winter 2–5 forecasts.

The general result from these analyses is very encouraging. However, two major issues have to be stated: First, the initialized decadal predictions (and the uninitialized runs) were adjusted to match the linear trend of the observations. To make sure that these positive findings are not purely a result from the statistical adjustment, we calculated the skill of the raw (not trend-adjusted) uninitialized simulations (4yr- and 8yr-running-means being the equivalents to the 2–5yr- and 2–9yr-forecasts) over climatological forecasts. We found that these raw uninitialized runs do exhibit skill patterns (not shown) generally similar to those of the initialized hindcasts as in Fig. 3.5. While the magnitudes of these skill scores for the 4yr-running-mean is overall considerably lower than those of the initialized 2–5yr-forecasts (Fig. 3.5(a)), RPSS-magnitudes for the 8yr-running-mean are comparable to those of the initialized 2–9yr-forecasts (Fig. 3.5(b)). Thus, the statistical adjustment is not the major source of skill in this context but the skill of the uninitialized simulations emphasizes the second issue regarding the interpretation of the results shown in Fig. 3.5: they are not sufficient to prove the success of initialized decadal predictions. At least part of the skill compared to a climatological forecast seems to arise from the response to external forcing, already included in the uninitialized simulations.

Hence, to answer the question about the added value of initialization, skill scores were additionally calculated with the (trend-adjusted) uninitialized runs as reference forecast. Fig. 3.6(a) shows the results for the winter 2–5 forecast of the *Grand Ensemble*. Obviously, only for some regions the initialized hindcasts are able to provide significantly added value. The most prominent example is the entrance of the North Pacific storm track over Eastern Asia and the Northwest Pacific. Similarly but less pronounced and coherent, decadal predictions for winter storm frequencies at the entrance of the North Atlantic storm track along the North-American east coast seem to profit from initialization. The only other region where significantly positive skill scores can be diagnosed over a larger area is the American sector of the Arctic Ocean. These results are generally in line with those for intense cyclones investigated in the study of Kruschke et al. (2014), as in most cases we find areas of significantly positive skill for winter storm frequency south of areas exhibiting skill with respect to intense cyclone frequency (Kruschke et al., 2014, Fig. 3c and 5c). This matches our expectation, as the winter storm tracks, identified with the here-applied method, are usually found south of the related extra-tropical cyclone track. Skill of winter 2–5 predictions over the North Pacific storm
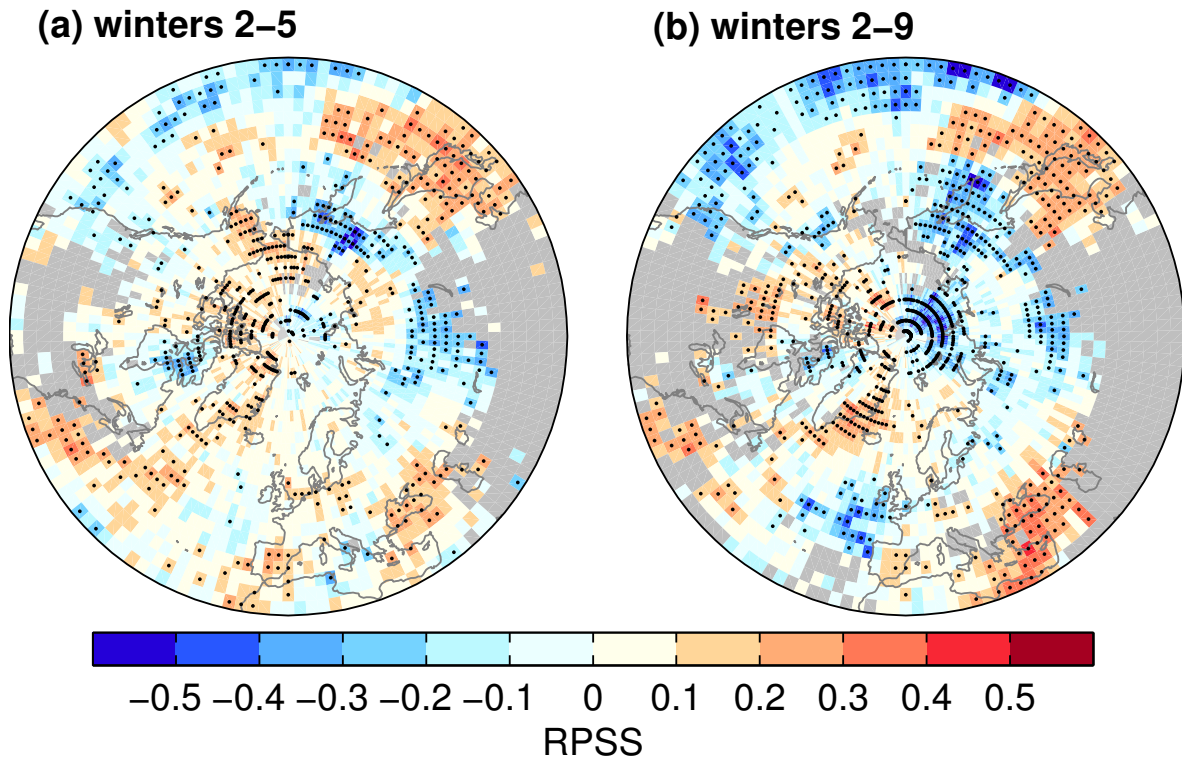
**(a) winters 2–5**　　　　　**(b) winters 2–9**



Figure 3.6.: **RPSS of *Grand Ensemble* over uninitialized simulations** regarding the average winter storm frequency (number of tracks per ONDJFM in the vicinity of 1000km) for (a) hindcast winters 2–5 and (b) hindcast winters 2–9 based on ERA-reanalyses as observational reference; significant skill scores ($\alpha < 5\%$) as black dots, areas of strong inconsistencies between ERA40 and ERA-Int. are masked out (grey)

track is, however, smaller than values found by Kruschke et al.. This may be – at least partially – explained by the different observational references used, as Kruschke et al., found a strong influence on the skill from the specific reanalysis dataset used. They state that analogous analyses regarding cyclone frequencies based on NCEP1-reanalysis (Kalnay et al., 1996) or a mix of ERA40 and ERA-Interim, as done in this paper, wipe out the skill they found in this area, using 20CR as observational reference. Considering predictions of the average winter storm frequency of winters 2–9 (Fig. 3.6(b)), only the positive skill pattern of the winter 2–5 forecast over the North-West Pacific prevails. On the other hand, an area over the Eastern Mediterranean and the Black Sea is marked by significantly positive skill scores for these forecast horizons. For all other regions of the NH, that is the subtropical North Pacific, the North-East Atlantic, large parts of

Eurasia and the adjacent Arctic Ocean, our initialized decadal predictions are not able to provide skill over the uninitialized transient simulations.
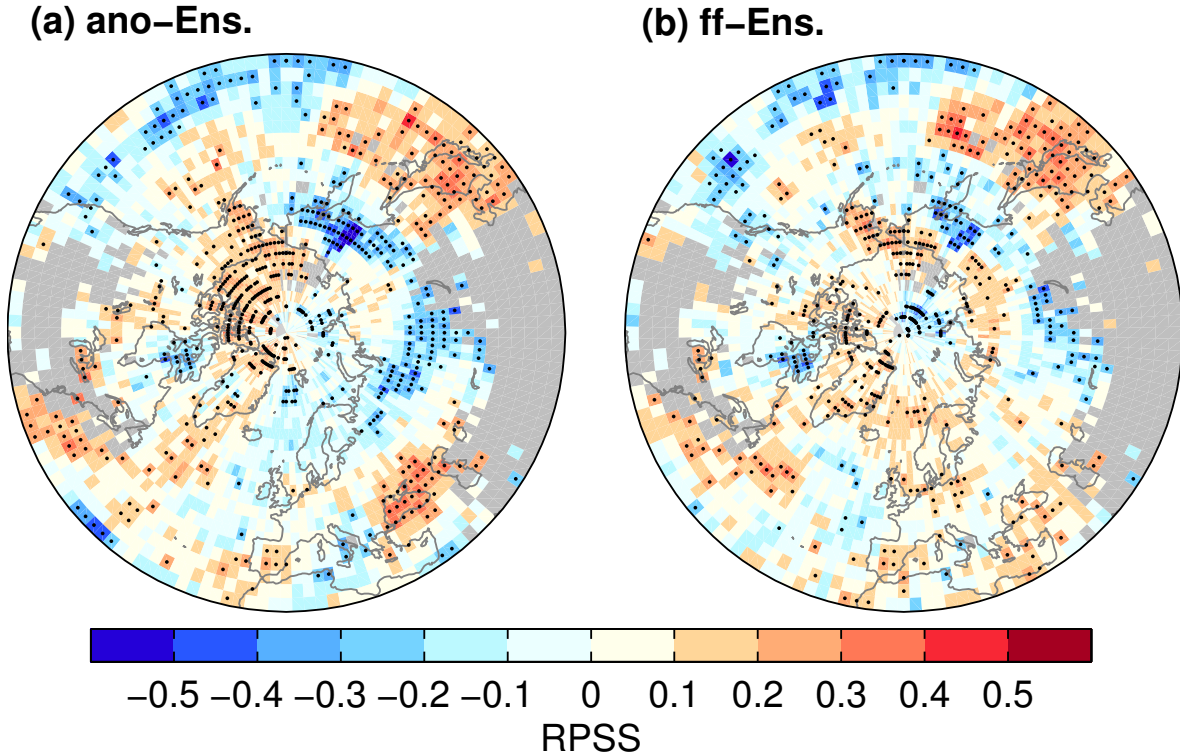


Figure 3.7.: **RPSS of (a) anomaly-initialized hindcasts (*ano-Ens.*) and (b) full-field-initialized hindcasts (*ff-Ens.*) over uninitialized simulations** regarding the average winter storm frequency (number of tracks per ONDJFM in the vicinity of 1000km) for hindcast winters 2–5 based on ERA-reanalyses as observational reference; significant skill scores ($\alpha < 5\%$) as black dots, areas of strong inconsistencies between ERA40 and ERA-Int. are masked out (grey)

To systematically evaluate the different initialization strategies followed so far, we compare the multiple system ensembles *ano-Ens.* and *ff-Ens.* as well as *ORA-Ens.* and *GECCO-Ens.* (only skill over uninitialized simulations for winter 2–5 hindcasts shown in Fig. 3.7 and 3.8, respectively). Generally, the differences with respect to the calculated skill scores between initializing from oceanic full-fields or anomalies are rather small. *ff-Ens.* seems to perform slightly better than *ano-Ens.*, though not significant, in predicting winter storm frequencies over the Northwest Pacific, while anomaly-initialization yields better results over the Arctic Ocean north of America and around the Black Sea. Most remarkable about the comparison of *ORA-Ens.* and *GECCO-Ens.* are the better

predictions of the latter regarding winter storm frequencies over the central and western North Pacific. The same seems to be the case over the Northwest Atlantic, however, hardly significant. Both, *ORA-Ens.* and *GECCO-Ens.* perform well over the American Arctic, but the area of significant skill is more coherent for *ORA-Ens.*. Overall, none of the initialization strategies is clearly superior to the others.
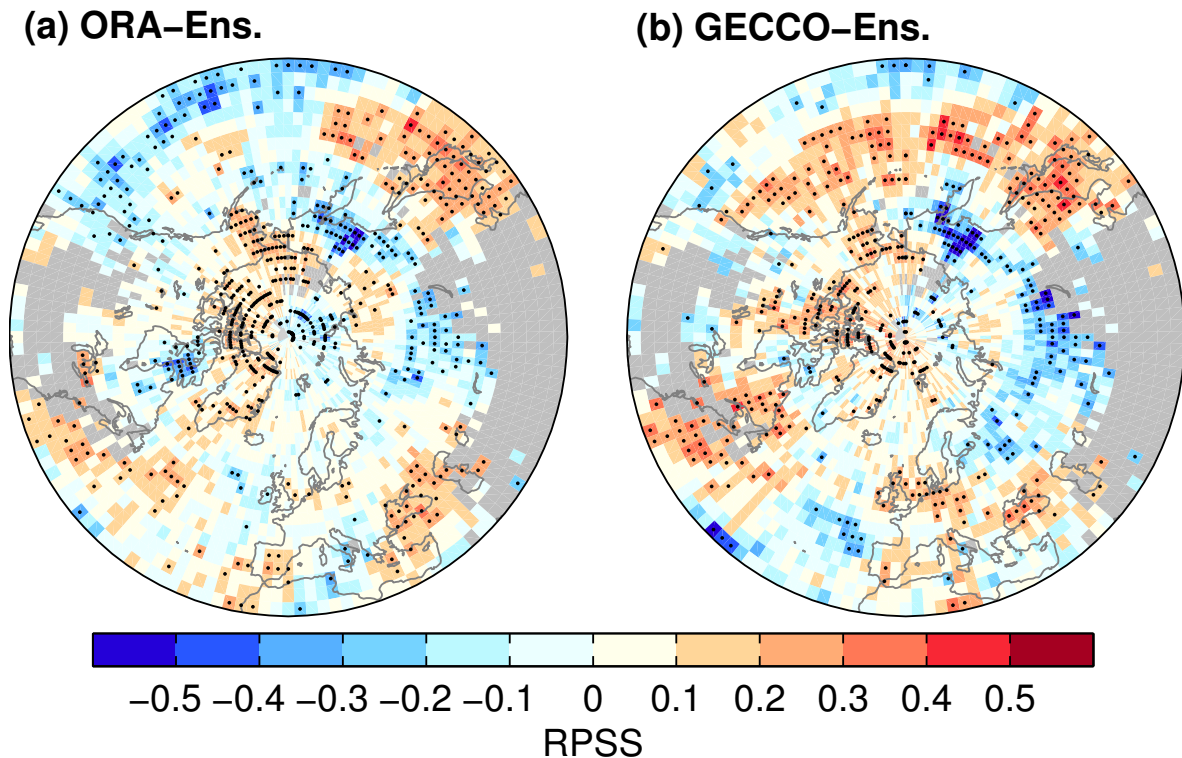


Figure 3.8.: **RPSS of hindcats initialized from (a) ORA-S4- (*ORA-Ens.*) and (b) GECCO2-ocean-reanalysis (*GECCO-Ens.*) over uninitialized simulations** regarding the average winter storm frequency (number of tracks per ONDJFM in the vicinity of 1000km) for hindcast winters 2–5 based on ERA-reanalyses as observational reference; significant skill scores ($\alpha < 5\%$) as black/white dots, areas of strong inconsistencies between ERA40 and ERA-Int. are masked out (grey)

Inspired by the study of Kharin et al. (2012), we chose a parametric approach to assess potential model drifts of the initialized hindcasts. The effect of this approach is exemplarily assessed for the winter 2–5 forecasts of the *ORAff* hindcast experiment (Fig. 3.9, skill over uninitialized simulations). The standard non-parametric procedure in this context, recommended by ICPO (2011), is to calculate a model drift principally constant over all initializations by means of a lead-time-dependent bias (implicitly also correcting

a climatological bias) and subsequently to assess the skill over uninitialized simulations that were corrected for their climatological bias (Fig. 3.9(a)). Our parametric approach estimates potential hindcast drifts as cubic polynomials with parameters varying linearly in time (implicitly also adjusting the linear long-term trend and the climatological bias). Subsequently, the skill is assessed over uninitialized simulations that were adjusted for their linear long-term trend and climatological bias (Fig. 3.9(b)). Hence, differences between Fig. 3.9(a) and (b) are due to the more or less appropriate approach regarding drift assessment only. These differences clearly show a generally higher skill of *ORAff*, when adjusted with the parametric drift-correction approach, while the overall pattern is not changed and local skill maxima/minima stay the same.
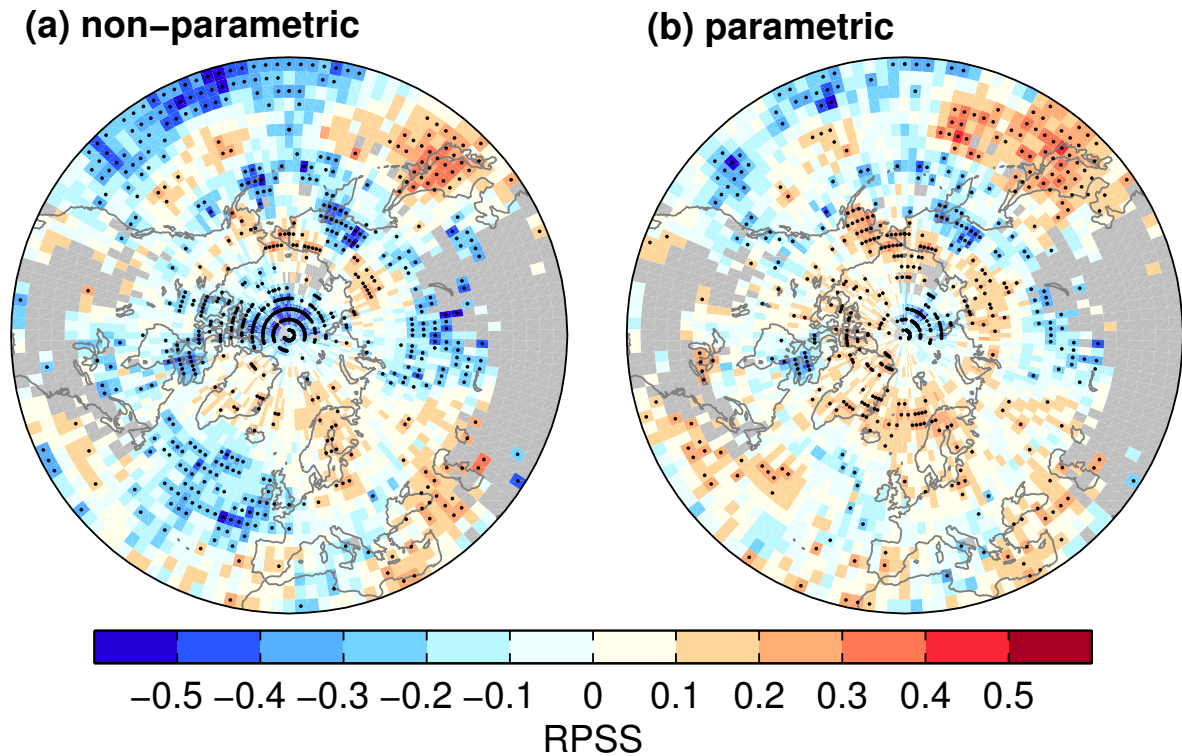


**(a) non–parametric**          **(b) parametric**

Figure 3.9.: **RPSS of *ORAff* over uninitialized simulations after (a) non-parametric and (b) parametric drift correction** regarding the average winter storm frequency (number of tracks per ONDJFM in the vicinity of 1000km) for hindcast winters 2–5 based on ERA-reanalyses as observational reference; significant skill scores ($\alpha < 5\%$) as black/white dots, areas of strong inconsistencies between ERA40 and ERA-Int. are masked out (grey)

The results of the other individual hindcast experiments can be found in Appendix A.1.

# 3.6. Summary and discussion

Basically five sets of decadal hindcasts produced within the MiKlip initiative are analyzed with respect to the skill of probabilistic three-category predictions regarding winter storm frequencies over the extra-tropical NH. Multiple-system ensembles were constructed by specific combinations of the original hindcast experiments to provide robust skill assessments (due to large ensemble sizes) and to permit systematic comparisons of different initialization strategies pursued so far.

It is shown that predictions of average winter storm frequency of winters 2–5 as well as winters 2–9 do exhibit significant skill (i.e. better than assuming climatological probabilities for each category and initialization; Fig. 3.5) for large parts of the extra-tropical NH, that is the whole North Pacific, the mid-latitudinal North Atlantic, the American sector of the Arctic and a region over the Eastern Mediterranean and the Black Sea.

However, a comparison to uninitialized (transient) simulations (Fig. 3.6) reveals that a substantial part of this skill is attributed to long-term trends, associated with imposed greenhouse gas and aerosol forcing. Hence, the additional value of initialization effort is restricted to smaller areas. These are the entrance regions of both NH storm tracks along the North-American East Coast and especially over East Asia and the Kuroshio Extension. Additionally, predictions for parts of the Arctic and the already-mentioned area over the Eastern Mediterranean and Black Sea profit from initialization. Regarding the latter region, skill for winters 2–9 is obviously higher than for winters 2–5.

A systematic comparison of hindcasts produced by anomaly- vs. full-field-initialization (Fig. 3.7) and initialized from ORA-S4- vs. GECCO2-ocean-reanalysis (Fig. 3.8) yields no clearly superior initialization strategy. The only remarkable difference seems to be existent over the North Pacific, where *GECCO-Ens.* provides higher predictive skill than *ORA-Ens.* for both winter 2–5 and 2–9 forecasts (the latter not shown).

The skill of decadal predictions regarding winter storm frequency in the NH storm track regions is generally lower than that regarding (intense) cyclone frequency as presented in the study of Kruschke et al. (2014). This is not caused by methodological differences since repeating the cyclone-related analyses of Kruschke et al. with the methods presented here, that is the adjustment of bias, long-term-trend and drift according to Sec. 3.4, yields even higher skill of the initialized predictions over uninitialized simulations

than presented in their study. As the vast majority of winter storms identified with the scheme applied here can be related to intense extra-tropical cyclones (not shown), it is basically not about completely different phenomena considered but about a specific subset of events. Kruschke et al. (2014) found promising results especially for intense cyclones´ frequencies. Their definition of such events led to climatological frequencies in the storm track regions that are approx. twice as high as the climatological frequencies of winter storms (see Fig. 3.1) in these regions. This means that we are considering more extreme events in these regions. We assume that the inter-annual to decadal variability of this subset exhibits lower signal-to-noise-ratios (beyond externally forced trends) for large parts of the storm tracks than the less extreme subset of Kruschke et al. (2014), resulting in lower predictability and hence prediction skill.

Comparison of skill assessments after using the standard drift-correction procedure, applied in Kruschke et al. (2014; recommended by ICPO, 2011) and the here-applied parametric method, inspired by Kharin et al. (2012) (Fig. 3.9), confirms the expectation that the latter is more adequate for estimating and subsequently eliminating model drifts. This leads to better skill assessments. However, a constraint of this approach is the inherent adjustment of linear long-term trends. First, such linear long-term trends will not be suitable for each parameter, region and period. Second, it surely is disputable to which degree (or lead time) decadal predictions are reliable if the underlying model exhibits deviating long-term trends due to external forcing, although we would like to point out, that the approach used in this study, effectively is very similar to the widely used detrending in this respect. More research is needed to come up with optimal solutions regarding drift correction and the understanding of model biases.

A major issue for a proper assessment of prediction skill in a "grid-point-wise manner" is a systematic misrepresentation of relevant features. The too zonal orientation of the North-Atlantic storm track and the northward shift of the North-Pacific storm track in the model (already diagnosed by Kruschke et al., 2014) will result in a deflation of prediction skill, as temporal variability of these features, even if perfectly forecasted, will produce related signals, e.g. winter storm frequencies, for locations differing between model and reality. As long as such systematic storm track deformations are existent in model simulations, approaches accounting for this feature would be more appropriate.

Not neglecting these constraints, we nevertheless consider the results found as being encouraging. It is shown that initialized decadal predictions do provide potentially

valuable information for several NH regions that go beyond externally forced long-term changes. Given the yet early stage of decadal prediction research and ongoing activities regarding model developments and improved initialization as well as adequate statistical post-processing and verification methods the prospects of furthermore improved decadal predictions regarding the frequency of winter storms and potentially related socio-economic value can be judged favorably.

# 4. Statistical regionalization of surface gusts within European winter storms

The work presented in this Chapter was done in the context of a research project funded by Munich RE (MR) and supported by the German Weather Service (DWD). The scientific progress presented here is naturally influenced by collaborations within this project. The dynamical regionalizations that form the basis of training and validating the statistical regionalization were conducted by Philip Lorenz and Robert Osinski. However, the objective procedure selecting which storm events to simulate was developed by the author of this thesis. The core of this Chapter, that is the statistical downscaling approach based on Multiple Linear Regressions and all analyses with respect to its validation are also an achievement of this thesis' author. The shortly mentioned alternative statistical downscaling approach, based on neural networks, was developed and tested by Maximilian Voigt as part of his Master's thesis. All intermediate and final results were discussed with the project's principal investigators Prof. Dr. Uwe Ulbrich and PD Dr. Gregor Leckebusch as well as the further project partners at MR and DWD. Hence, they certainly contributed to the overall advance of this study. In particular, Dr. Henning Rust gave the motivation to investigate the approaches of *stepwise linear regressions* and *quantile regressions* for their suitability in the context of this study. A manuscript for peer-reviewed publication based on this Chapter is currently in preparation.

## 4.1. Introduction

Winter wind storms are the most relevant natural hazard for the European continent (Munich RE Group, 2008a; Handmer et al., 2012), fortunately associated with rather few fatalities but tremendous economic losses ($\sim$25.2 bn€ overall, $\sim$12.8 bn€ insured losses in 1980–2006 according to Munich RE Group, 2007). These numbers are related to the fact that European winter storms are synoptic-scale events (related to extra-tropical cyclones), usually affecting several countries at the same time. Thus, they are associated with high risk of cumulative loss (Munich RE Group, 2007). Hence, they are of great relevance to the re-insurance industry. With respect to reasonable financial planning and legislative requirements such as the European *Solvency II* directive, re-insurance companies are particularly interested in probabilities and return periods of very rare events (Della-Marta et al., 2010). Given the comparably short period covered by sufficient meteorological observations, only a limited number of historic storm events is available in this respect. Many companies rely on Monte-Carlo-like sampling techniques (Dukes and Palutikof, 1995; Palutikof et al., 1999) to extend the observed events to a voluminous data set (numerically) appropriate for statistical studies on very rare events (Schwierz et al., 2010; Haylock, 2011). A different approach is to exploit comprehensive climate model simulations. Besides estimating impacts of potential climate change (e.g. Leckebusch et al., 2007; Rockel and Woth, 2007; Pinto et al., 2007; Donat et al., 2011a; Pinto et al., 2012; Held et al., 2013) such an approach is motivated by the expectation of physically more consistent results (e.g. Della-Marta et al., 2010; Haylock, 2011) eventually beneficial to loss modeling and related estimations of return periods. Given that standard insurance loss models require wind fields of high spatial resolution (Haylock, 2011), regional climate models (RCM) are employed for several studies (Schwierz et al., 2010; Donat et al., 2011a; Haylock, 2011; Held et al., 2013; Roberts et al., 2014). However, the applicability of dynamical downscaling methods is confined by its computational demands. Hence, in order to generate extensive data sets of spatially high resolution wind fields efficient statistical downscaling approaches are necessary. Purely statistical regionalization procedures generally relate some coarse scale predictors to the high resolution wind data. Such a procedure is presented by de Rooy and Kok (2004), downscaling RCM output to observing stations by distinguishing between large-scale model error and representation mismatch. Pryor et al. (2005) empirically derived a relationship between probabilistic surface wind distributions at various stations to larger

scale sea-level pressure gradients and relative vorticity (at 500 hPa) in a climatological context. The method presented by Orlowsky et al. (2008, application for wind speed and related economic loss shown by Held et al., 2013) consists of re-sampling measurements conditional to analogs of some explanatory variable. Equally applicable for climatological analyses are analog methods and clustering approaches (e.g. Leckebusch et al., 2008b), connecting some dependent variable to large-scale circulation patterns. Such approaches can also be related to a set of dynamical regionalizations representing the specific clusters (e.g. Pinto et al., 2010; Reyers et al., 2015).

Once, high temporal resolution is considered, few studies remain, targeting this issue. Salameh et al. (2009) statistically modeled six-hourly, daily, and weekly averages of wind components observed at several stations in southern France following a Generalized Additive Model approach based on surface winds (925, 850, and 700 hPa), surface pressure gradients, as well as geopotential height and relative vorticity at 500 hPa as large scale predictors. Haylock (2011) employed a combination of upward vertical extrapolation, horizontal interpolation, downward interpolation and an empirical gust model for estimating daily maximum gusts on a 7km-grid from coarser scale RCM integrations.

The objective of the current study is to develop a statistical downscaling procedure suitable for generating a comprehensive winter storm event set based on reanalyses (both, ERA40 – see Uppala et al., 2005 – and ERA-Interim – see Dee et al., 2011) and medium-range (10 days) weather forecasts produced by the *Ensemble Prediction System* (EPS, Molteni et al., 1996) of the *European Centre for Medium-Range Weather Forecasts* (ECMWF). This event set, containing multiple variations of past winter wind storms – each as physically consistent as the underlying model – is planned to be subsequently used for re-insurance loss modeling. For this purpose, high spatial (7 km) and temporal resolution (six-hourly) is targeted while covering the whole European continent. The basic concept is to perform dynamical regionalizations for a limited number (181) of selected storm episodes which are utilized subsequently to train and validate a statistical downscaling model eventually applicable to the totality of all storm events diagnosed in the reanalyses (∼3.700) and EPS data (∼280.000). The selection procedure and a description of the dynamical regionalization is found in Sec. 4.2.

In the face of the nature of this task – deriving a robust transfer function applicable to a wide range of wind storm situations represented in comparably coarse resolution based on a very limited extent of training data – the approaches of Salameh et al. (2009, because

of the predictands spread over a wide area) and Haylock (2011, because of the predictors being of considerably coarser scale) seem to be not suitable. Instead, a Multiple Linear Regression approach is used and thoroughly described in Sec. 4.3. Concurrently with the preparation of this thesis, Haas and Pinto (2012) conducted a study targeted at equivalent objectives, yielding a similar solution regarding the statistical downscaling. In this respect, an approach comparable to that of Haas and Pinto (2012) serves as a benchmark in Sec. 4.4 containing the validation of the method, including systematic comparisons of different (potential) predictors. The current Chapter of this thesis is completed by a summary, presented in Sec. 4.6.

## 4.2. Data

### 4.2.1. Impact-oriented compilation of training and validation data

The concept, to use a limited number of dynamically downscaled storm episodes for training and validation of a statistical downscaling procedure requires the basic selection to sample a subset which is representative for the totality of storm events the statistical downscaling is subsequently applied to. Representativity in this respect demands for spatial coverage of the whole analysis domain (Europe) and most intense storms (regarding the target variable, that is surface gusts) for each region to be part of the selection. The latter requirement is to ensure optimal suitability of the statistical downscaling function for highest wind speeds (potential lower quality for lower wind speeds is considered acceptable).

Thus, an objective selection of European winter storms based on the ESSI is performed by applying the WiTRACK-scheme, presented in Ch. 2 (see final paragraph of Sec. 2.3.3 for information on WiTRACK-features implemented in revision r105 which is used here) to the extended winter seasons (SONDJFMAM) of *ECMWF*'s reanalyses and the EPS forecasts. The EPS was introduced in December 1992. However, for this study we consider only forecasts, at least 6-hourly data is archived for. Hence, the EPS-related analyses are restricted to the period of January 2000 until January 2010[1]. The more

---

[1]A complementary analysis including training, validation, and application of the statistical downscaling procedure is conducted for the 12-hourly data. However, as the development of the statistical downscaling approach is primarily based on 6-hourly data, and the quality of the downscaling results

Table 4.1.: WiTRACK-configuration used for sectoral tracking of winter wind storms over Europe and the North Atlantic from ERA reanalyses and EPS-forecasts

| | |
|---|---|
| period covered: | ext. winters (SONDJFMAM) 1957/58–2001/02 (ERA40), 1979/80–2010/11 (ERA-Int.), and 2000/01–2010/11 (EPS) |
| domain: | Europe and North Atlantic (40°W–40°E, 25°N–80°N) |
| boundary sponge zone: | 35°W–40°W; 30°E–40°E; 25°N–35°N; 75°N–80°N |
| considered variable: | 6-hourly instantaneous 10m-wind speed |
| threshold: | empirical 98$^{th}$ percentile of 10m-winds, calculated from reference period 1989–2010 (all time steps) for ERA-Int., percentiles of other data sets climatologically scaled (see explanation in Tab. 3.3 and Osinski, 2014) |
| secondary thresh.: | empirical 95$^{th}$ percentile of 10m-winds, derived analogously to primary threshold |
| minimum size: | 160,000 km$^2$ |
| grid box weighting during tracking: | area multiplied by third power of relative threshold exceedances (SSI-contribution, see Eq. 2.1) |
| maximum translation distance: | 720 km |
| factor for max. internal relocation: | 0.5 |
| minimum lifetime: | 5 time steps $\sim$ 24 hours |
| grid box weighting for SSI-calculation: | standard SSI-contribution (see Eq. 2.1, equivalent to tracking) |

recent EPS predictions are archived in 3-hourly resolution for the first six forecast days, but in 6-hourly time steps for day 7–10. As the full 10-day forecasts should be used in a consistent manner, these additional time steps of the first six days are neglected. A thorough analysis of the storm events and associated properties identified from ERA-Interim and these EPS predictions can be found in the thesis of Osinski (2014) and the related paper of Osinski et al. (2015). Tab. 4.1 contains a short summary of the WiTRACK-configuration used for these analyses.

In the course of this WiTRACK-application the option of calculating Region-ESSIs – considering only grid boxes in certain domains – is utilized (see paragraph on "mask and multiplicator files..." in Sec. 2.3.3). Nine different regions are defined following geographical and socio-economic criteria as well as considerations of typical tracks of

---

for 12-hourly data is found to be significantly worse, the respective data set is neglected here.

extra-tropical cyclones and winter storms (Fig. 4.1). For each of these nine regions, the winter storm events identified from reanalyses are ranked according to their Region-ESSIs[2] and the most severe events are selected for dynamical downscaling. The number of storms selected is set differently for the individual regions, taking into account their climatological storm frequency but also their relevance for the portfolio of *Munich RE*. Eventually, the 30 most severe events for the GER-, FRA-, and UK-region are selected, respectively. The 24 most extreme storms of the SKA1-region, 18 for the IBER-region, as well as 15 for the SKA2-, BALT-, and ALP-region are chosen, respectively. The historical winter storm sample is completed by the 12 most severe events of the MEDI-region. Rectifying this sample for events that are selected based on more than one region yields a total number of 130 historical winter wind storms, subsequently downscaled dynamically for training and validating a statistical transfer function.

The selection of historical winter storms is complemented by "potential" winter storms, that is events diagnosed from EPS-forecasts representing more or less significant modifications of the synoptic situation observed in reality (and reanalyses). Different to reanalysis data, three-dimensional fields of the individual (perturbed) members of past EPS-forecasts are not completely available anymore. However, such fields are required for initializing the dynamical regionalization. Given that for the period since June 2005 at least the initial conditions for the EPS-members are archived, the selection procedure for "potential" winter storms is confined to the time between June 2005 and December 2010. As the dynamical downscaling is set up in free-running (global) hindcast mode based on completely different models (see Sec. 4.2.2) than the EPS (based on the *Integrated Forecasting System*, IFS), substantial deviations between these hindcasts and the original EPS-forecasts can be expected: the longer the simulation time, the larger the discrepancies. Hence, the selection procedure is additionally constrained to EPS-storms exhibiting their maximum intensity (in terms of ESSI-contributions of individual time steps) within the first two days after initialization. It is assumed that hindcasts of these events (primarily dependent on the initial conditions over that lead time) are in good agreement to the original EPS-forecasts and, hence, are beneficial for the development of a statistical downscaling method. Facing these restrictions, in particular the short period, less "potential" storms (one third of the numbers presented above) are

---

[2]For storm events occurring in the overlapping period of ERA40 and ERA-Interim, a spatio-temporal matching (described by Osinski, 2014) is performed, followed by an averaging of Region-ESSIs derived from the two datasets. These average Region-ESSIs are used for ranking and selection.
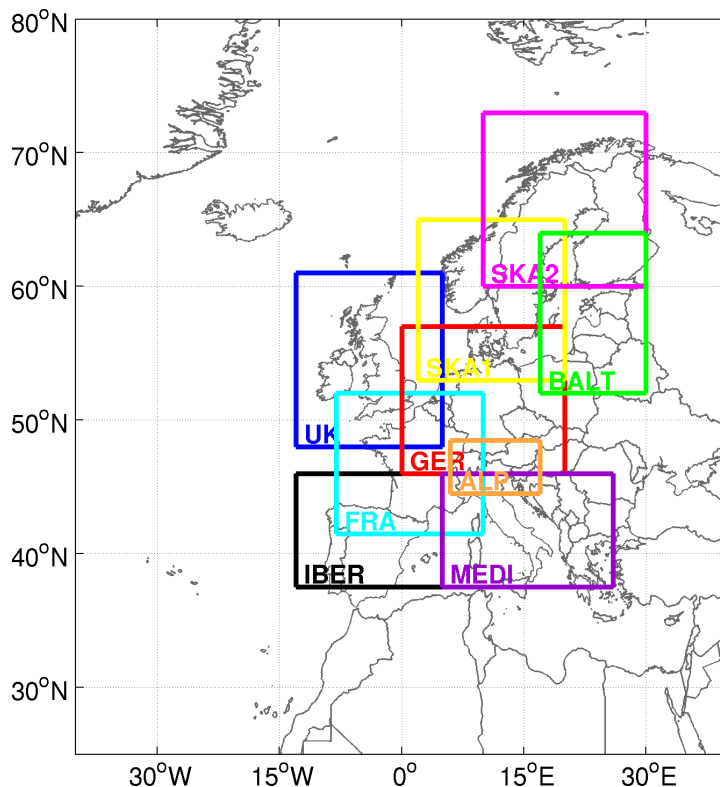
Figure 4.1.: Regions defined for calculation of Region-ESSIs, used for selecting storm
events, subsequently dynamically downscaled and utilized for training and validating
statistical downscaling

selected compared to the historical events. This yields a total number (after adjusting
for multiply selected events) of 51 "potential" winter wind storms that are dynamically
downscaled in the following.

With respect to the statistical transfer function, the spatial resolution of the predic-
tands ($\sim$7 km) is harmonized through dynamical regionalization (see Sec. 4.2.2). The
spatial resolution of any predictors – originating from ERA40 ($\sim$1.125°), ERA-Interim
($\sim$0.703°), or the EPS ($\sim$1.125°, $\sim$0.703°, and $\sim$0.45° for different cycles of the under-
lying IFS) – differs. Hence, a statistical transfer function, valid for all input data sets,
requires some spatial interpolation of the predictor fields. A first-order conservative
remapping (performed by means of the *cdo*s) procedure is used to interpolate all finer
grids to the coarsest grid considered here, that is the $\sim$1.125°-resolution of ERA40 and
the EPS forecasts for the period of January to November 2000. It is assumed that this
procedures introduces smaller errors than any interpolation from coarse to finer grids.

## 4.2.2. Dynamical regionalizations

The dynamical regionalization of the totality of 181 winter wind storms is conducted, employing the numerical weather prediction models of the German Weather Service (*Deutscher Wetterdienst*, DWD), that is the general circulation model *GME* (Majewski et al., 2002) driving the regional model *COSMO-EU* (Steppeler et al., 2002). With respect to the GME, model version 2.25 is used with a horizontal resolution of approx. 30 km and 60 layers. The non-hydrostatic COSMO-EU is run in model version 4.14, featuring a horizontal resolution of about 7 km and 40 layers and a dynamical core based on a two-level time-splitting integration following the *Runge-Kutta* scheme.

Wind gusts result from turbulent processes in the planetary boundary layer and, hence, are subject to various sub-grid scale influences. Thus, the estimation of gusts is not trivial and several approaches exist in this respect, most of them of statistical or empirical origin (Brasseur, 2001). The COSMO-EU model in version 4.14 contained a switch to choose between two different approaches, one being the estimate operationally used by DWD (Schulz, 2008) and the other developed by Brasseur (2001). Comparison to observations shows that both estimates yield useful results, depending on meteorological situation and the area considered (not shown here). Hence, the model code was adapted for the current study in order to calculate (and output) both estimates of turbulent surface gusts within any simulation (additionally, an estimate of convective gusts is calculated but neglected in the context of large scale winter storms for this study). The corresponding results of both gust estimation methods are consequently also used for training (and validating) statistical downscaling procedures.

All storm simulations are initialized shortly before the respective system made "landfall" over Europe. The global model GME is directly initialized with the (interpolated) fields of the respective reanalysis or (perturbed) EPS forecast. The simulation time for each episode is determined by the lifetime of the respective event. For the majority of storms this is two to three days. However, as already mentioned in Sec. 4.2.1 these hindcasts are completely free (within limitations of physical consistency represented by the model's equations) to develop differently than the according situations in the reanalyses and EPS forecasts, respectively. Such different evolutions introduce additional uncertainty to the relationships between any coarse scale predictors and high resolution gusts, quantitatively assessed by the statistical transfer function (yet to be found, see

Sec. 4.3 and 4.4). Hence, only the first two simulation days (precisely the first 51 hours, see Sec. 4.3) are used for training and validating the statistical downscaling, assuming that inconsistencies between predictors and predictands are limited to a tolerable level over this lead time. A thorough cost-benefit analysis with respect to the value offered by the predictands for different lead times would be desirable. However, given the computational demands of the statistical model training such an effort is omitted.

Finally, the differing temporal resolution has to be addressed. All predictor data sets considered are available in six-hourly resolution while the COSMO-EU-output hourly provides the maximum surface gust of the respective previous hour. Again, an aggregation to lower resolution seems to be less error-prone. Thus, maximum gusts of six-hourly time windows are calculated from the hourly COSMO-EU data. These time windows are centered over the respective predictor time steps to ensure maximum representativity of any predictors (instantaneously; see Se. 4.3) stored for a time step at half-time of these aggregation windows. That means for example a maximum of the COSMO-EU gusts stored at 10UTC, 11UTC, 12UTC, 13UTC, 14UTC, and 15UTC (each representing a maximum of the previous hour) which is related to coarse scale predictors stored as instantaneous values at 12UTC.

## 4.3. Methods

### 4.3.1. General remarks

In the course of this study, several approaches for statistically downscaling surface wind gusts have been tested. The thesis of Voigt (2012) depicts a neural network approach based on sea-level pressure (SLP) fields as coarse scale predictors. An alternative approach partly resembles the method of Salameh et al. (2009), conducting a *Principal Component Analysis* (PCA) again based on SLP and subsequently using the PC-timeseries as predictors for a Multiple Linear Regression. However, both approaches are obviously not competitive (comparative results not shown here) compared to the third alternative: Multiple Linear Regression, directly based on coarse scale grid box values. For the latter, SLP-gradients, 10m-wind speeds, and 10m-wind components were tested as explanatory variables. Results based on the former were found to be useful but far from optimal (not shown either). Regression models based on coarse scale wind speeds

and components perform significantly better. Hence, these are chosen to be the basis of the statistical downscaling presented here. In order to reduce potential inhomogeneities between the considered predictor data sets, all wind speeds and components are scaled by a characteristic parameter derived from the respective climatological distribution. In principle, any quantile, the mean value, or similar parameters would have been of equal suitability. Simply because the 98th percentile has been computed already (including the climatological scaling, explained in Tab. 3.3 and Osinski, 2014) it is this quantile used here, as well. Equivalent scalings are performed for both wind components – also used as predictors – and before the spatial interpolation mentioned in Sec. 4.3.1 is done.

The following Sec. 4.3.2 contains a description of the general method chosen in order to statistically model the high resolution gusts, while Sec. 4.4.1 constitutes a thorough comparison of various regression models, differing only with respect to the set of allocated (potential) predictors.

## 4.3.2. Stepwise linear regression

### Basic concept

The basic concept of the developed statistical downscaling procedure is that for each COSMO-EU grid box an individual regression model is set up. Each of these models (a total number of 436.905 equations, one for each grid box) represents the statistical relationship between, on the one hand, the wind gusts (modeled by the dynamical hindcasts of the 181 selected storm situations at the respective COSMO-EU grid box, and on the other hand, some predictor variables at surrounding coarse scale grid boxes from ERA40, ERA-Interim, and EPS forecasts, respectively.

As described in Sec. 4.2.2 and 4.3.1 the predictands are temporally aggregated to six-hour maximum gusts with the six-hour window centered over the time step of the respective predictors. These (potential) predictors are wind speeds as well as meridional and zonal components at 10 m above the surface, after being scaled by climatological 98th percentile of the respective quantity and subsequently interpolated to the common $\sim$1.125°-grid.

The basic approach for the statistical model is a *Stepwise Linear Regression* (SLR). This method includes not only the quantification of statistical relationships (by estimating regression coefficients) but also an objective selection of skillful predictors from a set of

allocated potential explanatory variables. Generally, SLR schemes consist of repeated testing for predictors to be added to (removed from) the regression model if these yield (no) significant skill. Different procedures are possible. The utilized algorithm (feature of the *MATLAB*-software) is a combination of *forward selection* and *backward selection*, suitable for applications where no a priori knowledge about a reasonable model is available (see e.g. Wilks, 2006, his Sec. 6.4.2 and 6.4.3, for a basic introduction to stepwise regression and both screening procedures). The scheme works as follows:

1. Fit the initial model. If no initial predictors are specified – as in the current study – the initial model consists of the mean value only (effectively a straight line with a slope of 0 and intercept according to the predictands' sample mean).

2. *Forward selection*:
   Check all available potential predictors individually whether they can improve the model significantly (*p-value*<0.05 according to *f-test* of the sum of squared residuals).

   Fit the new model including the predictor yielding the best improvement.

   Repeat until no more predictors can improve the model significantly.

3. *Backward selection*:
   Check whether any predictors can be removed without significant influence (*p-value*<0.1) to the model's quality.

   Fit the new model without the predictor most likely to be dispensable.

   Repeat until no predictors can be removed without significant loss of quality.

Generally it should be noted that statistical models resulting from SLR are depending on the chosen initial model and the respective sequence of adding and removing predictors. Hence, they constitute local optimizations only. It might be that other predictor combinations yield results of similar (or even better) quality. However, for most applications this is no important problem and as long as the aim is only producing statistical modeling results of reasonable accuracy, this empirical "black box" approach is quite adequate (Wilks, 2006, p. 212).

**Overview: reference model and potential predictors**

Tab. 4.2 contains a short summary of the statistical models under comparison for this study. As already mentioned, an approach similar to Haas and Pinto (2012, called M0 here) is used as a reference for all other approaches. This method does not contain an objective selection of skillful predictors. Instead for all COSMO-EU grid boxes, the scaled wind speeds of the nine surrounding coarse scale grid boxes are used as predictors[3].

Table 4.2.: Overview of tested statistical downscaling approaches

| | |
|---|---|
| M0 | reference model similar to Haas and Pinto (2012): **wind speeds of surrounding nine grid boxes** taken as predictors **without SLR** |
| M1 | wind speeds of surrounding nine grid boxes **and their squares** as potential predictors **for SLR** |
| M2 | wind speeds of surrounding nine grid boxes, squares, **and interactions** as potential predictors for SLR |
| M3 | wind speeds of **grid boxes within 300 km** and their squares as potential predictors for SLR |
| M4 | **transformed wind speeds** (Box and Cox, 1964) of surrounding nine grid boxes as potential predictors for SLR |
| M5 | **wind components** of surrounding nine grid boxes and their squares as potential predictors for SLR |
| M6 | wind components of surrounding nine grid boxes, squares, **and interactions** as potential predictors for SLR |
| M7 | wind components of **grid boxes within 300 km** and their squares as potential predictors for SLR |
| M8 | wind components of grid boxes within 300 km and their squares as potential predictors for SLR; **squared gusts as predictands** (square roots calculated before validating) |
| M9 | **wind speeds and components** of grid boxes within 300 km and their squares as potential predictors for SLR |

All other approaches – developed and tested in this study – contain the objective predictor selection by means of the SLR (see Sec. 4.3.2). The differences between these models are the allocated potential predictors, that is the pool from which the SLR-algorithm has to select appropriate predictors given the training data. As already mentioned,

---

[3]Haas and Pinto (2012) took the 16 surrounding grid boxes, though on the original ERA-Interim-grid of ∼0.7°. They did not perform any scaling of the predictor wind speeds as they were considering only one predictor data set

wind speeds and wind components are provided as potential predictors for all models, but differently treated. For most models it is not only the winds themselves but also their squares allocated to the SLR. Initial tests in the course of this study, not including such squared terms, revealed that wind gust maxima of the core storm field were often underestimated. Once, such squared terms are included, it is usual in statistical modeling to include all second-order terms, that is also interactions (products of different predictors). Hence, these interactions are additionally allocated to some models. As 6-hourly data is considered and the synoptic systems underlying the analyzed wind storms exhibit translation velocities of up to 100 km/h, wind speeds (and components) of more distant coarse scale grid boxes are presented to some models. Based on the mentioned translation velocity of 100 km/h and the time window of gust aggregation ($t \pm 3\text{h}$), grid boxes within a radius of 300 km are considered as potential predictors for some models. Two models differ more substantially: The first is M4, where the wind speeds (of surrounding nine coarse scale grid boxes) are not only scaled by the 98$^{\text{th}}$ percentile (as for all other models; see Sec. 4.3.1) but additionally transformed according to Box and Cox (1964). This power-law transformation is used to ensure a symmetric distribution of the predictor variables, potentially useful for regression approaches. The second deviating method is M8 with the wind components only used as predictors but the squares of the high resolution gusts being the predictands, motivated by the geometric relationship between wind components and wind magnitude.

### 4.3.3. Validation metrics

The main validation of the performance of the statistical downscaling is performed with the dynamical regionalizations utilized as reference. Once again, it should be noted that these simulations are run in hindcast mode. Hence, these simulations do not represent the respective storm events exactly as they are diagnosed from reanalyses or EPS forecasts. Instead, they are expected to increasingly deviate from the coarse scale predictor data sets the longer the lead time.

The primary validation metric chosen is the *Mean Squared Error* (MSE)

$$\text{MSE} = \frac{1}{K} \sum_{k=1}^{K} \left( \text{sm}_k - \text{dm}_k \right)^2 \tag{4.1}$$

averaging the squared differences between statistical (sm) and dynamical model (dm) results. These differences can be calculated for every grid box and (six-hourly) time step. However, the focus of this study is on extreme wind gusts potentially related to damage and loss. Thus, event "footprints" are calculated for validation, that is the spatial field of maximum gusts over the entire storm episode (restricted to a maximum of 51 hours, see Sec. 4.2.2). Regarding the MSE, this is a conservative approach as higher gusts are in most cases related to larger differences between statistical and dynamical downscaling. Hence, the MSE is the error averaged over all grid boxes ($k$). Comparing two statistical models can be done by calculating the *Mean Squared Error Skill Score* (MSESS)

$$\text{MSESS} = 1 - \frac{\text{MSE}_{\text{sm}}}{\text{MSE}_{\text{sm,ref}}} \qquad (4.2)$$

yielding the fractional error reduction compared to the reference model $\text{sm}_{\text{ref}}$. As already mentioned in Sec. 4.1, a statistical model similar to the approach of Haas and Pinto (2012) is used as the reference for all approaches developed in this study. According to Murphy (1988) the MSE can be split up into several components

$$\text{MSE} = \left(\overline{\text{sm}} - \overline{\text{dm}}\right)^2 + \sigma_{\text{sm}}^2 + \sigma_{\text{dm}}^2 - 2\sigma_{\text{sm}}\sigma_{\text{dm}}r_{\text{sm,dm}} \qquad (4.3)$$

The first term ($\left(\overline{\text{sm}} - \overline{\text{dm}}\right)^2$) is the squared bias, naturally contributing positively to the MSE. The last term is a product of the standard deviations ($\sigma_{\text{sm}}$ and $\sigma_{\text{dm}}$) calculated over all grid boxes and the *product-moment correlation* ($r_{\text{sm,dm}}$; Pearson, 1896), contributing negatively. Additionally, the two sample variances contribute positively to the MSE. This means that a comparably low variance of a model might be beneficial in terms of the MSE. This is certainly not desirable for the targeted result of realistic storm footprints (including, a distinct discrimination between storm-affected and non-affected regions). To get a better understanding of systematic differences between the statistical models under comparison, these parameters contributing to the MSE – the bias, the correlation, and the ratio of the two standard deviations – are additionally analyzed.

A fair validation requires the modeling result under consideration not to be part of the data the model was trained with. For the current study, all validation results presented are based on a three-fold cross-validation. The set of 181 storm episodes is randomly split into three subsets. Two subsets are used to train the respective downscaling procedure which is subsequently used to model wind gusts for the third subset. After three

permutations, results for all storm episodes are generated. Hence, an assessment of the statistical model's quality can be provided based on all 181 storm events without them being part of the respective training data.

## 4.4. Results

### 4.4.1. Validation for different sets of potential predictors with dynamical downscaling as reference

The primary validation result, comparing the tested statistical downscaling approaches are shown in Fig. 4.2, separately with respect to both gust estimation approaches. The DWD-gust (Schulz, 2008) yields lower gust estimates for most European regions compared to the Brasseur-gust (Brasseur, 2001). Hence, the former is also associated with smaller deviations between statistical and dynamical modeling. For both gust estimations, M4 yields the worst results of all methods. Analyses of the individual storm footprints reveal that this approach (based on transformed wind speeds) is not suitable at all. The reasons for this bad performance are unclear yet. All other methods provide similar quality as measured by the MSE. According to the MSESS – calculated based on average of all storm footprints which is not drawn in Fig. 4.2 – M8 and M9 are the best methods under comparison, 7.9% (5.5%) and 7.3% (8.4)% better than the simple approach equivalent to the method of Haas and Pinto (2012) for the DWD-gust (Brasseur-gust). The median footprint MSE of these two methods is 6.4 (12.8) m/s for M8 and 6.6 (12.4) m/s for M9. Using more grid boxes than the surrounding nine as potential predictors systematically provides better results (M3 vs. M1 and M7 vs. M5), whereas Providing interaction-terms as potential predictors additional to squares does not lead to better downscaling results. Instead, methods including interactions perform worse than their counterparts without these potential predictors (M2 vs. M1 and M6 vs. M5). This is a clear indication of overfitting, as the models with interaction-terms are provided with all potential predictors also allocated to their counterparts without interactions. However, the huge number of potential predictors to select from (90 and 342 for M2 and M6, resp., compared to 18 and 36 for M1 and M5, resp.) obviously leads to too complex models, fitted to some specific characteristics of the training data but not beneficial to other storm events. Most of the statistical model solely based on

the wind components perform worse than those based on absolute wind speeds. M9 combines wind speeds and components. Spatial analyses (not shown) comparing M9 and M3 reveal that wind components are particularly beneficial for regions with pronounced orography. As far as the MSE is considered, the approach similar to that of Haas and Pinto (2012) is competitive compared to the methods developed in this study, even without testing the predictors for their skill.
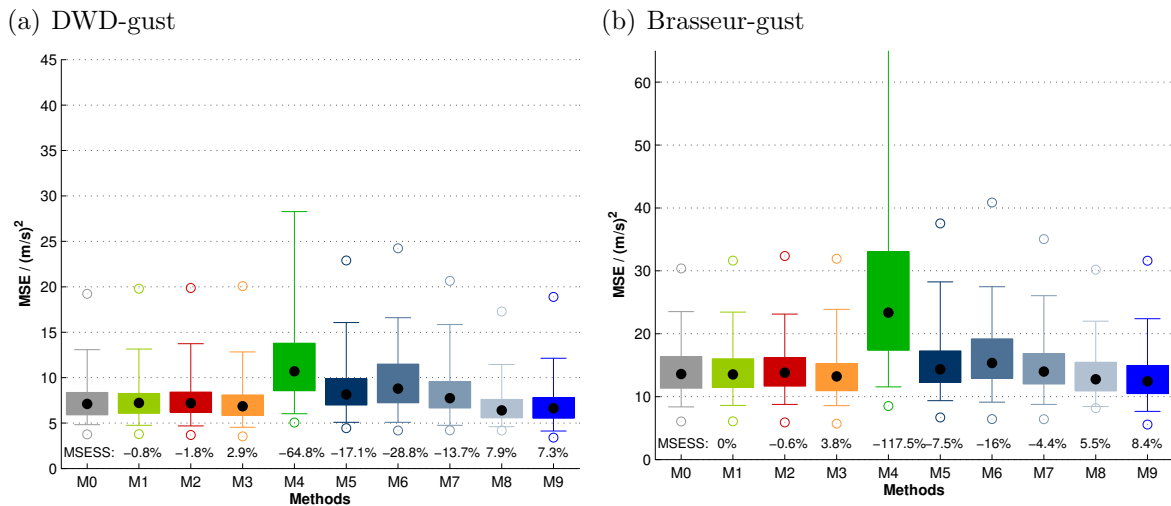


Figure 4.2.: Box-plots of MSE (compared to dynamical regionalization) with respect to 181 winter storm footprints according to nine different methods of statistically modeling the (a) DWD-gust, and (b) the Brasseur-gust: median (black dot), inter-quartile-range (box), 95%-interval (whiskers), minimum/maximum (circles); MSESS with M0 as reference model

Analyzing the bias in equivalent manner (Fig. 4.3) shows that the distributions of most procedures are shifted towards a negative bias. This is a result of the validation based on the footprints, that is the maximum event gust. These maximum gusts are systematically underestimated by most methods (compensated by overestimation of weak gusts by most methods). The methods incorporating wind components perform better in this respect than those methods based on wind magnitudes only. Again M9 improved M3 mainly in mountainous regions. The reference model similar to Haas and Pinto (2012) features the second largest bias of all methods for both gust estimations with a median of -0.62 m/s and -0.32 m/s for the Brasseur-gust and DWD-gust, respectively. Only M4 which was diagnosed non-competitive based on the MSE-results is worse. Especially the most extreme gusts are underestimated by this simple approach (not shown).
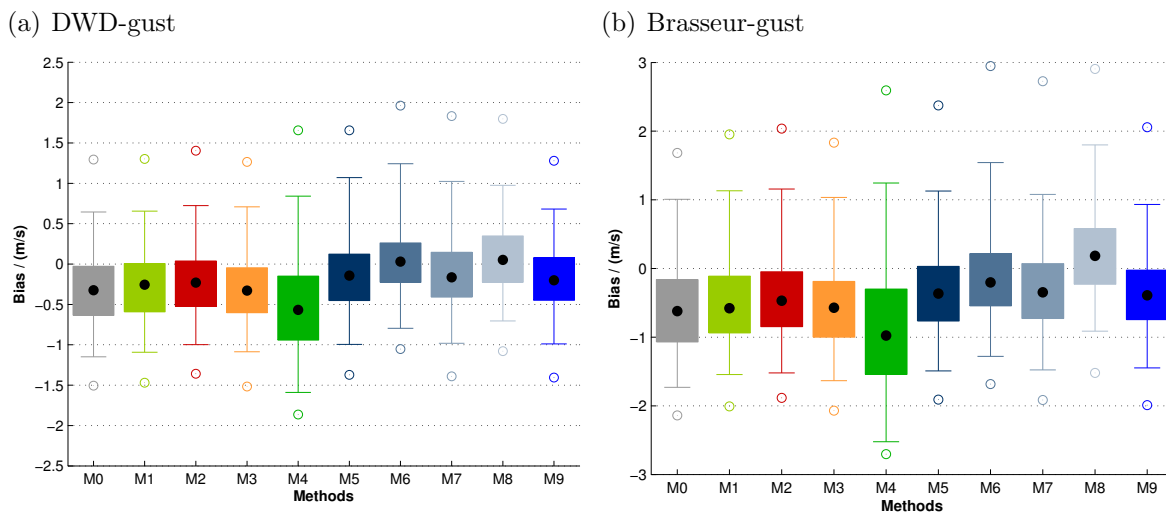
(a) DWD-gust

(b) Brasseur-gust



Figure 4.3.: Box-plots of the bias (compared to dynamical regionalization) with respect to 181 winter storm footprints according to nine different methods of statistically modeling the (a) DWD-gust, and (b) the Brasseur-gust: median (black dot), inter-quartile-range (box), 95%-interval (whiskers), minimum/maximum (circles)

Comparing the product-moment correlation regarding the storm footprints (Fig. 4.4) exhibits the overall level of statistical downscaling procedures based on Multiple Linear Regressions. All approaches yield a median correlation higher than 0.85 for the Brasseur-gust and beyond 0.9 for the DWD-gust. Methods solely based on wind components are again slightly worse than those based on wind magnitudes. The best method regarding the DWD-gust is M8 with a median correlation of 0.941 followed by M9 with a median correlation of 0.940. With respect to the Brasseur-gust it is M9 (0.926) followed by M3 (0.924).

Finally, the standard deviations of the statistical downscaling procedures are compared to those of the dynamical regionalization (Fig. 4.5). These results reveal that all methods based on wind magnitudes but also M8 (wind components predicting squared gusts) feature standard deviations shifted towards smaller values compared to the dynamical regionalizations. The other methods based on wind components (M5–M7) perform better regarding mean and median but feature a wider range over the 181 storm events. The reference model M0 – similar to the approach of Haas and Pinto (2012) – exhibits the smallest standard deviations of all methods. This is particularly obvious for the Brasseur-gust with the third quartile of the M0-distribution below the median of all
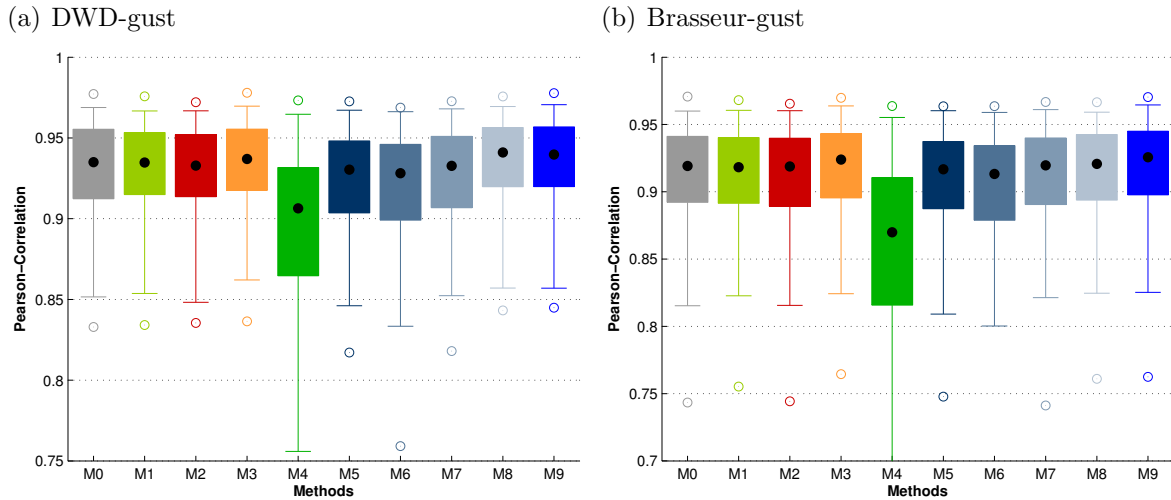
(a) DWD-gust

(b) Brasseur-gust



Figure 4.4.: Box-plots of the product-moment correlation (compared to dynamical re-gionalization) with respect to 181 winter storm footprints according to nine dif-ferent methods of statistically modeling the (a) DWD-gust, and (b) the Brasseur-gust: median (black dot), inter-quartile-range (box), 95%-interval (whiskers), mini-mum/maximum (circles)

methods but M8. For the DWD-gust, it is less extreme but still evident. This is due to distinct underestimation of maximum gusts on the one hand and overestimation of weak gusts on the other hand (not shown here). This property "helps" M0 achieving a quite competitive MSE, although a method less clearly discriminating wind storm fields against calm areas may be of limited use for approaches such as loss-modeling.

One fact to keep in mind is that the storm footprints being the basis for all validations presented so far are compiled from wind gusts of differing time steps. Given the set up of the reference data, this is equivalent to differing lead times of the hindcasts. For that reason, the agreement between statistical and dynamical regionalization is expected to be dependent on these lead times. Fig. 4.6 – evaluating MSEs calculated for the individual time steps – shows that this expectation is fulfilled. The average over all storm events approximately follows a linear increase of these disagreements, while the storm's distribution around these averages is clearly not symmetric. Beyond lead times of 24h discrepancies for individual storms grow much faster than those of the majority of such events. The linear fit and its intercept roughly indicate the order of model's performance without this effect of growing differences between statistical downscaling and dynamical regionalization as part of free running hindcasts.
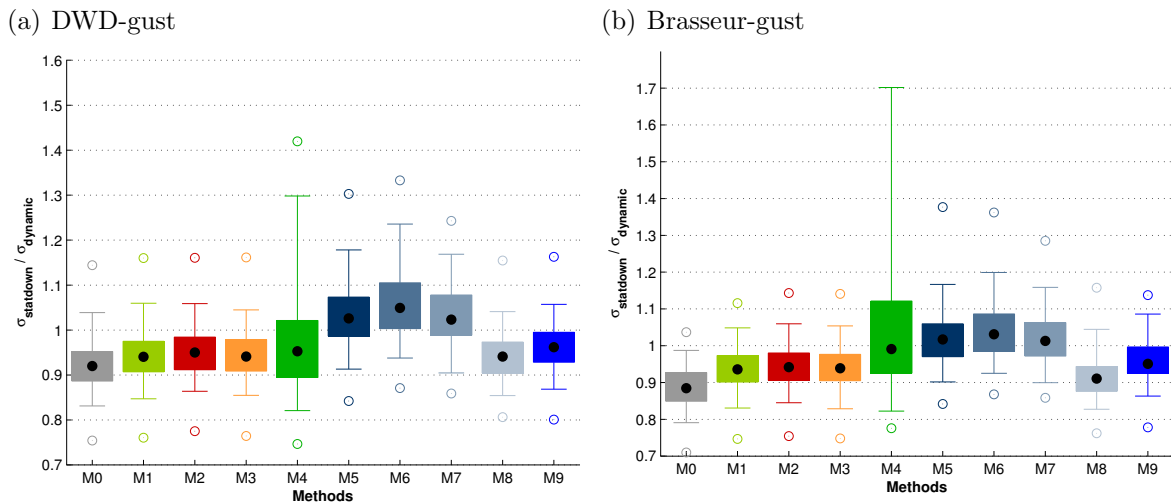
Figure 4.5.: Box-plots of the ratio of standard deviations (statistical modeling divided by dynamical downscaling) with respect to 181 winter storm footprints according to nine different methods of statistically modeling the (a) DWD-gust, and (b) the Brasseurgust: median (black dot), inter-quartile-range (box), 95%-interval (whiskers), minimum/maximum (circles)

The synthesis of all validation metrics and both gust estimation approaches is that M9 is the most suitable of all tested methods. It performs among the best models for all analyzed metrics. This synthesis is not only based on mean and median values but resulting from a thorough reflection of the total distributions over the 181 storm events.

Additional analyses (not shown) reveal that the validation metrics presented for statistical downscaling approaches such as M9, using the dynamical regionalizations as reference, are in the same order as comparisons of the two gust estimations originating from identical simulations and comparisons of identical gust estimations originating from dynamical regionalizations initialized with an offset of 24 hours. Hence, an additional result is that the performance of the statistical downscaling is within the range of reference uncertainty for these analyses.

Comparison against observations indicate that a statistical downscaling set up as M9 yields results of equal quality as the dynamical hindcasts for the historical storm events. However, such comparisons are constrained to few European areas and storm events during recent years given the very limited (free) availability of gust measurements. For this reason, these results are not shown here.
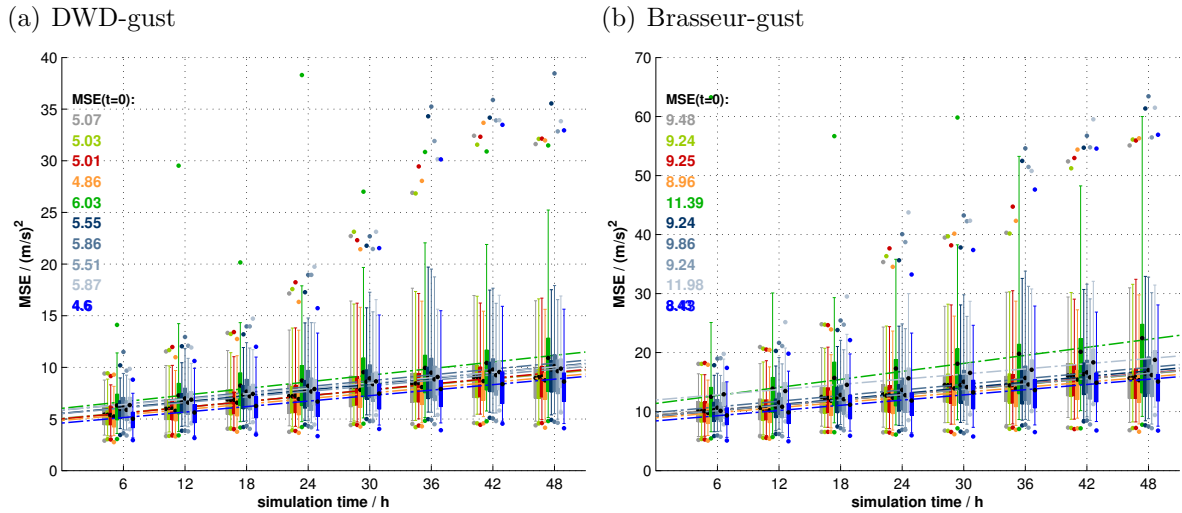
(a) DWD-gust

(b) Brasseur-gust



Figure 4.6.: Box-plots of the temporal evolution of the MSE (compared to dynamical regionalization) with respect to 181 winter storm footprints according to nine different methods of statistically modeling the (a) DWD-gust, and (b) the Brasseur-gust: median (black dot), inter-quartile-range (box), 95%-interval (whiskers), minimum/maximum (circles) and linear fit with intercept (extrapolation to lead time of 0h) written on the left

## 4.4.2. Qualitative results

After presenting a profound quantitative validation of the statistical downscaling procedures tested in the course of this study, this section shall be concluded by a qualitative but more tangible comparison of statistical and dynamical regionalization. For this purpose, the footprint according to dynamical and statistical regionalization (and their difference) for winter storm "Kyrill" is shown in Fig. 4.7. The overall pattern is in very good agreement between dynamical and statistical regionalization. The dynamical simulations feature strong convective activity locally enhancing wind gusts within the cold front leading to "streaks" in the footprint along the translation direction of the front. Such meso-scale features that are very specific to this unique storm event (observed in reality, though not necessarily at exactly those locations; see e.g. Gatzen et al., 2011), can not be expected to be represented by the statistical downscaling. Permanent effects on the wind field, such as orographic influences and effects from surface roughness are implicitly modeled by the statistical approach as different transfer functions have been trained for the individual COSMO-EU grid boxes. The sufficient representation of these
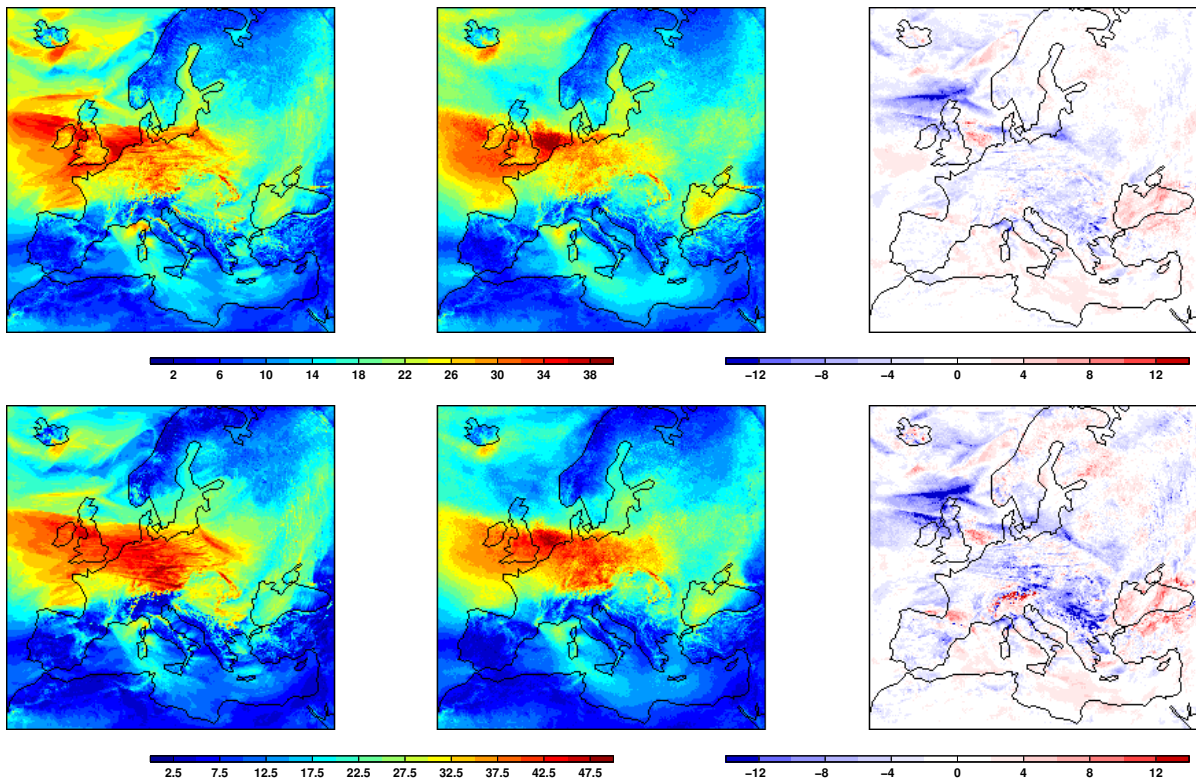
Figure 4.7.: Footprint of winter storm "Kyrill" (18./19.01.2007) according to dynamical (left) and statistical downscaling (center) as well as the difference between the two (right) with respect to the operational DWD gust estimation (Schulz, 2008, top) (top) and the approach of Brasseur (2001, bottom) (bottom) in m/s

factors can be estimated from land-sea contrasts and gust patterns following distinct orographic features such as the Rhine valley, the Carpathian Mountains, or (south of the core storm field) in the Alps, over Greece and Turkey. Even remote effects are represented if they are typical for many wind storms and as such trained to the statistical model. An example for such remote effects is visible in the high wind gusts from the Gulf of Lion towards Corsica and Sardinia with a steep gradient to much weaker gusts southwest of this area, resulting from the wake of the Pyrenees.

## 4.5. Quantification of uncertainties

Certainly the statistical downscaling procedure developed in this study is not perfect. The meso-scale convective activity within "Kyrill"'s cold front was just one example

of meteorological features that are not and can not be represented by the downscaling procedure. Another issue is that the downscaling naturally depends on the quality of the input data. Some features relevant for extreme wind speeds and related loss, such as "sting jets" can not be appropriately resolved by current reanalyses such as ERA-Interim and numerical weather prediction models of similar resolution (see e.g. Martinez-Alvarado et al., 2013). These issues underline the necessity of quantifying uncertainty of the results provided by the statistical downscaling (though facing the fact that standard gust diagnostics of dynamical regionalizations are not attributed with an uncertainty estimate).

For the current study, such an uncertainty estimate is derived from the comparison of dynamical and statistical downscaling, once again consulting the former in the sense of a reference. The residuals are analyzed depending on the statistically modeled gust. Fig. 4.8 shows such an analysis for a grid box in Offenbach, Germany. It is obvious that the uncertainty of the modeling grows with increasing values of the modeled gust itself. This phenomenon – called heteroscedasticity – is not desired for regression models. It does not affect the (mean) fit, that means the Multiple Linear Regression approach underlying the SLR is still appropriate, but standard error metrics derived from least-squares regression are not valid as they assume constant uncertainty (homoscedasticity). This issue is detectable for most COSMO-EU grid boxes. The solution developed for this study is to estimate the uncertainty of the statistical model by conducting a quantile regression (Koenker and Hallock, 2001) based on these residual analyses. For this purpose the residual 5%- and 95%-quantiles are estimated by the quantile regression based on a constant, a linear, and a logarithmic term. The latter is meant to model the saturation of uncertainty detectable for many grid boxes (not shown). The resulting 90%-confidence-interval is also shown in Fig. 4.8 (red lines). Such quantile regressions are conducted for each COSMO-EU grid box based on the three-fold cross-validation results for the final statistical downscaling method M9, yielding efficient uncertainty estimates (three regression coefficients for each, the 5%- and the 95%-quantile) related to the results of the downscaling itself.
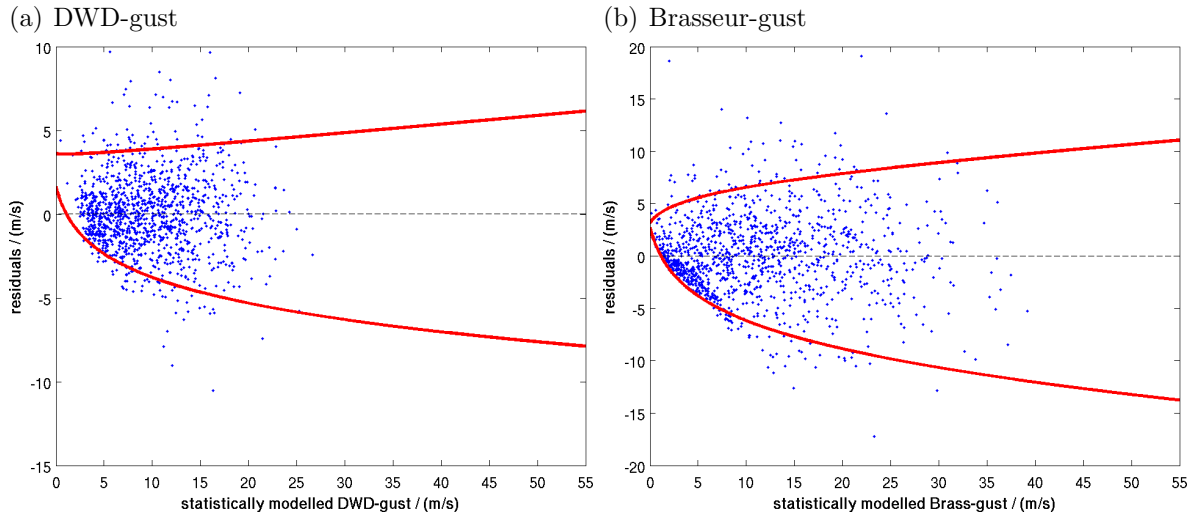
Figure 4.8.: Residuals (blue dots; difference between dynamical and statistical regionalization) depending on the statistically modeled gust according to (a) Schulz (2008) and (b) Brasseur (2001) for a COSMO-EU grid box at Offenbach, Germany (50.10°N, 8.75°E), and 90%-confidence-interval estimated by quantile regression approach

## 4.6. Summary and outlook

The current study developed a statistical downscaling procedure suitable for generating a comprehensive winter storm event set based on ECMWF's reanalyses (ERA40 and ERA-Interim) and the medium-range ensemble forecasts (EPS). The main purpose of this event set will be re-insurance loss modeling. The chosen downscaling approach estimates high spatial resolution (7 km) surface gust fields from coarse scale reanalysis and EPS wind magnitudes and the individual components (spat. resolution of approx. 1.125°, 0.7°, and 0.45°) covering the whole European continent. This is done by deriving individual transfer functions for each high resolution grid box (a total number of 436.905). The basic concept of the statistical downscaling – which was developed on the basis of 181 storm episodes, dynamically downscaled with the numerical weather prediction models (GME and COSMO-EU) of the DWD – is Multiple Linear Regressions. Various combinations of potential predictors were presented to a Stepwise Linear Regression algorithm in order to objectively select skillful (evaluated by three-fold cross-validation and several MSE-related metrics) predictors. It is shown that the combination of wind magnitudes and the individual components, as well as the squared values of these parameters yield the best results. Additionally, the statistical model is substantially im-

proved by including coarse scale predictors of grid boxes more distant (up to 300 km) from the respective high resolution grid box compared to including only the surrounding nine coarse scale grid boxes. These three characteristics of the final statistical model itself (including wind components, squared values, and more distant predictors) and the objective selection of skillful predictors eventually pose the difference to the simple approach published by Haas and Pinto (2012). The presented systematic analysis with respect to suitable predictor combinations – an additional difference to Haas and Pinto (2012), not motivating their model set up at all – shows that the statistical downscaling developed in this study performs better than that simple approach. Apart from a 5.5–8.4% lower MSE (Haas and Pinto, 2012, analyzing RMSE) it is shown that the approach of this thesis yields wind gust variances closer to those of dynamical simulations, considerably discriminating the storm field against areas of weak gusts. It should be noted that lower variance is a way to "cheat" at a lower MSE, although not desirable for the aimed follow-on applications, such as loss-modeling.

One constraint of the statistical downscaling results additional to the dependency on the quality of the predictor data (see Sec. 4.5), is that they reflect mean relationships between coarse scale wind fields and high resolution gusts. Apart from "unusual" storm-internal dynamics leading to extreme behavior of gusts there is also the issue of storm events being "unusual" on the larger scale. For example, wind storms exhibiting exceptionally high translation velocities cannot be modeled appropriately by the statistical downscaling procedure developed here. If such translation velocity is combined with a non-typical track, the problem is even worse. An example for the latter scenario is winter storm "Xynthia", being the worst statistically modeled historical storm event (see Fig. A.6 in Appendix A.3). More sophisticated statistical models are required to account for such non-typical behavior. However, for the huge majority of all considered events, the statistical downscaling yields completely satisfactory results.

The statistical downscaling procedure is trained and validated for two different approaches with respect to the estimation of surface gusts (Schulz, 2008; Brasseur, 2001). The obtained transfer functions are applied to the totality of all European winter storm events diagnosed in ERA40- and ERA-Interim-reanalyses ($\sim$3.700) and EPS data ($\sim$280.000), providing the comprehensive event set this effort was aiming at. These numbers highlight the necessity of a statistical downscaling procedure, as dynamical downscaling to the targeted resolution would have been infeasible within acceptable

time. In contrast to this, the statistical downscaling application for this huge event set took approx. two weeks to be completed, providing estimates equivalent to both considered gust diagnostics. The approach developed is very flexible and, hence, applicable to a wide range of model simulations, from medium-range weather forecasts (as in this study) over decadal climate predictions to centennial climate projections.

Future work is dedicated to the already-mentioned loss-modeling, as well as climatological analyses. Regarding the latter, studies analog to those based on the coarse predictor data (see Osinski, 2014; Osinski et al., 2015) can now be conducted on high spatial resolution. In particular, this is planned with respect to the estimation of return periods and return levels of wind gusts within European winter storms. Such analyses do profit substantially from incorporation of multiple variations of past winter wind storms originating from the EPS, as they constitute alternative scenarios as they could have happened following the principles of physical consistency as guaranteed by the dynamical models. In the context of loss-modeling not only the follow-on application for the specific purposes of the re-insurance industry are conducted but also the development of probabilistic loss-event predictions, based on ensemble forecasts such as the EPS.

# 5. Synthesis

## 5.1. Summary

The current thesis is dedicated to European winter wind storms and questions regarding (i) an optimal impact-oriented identification, (ii) any predictive skill on perennial to decadal time scales potentially relevant for socio-economic stakeholders, and (iii) a statistical downscaling approach suitable for surface wind gusts in this respect.

Regarding the objective event identification, the impact-oriented WiTRACK-algorithm (Leckebusch et al., 2008a; Renggli, 2011) was substantially improved in the course of this thesis (Sec. 2.3.3). These improvements tackle errors and inconsistencies of previous versions but more importantly, they contain significant advances regarding the spatio-temporal tracking of these events. Additionally (in collaboration with several colleagues), the scheme was made more flexible in several aspects, widening the range of potential applications. In this context, the first globally consistent climatological analysis of potentially damaging wind storms was presented (Sec. 2.4.3).

The underlying approach still must be considered as experimental with respect to the question, whether the principle of identifying such events based on exceedances of the local $98^{\text{th}}$ percentile of wind speeds (Klawa and Ulbrich, 2003; Leckebusch et al., 2008a) is transferable to other regions than Europe and the North Atlantic. The analyses of this thesis indicate that this is generally the case for extra-tropical regions on both hemispheres. Results for the tropics seem to be acceptable, however, they suggest that a higher threshold might be more appropriate if explicitly targeting wind fields of tropical cyclones.

The global climatology of wind storm events based on the available 35 years of the *ERA-Interim*-reanalysis overall matches the results of other studies focused on specific regions and storm types. In contrast to these tailored studies, the current thesis performed a

consistent comparison of characteristics of very different storm types from various regions spread over the globe. A variety of insights is provided by this comparison: Wind storms identified over the tropical Northern Indian Ocean feature characteristics significantly differing from other tropical regions. Preliminary analyses lead to the hypothesis, that the majority of the identified events are related to synoptic-scale monsoon low-pressure systems (see e.g. Hurley and Boos, 2014), while event samples of other tropical regions are more influenced by meso-scale tropical cyclones. Comparing tropical storm property distributions of different basins yield events over the North Atlantic tending to be smaller than the others but featuring a wider spread than all other storm types (including extra-tropical events) with respect to potential damage as measured by the ESSI (Leckebusch et al., 2008a). Comparing extra-tropical wind storms shows that southern-hemispheric systems generally travel faster than Northern Hemisphere storms, while the latter feature overall higher loss-potentials estimated by the ESSI. Intra-hemispheric comparisons generally yield minor differences. Remarkable in this respect is the systematic difference in translation velocities between the North Pacific (faster) and the North Atlantic (slower) which is the primary reason of lower track densities over the North Atlantic but higher storm geneses densities, when compared to the North Atlantic.

Identifying Northern Hemisphere winter wind storms from initialized decadal climate predictions is the starting point in evaluating potential forecast skill of such simulations with respect to the frequency of one of the most relevant types of natural hazards. By doing that, Ch. 3 of this thesis continues the study of Kruschke et al. (2014), more physically motivated by considering extra-tropical cyclones. Compared to this predecessor study, the work presented in this thesis analyzes more decadal hindcast experiments and combines these simulations according to underlying initialization strategies in order to systematically examine these strategies. In the context of *MiKlip* – the German initiative for decadal climate prediction – the current study can be seen as pioneering in this respect, facing that the next development stage of the *MiKlip*-system (the *prototype*-system) will be such a combination of experiments based on slightly different initialization strategies. Related to that, this thesis (and the mostly identical manuscript of Kruschke et al., 2015) represents the first analysis of a substantial part of this new system – the *ORAff*-ensemble (see Sec. 3.2.1) consisting of 10 out of 30 members of the final *prototype*-system. Additionally, this study contains a novel drift correction approach – inspired by the work of Kharin et al. (2012) and Gangstø et al. (2013) – accounts for

drifts of (full-field-)initialized simulations that change over time. This drift correction is found to be more appropriate than standard procedures (International CLIVAR Project Office, 2011), assuming a constant model drift, independent of the climate state at the time of initialization. All these features – experiment combinations and related larger ensembles as well as the better drift correction – are beneficial for the evaluation of decadal prediction skill yielding more robust results than without these developments.

The result of this (probabilistic) assessment is that significant skill exists for 2–5y- and 2–9y-forecasts with respect to winter storm frequency over large parts of the extra-tropical Northern Hemisphere. However, much of this skill is associated with long-term changes already contained in uninitialized transient simulations. The added value of the enormous initialization effort is restricted to a few areas, namely the entrance regions of the stormtracks, the Eastern Mediterranean and parts of the Arctic. Interestingly, these results do not differ substantially over the considered initialization strategies. On the one hand, this is a positive result further fostering confidence in derived skill estimates. On the other hand, this queries the progress made with recent development stages of the *MiKlip*-system. However, the synthesis of the results presented by Kruschke et al. (2014) and Ch. 3 is that multi-annual prediction skill for parts of the extra-tropics is better than expected in the light of studies on oceanic forcing of the atmosphere (e.g. Frankignoul, 1985; Kushnir et al., 2002) facing that such oceanic forcing is assumed to be the primary source of predictability on these time scales. This surprisingly positive result may be related to the findings of Eade et al. (2014), revealing that prediction skill and predictability of phenomena in the "real world" may be larger than estimated from (imperfect) model experiments. Hence, the author of this thesis likes to see his work presented in Kruschke et al. (2014) and Ch. 3 in the optimistic, though factually founded spirit of Urbain Le Verrier, hopefully contributing to the advance of climate prediction.

Finally, a statistical downscaling approach for surface gusts within European winter storms was presented in Ch. 4. This downscaling procedure is suitable for efficiently estimating high resolution (7 km, 6-hourly) wind gusts according to two different estimations (Schulz, 2008; Brasseur, 2001) from spatially low-resolved results of reanalyses, numerical weather prediction or climate simulations. Such an approach is the only feasible solution when aiming at analyzing a great number of winter storm situations as necessary for (re-)insurance loss-modeling. The method developed in this thesis is based on a Stepwise Linear Regression algorithm, objectively selecting skillful predictors from

a larger set of potentially explaining variables. Such a transfer function is built for each individual high resolution grid box, eventually constituting 436.905 downscaling equations (and respective uncertainty estimates). Several variants of the statistical model were trained and validated on the basis of 181 dynamically regionalized (applying the numerical weather prediction models of DWD) winter storm episodes. It is shown, that such the statistical downscaling yields results of equal quality as dynamical regionalization when provided with the following potential predictors: wind speeds and components as well as their squares of coarse scale grid boxes located in the wider environment (up to a distance of 300 km) of the respective high resolution grid box. This set up of the statistical model is a result of a systematic comparison of different configurations. This comparison includes a benchmark approach, equivalent to a more simple procedure published by Haas and Pinto (2012). It is shown, that the simple model is outperformed by the method developed here (MSESS of 7.3% and 8.4% with respect to the two gust estimations), although the simple model "cheats" at a comparably low MSE by exhibiting a smaller variance (underestimating peak gusts, overestimating weak gusts) than all statistical model configurations considered in this study. The statistically modeled storm fields are complemented by polynomial estimates of uncertainty with respect to the statistical modeling, quantifying the 90%-confidence-interval depending on the modeled gust itself by means of six regression parameters (three for each, the upper and lower bound of the interval). Application of this downscaling approach to ECMWF's reanalyses and ensemble weather forecasts (EPS) yields the two alternative surface gust estimates in high spatial resolution over the whole European continent for a total number of ∼284.000 winter storm events, subsequently meant to be used for re-insurance loss-modeling.

## 5.2. Outlook

The current thesis constitutes a comprehensive effort regarding research related to European winter wind storms. Though representing substantial advances in several aspects, naturally various opportunities for future work remain.

Regarding the impact-oriented identification of wind storm events, several possibilities for optimizing the algorithm and the tracking routine in particular exist. On the other hand, such optimizations will be compulsorily related to some focusing to specific events and their behavior in this respect. Such a development would (at least partially) sacrifice the WiTRACK-property of being extremely flexible due to its rather simple design. This flexibility allows application to other variables in a more general context. Such "foreign" WiTRACK-analyses are conducted on a experimental basis for parameters such as surface pressure, potential vorticity (on isentropic levels), precipitation fields derived from radar data, and sea-surface temperatures. The latter application allowed for analyses equivalent to that of Ray and Giese (2012) regarding potential shifts of ElNiño- and LaNiña-patterns. Such examinations are not scientifically relevant for the content of this thesis, that is (European winter) wind storms. But if such "foreign" WiTRACK-applications are generally desired, the methodological implication for further development of the algorithm is to retain the flexible design of the scheme. Closer to the core purpose of WiTRACK is the analysis of tropical wind storms. In this respect, clear indications of the suitability of the scheme exist, many of them presented in this thesis. However, a basic issue remains the existence of potentially more appropriate thresholds for the identification of these events. Potential future studies targeting these storm types by means of WiTRACK do have to tackle this issue, performing a thorough analysis of wind speeds leading to damage in tropical regions. One specific further analyses regarding the global climatology of wind storms presented in Sec. 2.4.3 is already prepared. The respective WiTRACK-application (see Sec. 2.4.1) was conducted using the option of specifying mask files for the calculation of Region-SSIs based on 26 regions spread over the globe as defined by the IPCC (2012). Hence, a comparison of wind storm intensity affecting these regions, respective variabilities and changes over the period covered by ERA-Interim could be started right away. Similar analyses based on long-term climate simulations would complement such an examination.

Revisiting the results presented in Ch. 3 and the manuscript of Kruschke et al. (2015)

must be related to accounting for the yet early stage of decadal climate prediction. Hence, multiple issues regarding potential future work exist. Generally, the most relevant challenges are associated with research on optimal initialization, i.e. assimilation strategies better enabling the models to catch the actual climate state and phase. However, analyses on these strategies are hampered by the still evident lack of observational data for the ocean probably being the most important climate subsystem in this respect. Considering the presented analyses on winter storm frequencies (as well as cyclone frequencies tackled in Kruschke et al., 2014) and the skill of decadal predictions in this respect, many open questions remain regarding the underlying processes responsible for evident skill. Similarly, further analyses are necessary to answer the question whether such frequencies in other regions are simply not predictable on these time scales or if it is any shortcomings of the model systems (including initialization, ensemble generation, and post-processing strategies) hampering skill of current experiments. The analyses presented in both studies, Kruschke et al. (2014) and Ch. 3 (Kruschke et al., 2015) do represent pioneering research beyond current state of the art (see Reviewer comments in Appendix A.2) but additional effort on these process-oriented issues is necessary to foster confidence in found prediction skill.

In contrast to coarse spatial and temporal scales addressed by the research on decadal predictions, Ch. 4 deals with providing high resolution information with respect to (European) winter storms. Overall, the author of this thesis is confident having developed a statistical downscaling procedure which is hard to outperform regarding equivalent temporal and spatial resolution. Only few methodological issues remain open. A very general point in that respect is the question of additional/alternative predictors. In principle, a long list of candidates could be tested. However, few parameters seem to offer large potential. According to the author's expectation, one specific variable looks most promising, that is some quantitative measure of atmospheric stability such as the Brunt-Väisälä-frequency or simply the vertical (potential) temperature gradient. A respective refinement of the statistical model is easy to implement. For the current study, it was not tested simply because of the fact that respective data was not part of the initial comprehensive download regarding archived EPS forecasts. Planned applications of the developed statistical downscaling approach and subsequent analyses are climatological analyses based on the produced winter storm event set, equivalent to coarse scale examinations performed by Osinski (2014) and Osinski et al. (2015), such as deriving re-

turn periods of high resolution gusts. Additionally, research on probabilistic predictions of economic loss-events based on downscaled ensemble weather forecasts has already been started. Apart from these topics based on numerical weather prediction simulations, further application of the method to climate simulations is conceivable. Given the encouraging results regarding skill of decadal forecasts with respect to winter wind storms, connecting the scientific content of Ch. 3 and Ch. 4 by statistically downscaling such skillful predictions of wind speeds might be an application relevant for some socio-economic stakeholders. In fact, a first analysis in this context was performed by Reyers et al. (2015) and the already-mentioned approach of Haas and Pinto (2012) – though found to be suboptimal according to this thesis – is also tested (Haas et al., 2015).

# Bibliography

Balmaseda, M. A., K. Mogensen, and A. T. Weaver 2013: Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q.J.R. Meteorol. Soc.* **139**(674): 1132–1161, doi: 10.1002/qj.2063

Barredo, J. I. 2010: No upward trend in normalised windstorm losses in Europe: 1970-2008. *Nat. Hazard Earth Sys.* **10**(1): 97–104

Befort, D., M. Fischer, G. Leckebusch, U. Ulbrich, A. Ganske, G. Rosenhagen, and H. Heinrich 2014: Identification of storm surge events over the German Bight from atmospheric reanalysis and climate model data. *Nat. Hazards Earth Syst. Sci. Discuss.* **2**: 3935–3963, doi: 10.5194/nhessd-2-3935-2014

Bell, G. D., M. S. Halpert, R. C. Schnell, R. W. Higgins, J. Lawrimore, V. E. Kousky, R. Tinker, W. Thiaw, M. Chelliah, and A. Artusa 2000: Climate Assessment for 1999. *Bull. Amer. Meteor. Soc.* **81**: 1328–1378

Bengtsson, L., K. I. Hodges, and M. Esch 2007: Tropical cyclones in a T159 resolution global climate model: comparison with observations and re-analyses. *Tellus Series A-Dynamic Meteorology And Oceanography* **59**(4): 396–416

Berry, G., M. J. Reeder, and C. Jakob 2011: A global climatology of atmospheric fronts. *Geophys. Res. Lett.* **38**: L04809, doi: 10.1029/2010GL046451

Bjerknes, J. A. B. 1964: Atlantic air–sea interaction. *Adv. Geophys,* **10**: 1–82

Blake, E. S., T. B. Kimberlain, R. J. Berg, J. P. Cangialosi, and J. L. Beven, II 2013: Hurricane Sandy. Tropical Cyclone Report AL182012, NOAA – National Hurricane Center, Miami, FL, USA. Published online via `http://www.nhc.noaa.gov/data/tcr/AL182012_Sandy.pdf`

Blechschmidt, A. M. 2008: A 2-year climatology of polar low events over the Nordic Seas from satellite remote sensing. *Geophys. Res. Lett.* **35**(9): L09815

Blender, R., K. Fraedrich, and F. Lunkeit 1997: Identification of cyclone-track regimes in the North Atlantic. *Quart. J. Roy. Meteor. Soc.* **123**(539): 727–741, doi: 10.1002/qj.49712353910

Boer, G., V. Kharin, and W. Merryfield 2013: Decadal predictability and forecast skill. *Climate Dyn.* **41**: 1817–1833, doi: 10.1007/s00382-013-1705-0

Box, G. E. P. and D. R. Cox 1964: An analysis of transformations. *Journal of the Royal Statistical Society* **B26**: 211–243

Brasseur, O. 2001: Development and application of a physical approach to estimating wind gusts. *Mon. Wea. Rev.* **129**(1): 5–25, doi: 10.1175/1520-0493(2001)129⟨0005: DAAOAP⟩2.0.CO;2

Buckingham, C., T. Marchok, I. Ginis, L. Rothstein, and D. Rowe 2010: Short- and Medium-Range Prediction of Tropical and Transitioning Cyclone Tracks within the NCEP Global Ensemble Forecasting System. *Wea. Forecasting* **25**(6): 1736–1754, doi: 10.1175/2010WAF2222398.1

Buizza, R. and A. Hollingsworth 2002: Storm prediction over Europe using the ECMWF ensemble prediction system. *Meteorol. Appl.* **9**(3): 289–305, doi: 10.1017/S1350482702003031

Caron, L.-P., C. G. Jones, and F. Doblas-Reyes 2014: Multi-year prediction skill of Atlantic hurricane activity in CMIP5 decadal hindcasts. *Climate Dyn.* **42**(9-10): 2675–2690, doi: 10.1007/s00382-013-1773-1

Chan, J. C. L., J. E. Shi, and C. M. Lam 1998: Seasonal forecasting of tropical cyclone activity over the western North Pacific and the South China Sea. *Wea. Forecasting* **13**(4): 997–1004, doi: 10.1175/1520-0434(1998)013⟨0997:SFOTCA⟩2.0.CO;2

Christensen, J., K. Krishna Kumar, E. Aldrian, S.-I. An, I. Cavalcanti, M. de Castro, W. Dong, P. Goswami, A. Hall, J. Kanyanga, A. Kitoh, J. Kossin, N.-C. Lau, J. Renwick, D. Stephenson, S.-P. Xie, and T. Zhou 2013: Climate Phenomena and their

Relevance for Future Regional Climate Change. In: T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, chapter 14, 1217–1308, Cambridge Univ. Press, Cambridge, United Kingdom and New York, NY, USA

Colle, B. A. and M. E. Charles 2011: Spatial Distribution and Evolution of Extratropical Cyclone Errors over North America and its Adjacent Oceans in the NCEP Global Forecast System Model. *Wea. Forecasting* **26**(2): 129–149, doi: 10.1175/2010WAF2222422.1

Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, J. Gleason, B. E., R. S. Vose, G. Rutledge, P. Bessemoulin, S. Broennimann, M. Brunet, R. I. Crouthamel, A. N. Grant, P. Y. Groisman, P. D. Jones, M. C. Kruk, A. C. Kruger, G. J. Marshall, M. Maugeri, H. Y. Mok, O. Nordli, T. F. Ross, R. M. Trigo, X. L. Wang, S. D. Woodruff, and S. J. Worley 2011: The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.* **137**(654): 1–28, doi: 10.1002/qj.776

Counillon, F., I. Bethke, N. Keenlyside, M. Bentsen, L. Bertino, and F. Zheng 2014: Seasonal-to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System Model: a twin experiment. *Tellus A* **66**: 21074, doi: 10.3402/tellusa.v66.21074

Crompton, R. P. and K. J. McAneney 2008: Normalised Australian insured losses from meteorological hazards: 1967-2006. *Environmental Science & Policy* **11**(5): 371–378, doi: 10.1016/j.envsci.2008.01.005

Dacre, H. F., M. K. Hawcroft, M. A. Stringer, and K. I. Hodges 2012: An Extratropical Cyclone Atlas A Tool for Illustrating Cyclone Structure and Evolution Characteristics. *Bull. Amer. Meteor. Soc.* **93**(10): 1497–1502, doi: 10.1175/BAMS-D-11-00164.1

Daloz, A. S., F. Chauvin, K. Walsh, S. Lavender, D. Abbs, and F. Roux 2012: The ability of general circulation models to simulate tropical cyclones and their precursors over the North Atlantic main development region. *Climate Dyn.* **39**(7-8): 1559–1576, doi: 10.1007/s00382-012-1290-7

Dare, R. A. and N. E. Davidson 2004: Characteristics of tropical cyclones in the Australian region. *Mon. Wea. Rev.* **132**(12): 3049–3065, doi: 10.1175/MWR2834.1

de Rooy, W. C. and K. Kok 2004: A combined physical-statistical approach for the downscaling of model wind speed. *Wea. Forecasting* **19**(3): 485–495

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Holm, L. Isaksen, P. Kallberg, M. Koehler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J. . J. Morcrette, B. . K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J. . N. Thepaut, and F. Vitart 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.* **137**(656): 553–597, doi: 10.1002/qj.828

DeGaetano, A. T., M. E. Hirsch, and S. J. Colucci 2002: Statistical prediction of seasonal East Coast winter storm frequency. *J. Climate* **15**(10): 1101–1117, doi: 10.1175/ 1520-0442(2002)015⟨1101:SPOSEC⟩2.0.CO;2

Della-Marta, P. M., M. A. Liniger, C. Appenzeller, D. N. Bresch, P. Kollner-Heck, and V. Muccione 2010: Improved Estimates of the European Winter Windstorm Climate and the Risk of Reinsurance Loss Using Climate Model Data. *Journal of Applied Meteorology and Climatology* **49**(10): 2092–2120, doi: 10.1175/2010JAMC2133.1

Dixon, M. and G. Wiener 1993: Titan - Thunderstorm Identification, Tracking, Analysis, and Nowcasting - A Radar-based Methodology. *J. Atmos. Oceanic Technol.* **10**(6): 785–797, doi: 10.1175/1520-0426(1993)010⟨0785:TTITAA⟩2.0.CO;2

Doblas-Reyes, F. J., I. Andreu-Burillo, Y. Chikamoto, J. Garcia-Serrano, V. Guemas, M. Kimoto, T. Mochizuki, L. R. L. Rodrigues, and G. J. van Oldenborgh 2013: Initialized near-term regional climate change prediction. *Nat. Commun.* **4**: 1715, doi: 10.1038/ncomms2704

Donat, M. G., G. C. Leckebusch, S. Wild, and U. Ulbrich 2011a: Future changes in European winter storm losses and extreme wind speeds inferred from GCM and RCM multi-model simulations. *Nat. Hazard Earth Sys.* **11**(5): 1351–1370, doi: 10.5194/ nhess-11-1351-2011

Donat, M. G., D. Renggli, S. Wild, L. V. Alexander, G. C. Leckebusch, and U. Ulbrich 2011b: Reanalysis suggests long-term upward trends in European storminess since 1871. *Geophys. Res. Lett.* **38**: L14703, doi: 10.1029/2011GL047995

Dukes, M. D. G. and J. P. Palutikof 1995: Estimation of extreme wind speeds with very long return periods. *J. Appl. Meteor.* **34**: 1950–1961

Eade, R., E. Hamilton, D. M. Smith, R. J. Graham, and A. A. Scaife 2012: Forecasting the number of extreme daily events out to a decade ahead. *J. Geophys. Res.-Atmos.* **117**: D21110, doi: 10.1029/2012JD018015

Eade, R., D. Smith, A. Scaife, E. Wallace, N. Dunstone, L. Hermanson, and N. Robinson 2014: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.* **41**(15): 5620–5628, doi: 10.1002/2014GL061146

Ebert, E. E. and J. L. McBride 2000: Verification of precipitation in weather systems: determination of systematic errors. *J. Hydrol.* **239**(1-4): 179–202, doi: 10.1016/S0022-1694(00)00343-7

Emanuel, K. 2000: A statistical analysis of tropical cyclone intensity. *Mon. Wea. Rev.* **128**(4): 1139–1152, doi: 10.1175/1520-0493(2000)128⟨1139:ASAOTC⟩2.0.CO;2

Emanuel, K. 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature* **436**: 686–688

Etienne, C. and M. Beniston 2012: Wind storm loss estimations in the Canton of Vaud (Western Switzerland). *Nat. Hazard Earth Sys.* **12**(12): 3789–3798, doi: 10.5194/nhess-12-3789-2012

Ferro, C. A. T. 2007: Comparing Probabilistic forecasting systems with the brier score. *Wea. Forecasting* **22**(5): 1076–1088, doi: 10.1175/WAF1034.1

Ferro, C. A. T., D. S. Richardson, and A. P. Weigel 2008: On the effect of ensemble size on the discrete and continuous ranked probability scores. *Meteorol. Appl.* **15**(1): 19–24, doi: 10.1002/met.45

Fiero, A. 1991: *Histoire de la météorologie.* Denoël, Paris, France

Fink, A. H., T. Brucher, V. Ermert, A. Kruger, and J. G. Pinto 2009: The European storm Kyrill in January 2007: synoptic evolution, meteorological impacts and some considerations with respect to climate change. *Nat. Hazard Earth Sys.* **9**(2): 405–423

Fink, A. H., S. Pohle, J. G. Pinto, and P. Knippertz 2012: Diagnosing the influence of diabatic processes on the explosive deepening of extratropical cyclones. *Geophys. Res. Lett.* **39**: L07803, doi: 10.1029/2012GL051025

Frankignoul, C. 1985: Sea surface temperature anomalies, planetary waves and air-sea feedback in the middle latitudes. *Rev. Geophys.* **23**(4): 357–390, doi: 10.1029/RG023i004p00357

Froude, L. S. R. 2009: Regional Differences in the Prediction of Extratropical Cyclones by the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.* **137**(3): 893–911, doi: 10.1175/2008MWR2610.1

Froude, L. S. R. 2010: TIGGE: Comparison of the Prediction of Northern Hemisphere Extratropical Cyclones by Different Ensemble Prediction Systems. *Wea. Forecasting* **25**(3): 819–836, doi: 10.1175/2010WAF2222326.1

Froude, L. S. R. 2011: TIGGE: Comparison of the Prediction of Southern Hemisphere Extratropical Cyclones by Different Ensemble Prediction Systems. *Wea. Forecasting* **26**(3): 388–398, doi: 10.1175/2010WAF2222457.1

Froude, L. S. R., L. Bengtsson, and K. I. Hodges 2007a: The predictability of extratropical storm tracks and the sensitivity of their prediction to the observing system. *Mon. Wea. Rev.* **135**(2): 315–333, doi: 10.1175/MWR3274.1

Froude, L. S. R., L. Bengtsson, and K. I. Hodges 2007b: The prediction of extratropical storm tracks by the ECMWF and NCEP ensemble prediction systems. *Mon. Wea. Rev.* **135**(7): 2545–2567, doi: 10.1175/MWR3422.1

Galarneau, T. J., Jr., C. A. Davis, and M. A. Shapiro 2013: Intensification of Hurricane Sandy (2012) through Extratropical Warm Core Seclusion. *Mon. Wea. Rev.* **141**(12): 4296–4321, doi: 10.1175/MWR-D-13-00181.1

Gangstø, R., A. P. Weigel, M. A. Liniger, and C. Appenzeller 2013: Methodological aspects of the validation of decadal predictions. *Climate Res.* **55**(3): 181–200, doi: 10.3354/cr01135

Garcia-Serrano, J. and F. J. Doblas-Reyes 2012: On the assessment of near-surface global temperature and North Atlantic multi-decadal variability in the ENSEMBLES decadal hindcast. *Climate Dyn.* **39**(7-8): 2025–2040, doi: 10.1007/s00382-012-1413-1

Gatzen, C., T. Púčik, and D. Ryva 2011: Two cold-season derechoes in Europe. *Atmospheric Research* **100**(2011): 740–748, doi: 10.1016/j.atmosres.2010.11.015

Giorgetta, M. A., J. Jungclaus, C. H. Reick, S. Legutke, J. Bader, M. Böttinger, V. Brovkin, T. Crueger, M. Esch, K. Fieg, K. Glushak, V. Gayler, H. Haak, H.-D. Hollweg, T. Ilyina, S. Kinne, L. Kornblueh, D. Matei, T. Mauritsen, U. Mikolajewicz, W. Mueller, D. Notz, F. Pithan, T. Raddatz, S. Rast, R. Redler, E. Roeckner, H. Schmidt, R. Schnur, J. Segschneider, K. D. Six, M. Stockhause, C. Timmreck, J. Wegner, H. Widmann, K.-H. Wieners, M. Claussen, J. Marotzke, and B. Stevens 2013: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5. *J. Adv. Model. Earth Syst.* **5**(3): 572–597, doi: 10.1002/jame.20038

Goddard, L., A. Kumar, A. Solomon, D. Smith, G. Boer, P. Gonzalez, V. Kharin, W. Merryfield, C. Deser, S. Mason, B. Kirtman, R. Msadek, R. Sutton, E. Hawkins, T. Fricker, G. Hegerl, C. Ferro, D. Stephenson, G. Meehl, T. Stockdale, R. Burgman, A. Greene, Y. Kushnir, M. Newman, J. Carton, I. Fukumori, and T. Delworth 2013: A verification framework for interannual-to-decadal predictions experiments. *Climate Dyn.* **40**(1-2): 245–272, doi: 10.1007/s00382-012-1481-2

Goerss, J. S. 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.* **128**(4): 1187–1193, doi: 10.1175/1520-0493(2000)128⟨1187: TCTFUA⟩2.0.CO;2

Gray, S. L. and H. F. Dacre 2006: Classifying dynamical forcing mechanisms using a climatology of extratropical cyclones. *Quart. J. Roy. Meteor. Soc.* **132**(617): 1119–1137

Gray, W. M., C. W. Landsea, P. W. Mielke, and K. J. Berry 1994: Predicting Atlantic Basin Seasonal Tropical Cyclone Activity By 1 June. *Wea. Forecasting* **9**(1): 103–115, doi: 10.1175/1520-0434(1994)009⟨0103:PABSTC⟩2.0.CO;2

Grieger, J., G. C. Leckebusch, M. G. Donat, M. Schuster, and U. Ulbrich 2014: Southern Hemisphere winter cyclone activity under recent and future climate conditions in multi-model AOGCM simulations. *Int. J. Climatol.* **34**(12): 3400–3416, doi: {10.1002/joc.3917}

Gulev, S. K., O. Zolina, and S. Grigoriev 2001: Extratropical cyclone variability in the Northern Hemisphere winter from the NCEP/NCAR reanalysis data. *Climate Dyn.* **17**(10): 795–809

Haas, R. and J. G. Pinto 2012: A combined statistical and dynamical approach for downscaling large-scale footprints of European windstorms. *Geophys. Res. Lett.* **39**: L23804, doi: 10.1029/2012GL054014

Haas, R., M. Reyers, and J. G. Pinto 2015: Decadal predictability of regional-scale peak winds over Europe using the Earth System Model of the Max-Planck-Institute for Meteorology. *Meteorol. Z.* (accepted)

Hall, T. M. and A. H. Sobel 2013: On the impact angle of Hurricane Sandy's New Jersey landfall. *Geophys. Res. Lett.* **40**(10): 2312–2315, doi: 10.1002/grl.50395

Handmer, J., Y. Honda, Z. Kundzewicz, N. Arnell, G. Benito, J. Hatfield, I. Mohamed, P. Peduzzi, S. Wu, B. Sherstyukov, K. Takahashi, and Z. Yan 2012: Changes in impacts of climate extremes: human systems and ecosystems. In: C. B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, M. Tignor, and P. M. Midgley (Eds.), *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation – A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)*, 231–290, Cambridge Univ. Press, Cambridge, UK, and New York, USA

Hanley, J. and R. Caballero 2012: Objective identification and tracking of multicentre cyclones in the ERA-Interim reanalysis dataset. *Quart. J. Roy. Meteor. Soc.* **138**(664): 612–625, doi: 10.1002/qj.948

Hanlon, H. M., G. C. Hegerl, S. F. B. Tett, and D. M. Smith 2013: Can a Decadal Forecasting System Predict Temperature Extreme Indices? *J. Climate* **26**(11): 3728–3744, doi: 10.1175/JCLI-D-12-00512.1

Hartmann, D., A. Klein Tank, M. Rusticucci, L. Alexander, S. Brönnimann, Y. Charabi, F. Dentener, E. Dlugokencky, D. Easterling, A. Kaplan, B. Soden, P. Thorne, M. Wild, and P. Zhai 2013: Observations: Atmosphere and Surface. In: T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, chapter 2, 159–254, Cambridge Univ. Press, Cambridge, United Kingdom and New York, NY, USA

Hawcroft, M. K., L. C. Shaffrey, K. I. Hodges, and H. F. Dacre 2012: How much Northern Hemisphere precipitation is associated with extratropical cyclones? *Geophys. Res. Lett.* **39**: L24809, doi: 10.1029/2012GL053866

Hawkins, E., B. W. Dong, J. Robson, R. Sutton, and D. Smith 2014: The Interpretation and Use of Biases in Decadal Climate Predictions. *J. Climate* **27**(8): 2931–2947, doi: 10.1175/JCLI-D-13-00473.1

Haylock, M. R. 2011: European extra-tropical storm damage risk from a multi-model ensemble of dynamically-downscaled global climate models. *Nat. Hazard Earth Sys.* **11**(10): 2847–2857, doi: {10.5194/nhess-11-2847-2011}

Hazeleger, W., V. Guemas, B. Wouters, S. Corti, I. Andreu-Burillo, F. J. Doblas-Reyes, K. Wyser, and M. Caian 2013: Multiyear climate predictions using two initialization strategies. *Geophys. Res. Lett.* **40**(9): 1794–1798, doi: 10.1002/grl.50355

Held, H., F.-W. Gerstengarbe, T. Pardowitz, J. G. Pinto, U. Ulbrich, K. Born, M. G. Donat, M. K. Karremann, G. C. Leckebusch, P. Ludwig, K. M. Nissen, H. Österle, B. F. Prahl, P. C. Werner, D. J. Befort, and O. Burghoff 2013: Projections of global warming-induced impacts on winter storm losses in the German private household sector. *Climatic Change* **121**(2): 195–207, doi: 10.1007/s10584-013-0872-7

Hewson, T. D. and H. A. Titley 2010: Objective identification, typing and tracking of the complete life-cycles of cyclonic features at high spatial resolution. *Meteorol. Appl.* **17**(3): 355–381, doi: 10.1002/met.204

Hodges, K. 1994: A General-Method For Tracking Analysis And Its Application To Meteorological Data. *Mon. Wea. Rev.* **122**(11): 2573–2586

Hodges, K., R. Lee, and L. Bengtsson 2011: A Comparison of Extra-tropical Cyclones in Recent Re-analyses; ERA-INTERIM, NASA-MERRA, NCEP-CFSR and JRA25. *J. Climate* **24**(18): 4888–4906, doi: 10.1175/2011JCLI4097.1

Hodges, K. I. 1996: Spherical nonparametric estimators applied to the UGAMP model integration for AMIP. *Mon. Wea. Rev.* **124**(12): 2914–2932, doi: 10.1175/1520-0493(1996)124⟨2914:SNEATT⟩2.0.CO;2

Hoskins, B. J. and K. I. Hodges 2005: A new perspective on Southern Hemisphere storm tracks. *Journal of Climate* **18**(20): 4108–4129, doi: 10.1175/JCLI3570.1

Hurley, J. V. and W. R. Boos 2014: A global climatology of monsoon low-pressure systems. *Q.J.R. Meteorol. Soc.* doi: 10.1002/qj.2447. Published online first: `http://dx.doi.org/10.1002/qj.2447`

Inatsu, M. 2009: The neighbor enclosed area tracking algorithm for extratropical wintertime cyclones. *Atmos. Sci. Let.* **10**(4): 267–272, doi: 10.1002/asl.238

International CLIVAR Project Office 2011: Data and Bias Correction for Decadal Climate Predictions. published online. URL `http://www.wcrp-climate.org/decadal/references/DCPP_Bias_Correction.pdf`, compiled by CMIP-WGCM-WGSIP Decadal Climate Prediction Panel

IPCC 2012: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, UK, and New York, USA

Jung, T., E. Klinker, and S. Uppala 2004: Reanalysis and reforecast of three major European storms of the twentieth century using the ECMWF forecasting system. Part I: Analyses and deterministic forecasts. *Meteorol. Appl.* **11**(4): 343–361

Jung, T., E. Klinker, and S. Uppala 2005: Reanalysis and reforecast of three major European storms of the twentieth century using the ECMWF forecasting system. Part II: Ensemble forecasts. *Meteorol. Appl.* **12**(2): 111–122

Jungclaus, J. H., N. Fischer, H. Haak, K. Lohmann, J. Marotzke, D. Matei, U. Mikolajewicz, D. Notz, and J. S. von Storch 2013: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component

of the MPI-Earth system model. *J. Adv. Model. Earth Syst.* **5**(2): 422–446, doi: 10.1002/jame.20023

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.* **77**(3): 437–471, doi: 10.1175/1520-0477(1996)077⟨0437:TNYRP⟩2.0.CO;2

Kantha, L. 2006: Time to replace the Saffir-Simpson hurricane scale? *Eos Trans. AGU* **87**(1): 3–6, doi: 10.1029/2006EO010003

Kew, S. F., M. Sprenger, and H. C. Davies 2010: Potential Vorticity Anomalies of the Lowermost Stratosphere: A 10-Yr Winter Climatology. *Mon. Wea. Rev.* **138**(4): 1234–1249, doi: 10.1175/2009MWR3193.1

Kharin, V. V., G. J. Boer, W. J. Merryfield, J. F. Scinocca, and W. . S. Lee 2012: Statistical adjustment of decadal predictions in a changing climate. *Geophys. Res. Lett.* **39**: L19705, doi: 10.1029/2012GL052647

Klawa, M. and U. Ulbrich 2003: A model for the estimation of storm losses and the identification of severe winter storms in Germany. *Nat. Hazard Earth Sys.* **3**(6): 725–732

Knaff, J. A., S. P. Longmore, and D. A. Molenar 2014: An Objective Satellite-Based Tropical Cyclone Size Climatology. *J. Climate* **27**(1): 455–476, doi: 10.1175/JCLI-D-13-00096.1

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann 2010: THE INTERNATIONAL BEST TRACK ARCHIVE FOR CLIMATE STEWARD-SHIP (IBTrACS) Unifying Tropical Cyclone Data. *Bull. Amer. Meteor. Soc.* **91**(3): 363–+, doi: 10.1175/2009BAMS2755.1

Kober, K. and A. Tafferner 2009: Tracking and nowcasting of convective cells using remote sensing data from radar and satellite. *Meteor. Z.* **18**(1): 75–84, doi: 10.1127/0941-2948/2009/359

Koenker, R. and K. F. Hallock 2001: Quantile regression. *Journal of Economic Perspectives* **15**(4): Allied Soc Sci Assoc, doi: 10.1257/jep.15.4.143

Köhl, A. 2015: Evaluation of the GECCO2 ocean synthesis: transports of volume, heat and freshwater in the Atlantic. *Quart. J. Roy. Meteor. Soc.* **141**(686): 166–181, doi: 10.1002/qj.2347

Kristiansen, J., S. L. Sorland, T. Iversen, D. Bjorge, and M. O. Koltzow 2011: High-resolution ensemble prediction of a polar low development. *Tellus A* **63**(3): 585–604, doi: 10.1111/j.1600-0870.2010.00498.x

Kruschke, T. 2008: *Zusammenhang zwischen verschiedenen meteorologischen Eigenschaften von Winterstürmen und resultierenden Schäden in Europa.* Bachelor's thesis, Institute of Meteorology – Dept. of Earth Sciences – Freie Universität Berlin, Berlin, Germany

Kruschke, T. 2010: *Dekadische Variabilität der Anzahl und Intensität von Winterstürmen der extratropischen Nordhemisphäre.* Master's thesis, Institute of Meteorology – Dept. of Earth Sciences – Freie Universität Berlin, Berlin, Germany

Kruschke, T., H. W. Rust, C. Kadow, G. C. Leckebusch, and U. Ulbrich 2014: Evaluating decadal predictions of northern hemispheric cyclone frequencies. *Tellus A* **66**: 22830, doi: 10.3402/tellusa.v66.22830

Kruschke, T., H. W. Rust, C. Kadow, W. A. Müller, H. Pohlmann, G. C. Leckebusch, and U. Ulbrich 2015: Probabilistic evaluation of decadal prediction skill regarding Northern Hemisphere winter storms. *Meteor. Z.* doi: 10.1127/metz/2015/0641. (early online release)

Kröger, J., W. Müller, and J.-S. von Storch 2012: Impact of different ocean reanalyses on decadal climate prediction. *Climate Dyn.* **39**: 795–810, doi: 10.1007/s00382-012-1310-7

Kurihara, Y., M. A. Bender, R. E. Tuleya, and R. J. Ross 1995: Improvements In the GFDL Hurricane Prediction System. *Mon. Wea. Rev.* **123**(9): 2791–2801, doi: 10.1175/1520-0493(1995)123⟨2791:IITGHP⟩2.0.CO;2

Kushnir, Y., W. A. Robinson, I. Blade, N. M. J. Hall, S. Peng, and R. Sutton 2002: Atmospheric GCM response to extratropical SST anomalies: Synthesis and evaluation. *J. Climate* **15**(16): 2233–2256, doi: 10.1175/1520-0442(2002)015⟨2233:AGRTES⟩2.0. CO;2

Landsberg, H. 1954: Storm of Balaklava and the Daily Weather Forecast. *The Scientific Monthly* **79**(6): 347–352

LaRow, T. E., Y. . K. Lim, D. W. Shin, E. P. Chassignet, and S. Cocke 2008: Atlantic basin seasonal hurricane simulations. *J. Climate* **21**(13): 3191–3206, doi: 10.1175/ 2007JCLI2036.1

Leckebusch, G. C., D. Renggli, and U. Ulbrich 2008a: Development and Application of an Objective Storm Severity Measure for the Northeast Atlantic Region. *Meteor. Z.* **17**(5): 575–587

Leckebusch, G. C., U. Ulbrich, L. Froehlich, and J. G. Pinto 2007: Property loss potentials for European midlatitude storms in a changing climate. *Geophys. Res. Lett.* **34**(5): L05703, doi: 10.1029/2006GL027663

Leckebusch, G. C., A. Weimer, J. G. Pinto, M. Reyers, and P. Speth 2008b: Extreme wind storms over Europe in present and future climate: a cluster analysis approach. *Meteor. Z.* **17**(1): 67–82, doi: 10.1127/0941-2948/2008/0266

Lim, E.-P. and I. Simmonds 2007: Southern Hemisphere winter extratropical cyclone characteristics and vertical organization observed with the ERA-40 data in 1979-2001. *J. Climate* **20**(11): 2675–2690, doi: 10.1175/JCLI4135.1

Lindgrén, S. and J. Neumann 1980: Great Historical Events That Were Significantly Affected by the Weather: 5, Some Meteorological Events of the Crimean War and Their Consequences. *Bull. Amer. Meteor. Soc.* **61**, doi: 10.1175/1520-0477(1980)061⟨1570: GHETWS⟩2.0.CO;2

Lionello, P., F. Dalan, and E. Elvini 2002: Cyclones in the Mediterranean region: the present and the doubled CO2 climate scenarios. *Climate Res.* **22**(2): 147–159, doi: 10.3354/cr022147

Luksch, U., C. Raible, R. Blender, and K. Fraedrich 2005: Decadal cyclone variability in the North Atlantic. *Meteor. Z.* **14**(6): 747–753, doi: 10.1127/0941-2948/2005/0075

Magnusson, L., J.-R. Bidlot, S. T. K. Lang, A. Thorpe, N. Wedi, and M. Yamaguchi 2014: Evaluation of Medium-Range Forecasts for Hurricane Sandy. *Mon. Wea. Rev.* **142**(5): 1962–1981, doi: 10.1175/MWR-D-13-00228.1

Mailier, P. J., D. B. Stephenson, and C. A. Ferro 2006: Serial Clustering of Extratropical Cyclones. *Mon. Wea. Rev.* **134**(8): 2224–2240

Majewski, D., D. Liermann, P. Prohl, B. Ritter, M. Buchhold, T. Hanisch, G. Paul, W. Wergen, and J. Baumgardner 2002: The operational global icosahedral-hexagonal gridpoint model GME: Description and high-resolution tests. *Mon. Wea. Rev.* **130**(2): 319–338, doi: 10.1175/1520-0493(2002)130⟨0319:TOGIHG⟩2.0.CO;2

Manganello, J. V., K. I. Hodges, J. L. Kinter, B. A. Cash, L. Marx, T. Jung, D. Achuthavarier, J. M. Adams, E. L. Altshuler, B. H. Huang, E. K. Jin, C. Stan, P. Towers, and N. Wedi 2012: Tropical Cyclone Climatology in a 10-km Global Atmospheric GCM: Toward Weather-Resolving Climate Modeling. *J. Climate* **25**(11): 3867–3893, doi: 10.1175/JCLI-D-11-00346.1

Martinez-Alvarado, O., S. L. Gray, J. L. Catto, and P. A. Clark 2012: Sting jets in intense winter North-Atlantic windstorms. *Environmental Research Letters* **7**(2): 024014, doi: 10.1088/1748-9326/7/2/024014

Martinez-Alvarado, O., S. L. Gray, P. A. Clark, and L. H. Baker 2013: Objective detection of sting jets in low-resolution datasets. *Met. Apps.* **20**(1): 41–55, doi: 10.1002/met.297

Matei, D., H. Pohlmann, J. Jungclaus, W. Müller, H. Haak, and J. Marotzke 2012: Two Tales of Initializing Decadal Climate Prediction Experiments with the ECHAM5/MPI-OM Model. *J. Climate* **25**(24): 8502–8523, doi: 10.1175/JCLI-D-11-00633.1

Meehl, G. A., L. Goddard, G. Boer, R. Burgman, G. Branstator, C. Cassou, S. Corti, G. Danabasoglu, F. Doblas-Reyes, E. Hawkins, A. Karspeck, M. Kimoto, A. Kumar, D. Matei, J. Mignot, R. Msadek, H. Pohlmann, M. Rienecker, T. Rosati, E. Schneider, D. Smith, R. Sutton, H. Teng, G. J. van Oldenborgh, G. Vecchi, and S. Yeager 2014:

Decadal Climate Prediction: An Update from the Trenches. *Bull. Amer. Meteor. Soc.* **95**(2): 243–267, doi: 10.1175/BAMS-D-12-00241.1

Meehl, G. A., L. Goddard, J. Murphy, R. J. Stouffer, G. Boer, G. Danabasoglu, K. Dixon, M. A. Giorgetta, A. M. Greene, E. Hawkins, G. Hegerl, D. Karoly, N. Keenlyside, M. Kimoto, B. Kirtman, A. Navarra, R. Pulwarty, D. Smith, D. Stammer, and T. Stockdale 2009: Decadal Prediction - Can It Be Skillful? *Bull. Amer. Meteor. Soc.* **90**(10): 1467–1485, doi: 10.1175/2009BAMS2778.1

Meehl, G. A. and H. Teng 2014: CMIP5 multi- model hindcasts for the mid-1970s shift and early 2000s hiatus and predictions for 2016-2035. *Geophys. Res. Lett.* **41**(5): 1711–1716, doi: 10.1002/2014GL059256

Mendes, D., E. P. Souza, J. A. Marengo, and M. C. D. Mendes 2010: Climatology of extratropical cyclones over the South American-southern oceans sector. *Theor. Appl. Climatol.* **100**(3-4): 239–250, doi: 10.1007/s00704-009-0161-6

Mesquita, M. D. S., D. E. Atkinson, and K. I. Hodges 2010: Characteristics and Variability of Storm Tracks in the North Pacific, Bering Sea, and Alaska. *J. Climate* **23**(2): 294–311

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.* **122**(529): 73–119, doi: 10.1002/qj.49712252905

Monmonier, M. 1999: *Air Apparent – How meteorologists learned to map, predict, and dramatize weather*. The University of Chicago Press, Chicago (USA) and London (UK)

Mori, M., M. Kimoto, M. Ishii, S. Yokoi, T. Mochizuki, Y. Chikamoto, M. Watanabe, T. Nozawa, H. Tatebe, T. T. Sakamoto, Y. Komuro, Y. Imada, and H. Koyama 2013: Hindcast Prediction and Near-Future Projection of Tropical Cyclone Activity over the Western North Pacific Using CMIP5 Near-Term Experiments with MIROC. *J. Meteor. Soc. Japan* **91**(4): 431–452, doi: 10.2151/jmsj.2013-402

Mueller, C., T. Saxen, R. Roberts, J. Wilson, T. Betancourt, S. Dettling, N. Oien, and J. Yee 2003: NCAR Auto-Nowcast System. *Wea. Forecasting* **18**(4): 545–561, doi: 10.1175/1520-0434(2003)018⟨0545:NAS⟩2.0.CO;2

Munich RE Group 2001: Winterstürme in Europa: Schadenanalyse 1999 - Schaden-potenziale. Publication of Munich Re, No. 302-03108, `www.munichre.com`

Munich RE Group 2007: Zwischen Hoch und Tief - Wetterrisiken in Mitteleuropa. URL `www.munichre.com`, veröffentlichung der Münchener Rück, `www.munichre.com`, Bestell.-Nr. 302-05481

Munich RE Group 2008a: Knowledge series: Highs and lows – Weather risks in central Europe. published online, e.g. `http://www.mroc.com/publications.html`

Munich RE Group 2008b: Schadenspiegel 1/2008 – Special feature issue: Risk factor of air. published online, e.g. `http://www.mroc.com/publications.html`

Munich RE Group 2013: Topics Geo: Naturkatastrophen 2012 – Analysen, Bewertungen, Positionen. published online via `https://www.munichre.com/de/reinsurance/magazine/topics-online/2013/02/natural-catastrophes`, Munich, Germany

Munich RE Group 2014a: NatCatSERVICE – Schadenereignisse in Europa 1980–2013 – 10 teuerste Winterstürme für die Gesamtwirtschaft. Published online via `https://www.munichre.com/touch/naturalhazards/de/natcatservice/significant-natural-catastrophes`

Munich RE Group 2014b: NatCatSERVICE – Schadenereignisse weltweit 1980–2013. Published online via `https://www.munichre.com/touch/naturalhazards/de/natcatservice/focus-analyses`

Murphy, A. H. 1988: Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. *Mon. Wea. Rev.* **116**(12): 2417–2425, doi: 10.1175/1520-0493(1988)116⟨2417:SSBOTM⟩2.0.CO;2

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. *Aust. Meteor. Mag.* **39**: 155–166

Müller, W. A., J. Baehr, H. Haak, J. H. Jungclaus, J. Kröger, D. Matei, D. Notz, H. Pohlmann, J. S. von Storch, and J. Marotzke 2012: Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology. *Geophys. Res. Lett.* **39**: L22707, doi: 10.1029/2012GL053326

Müller, W. A., H. Pohlmann, F. Sienz, and D. Smith 2014: Decadal climate predictions for the period 1901-2010 with a coupled climate model. *Geophys. Res. Lett.* **41**(6): 2100–2107, doi: 10.1002/2014GL059259

Neu, U., M. G. Akperov, N. Bellenbaum, R. S. Benestad, R. Blender, R. Caballero, A. Cocozza, H. F. Dacre, Y. Feng, K. Fraedrich, J. Grieger, S. Gulev, J. Hanley, T. Hewson, M. Inatsu, K. Keay, S. F. Kew, I. Kindem, G. C. Leckebusch, M. L. R. Liberato, P. Lionello, I. I. Mokhov, J. G. Pinto, C. C. Raible, M. Reale, I. Rudeva, M. Schuster, I. Simmonds, M. Sinclair, M. Sprenger, N. D. Tilinina, I. F. Trigo, S. Ulbrich, U. Ulbrich, X. L. Wang, and H. Wernli 2013: IMILAST A Community Effort to Intercompare Extratropical Cyclone Detection and Tracking Algorithms. *Bull. Amer. Meteor. Soc.* **94**(4): 529–547, doi: 10.1175/BAMS-D-11-00154.1

Nissen, K., U. Ulbrich, G. Leckebusch, and I. Kuhnel 2014a: Decadal windstorm activity in the North Atlantic-European sector and its relationship to the meridional overturning circulation in an ensemble of simulations with a coupled climate model. *Climate Dyn.* **43**: 1545–1555, doi: 10.1007/s00382-013-1975-6

Nissen, K. M., G. C. Leckebusch, J. G. Pinto, D. Renggli, S. Ulbrich, and U. Ulbrich 2010: Cyclones causing wind storms in the Mediterranean: characteristics, trends and links to large-scale patterns. *Nat. Hazard Earth Sys.* **10**(7): 1379–1391, doi: 10.5194/nhess-10-1379-2010

Nissen, K. M., G. C. Leckebusch, J. G. Pinto, and U. Ulbrich 2014b: Mediterranean cyclones and windstorms in a changing climate. *Reg. Environ. Change* **14**: 1873–1890, doi: 10.1007/s10113-012-0400-8

Orlowsky, B., F.-W. Gerstengarbe, and P. C. Werner 2008: A resampling scheme for regional climate simulations and its performance compared to a dynamical RCM. *Theor. Appl. Climatol.* **92**(3-4): 209–223, doi: 10.1007/s00704-007-0352-y

Osinski, R. D. 2014: *Untersuchungen zu Atlantisch-Europäischen Winterstürmen anhand des operationellen Ensemblevorhersagesystems des EZMW*. Ph.D. thesis, Department of Earth Sciences, Freie Universität Berlin, Berlin, Germany

Osinski, R. D., P. Lorenz, T. Kruschke, M. Voigt, U. Ulbrich, G. C. Leckebusch, E. Faust, T. Hofherr, and D. Majewski 2015: An approach to build an event set of European

wind storms based on ECMWF EPS. *Nat. Hazard Earth Sys. Discuss.* **3**: 1231–1268, doi: 10.5194/nhessd-3-1231-2015

Palutikof, J. and A. Skellern 1991: Storm Severity over Britain. Report to Commercial Union General Insurance, Climatic Research Unit, School of Environmental Studies, University of East Anglia, Norwich, UK

Palutikof, J. P., B. B. Brabson, D. H. Lister, and S. T. Adcock 1999: A review of methods to calculate extreme wind speeds. *Meteor. Appl.* **6**: 119–132

Pantillon, F. P., J.-P. Chaboureau, P. J. Mascart, and C. Lac 2013: Predictability of a Mediterranean Tropical-Like Storm Downstream of the Extratropical Transition of Hurricane Helene (2006). *Mon. Wea. Rev.* **141**(6): 1943–1962, doi: 10.1175/MWR-D-12-00164.1

Papritz, L., S. Pfahl, I. Rudeva, I. Simmonds, H. Sodemann, and H. Wernli 2014: The Role of Extratropical Cyclones and Fronts for Southern Ocean Freshwater Fluxes. *J. Climate* **27**(16): 6205–6224, doi: 10.1175/JCLI-D-13-00409.1

Pardowitz, T. 2014: *Anthropogenic Changes in the Frequency and Severity of European Winter Storms: Mechanisms, Impacts and their Uncertainties.* Ph.D. thesis, Department of Earth Sciences, Freie Universität Berlin, Berlin, Germany

Pardowitz, T., D. J. Befort, G. C. Leckebusch, and U. Ulbrich 2014: Estimating uncertainties from high resolution simulations of extreme wind storms and consequences for impacts. *Meteor. Z.* (submitted)

Paredes, D., R. M. Trigo, R. Garcia-Herrera, and I. F. Trigo 2006: Understanding precipitation changes in Iberia in early spring: Weather typing and storm-tracking approaches. *J. Hydrometeorol.* **7**(1): 101–113, doi: 10.1175/JHM472.1

Pearson, K. 1896: Contributions to the Mathematical Theory of Evolution. III. Regression, Heredity, and Panmixia. *Philosophical Transactions of the Royal Society of London. Series A* **187**: 253–318

Peterson, T. C., D. M. Anderson, S. J. Cohen, M. Cortez-Vázquez, R. J. Murnane, C. Parmesan, D. Phillips, R. S. Pulwarty, and J. M. R. Stone 2008: Why Weather and Climate Extremes Matter. In: T. R. Karl, G. A. Meehl, C. D. Miller, S. J. Hassol,

A. M. Waple, and W. L. Murray (Eds.), *Weather and Climate Extremes in a Changing Climate. Regions of Focus: North America, Hawaii, Caribbean, and U.S. Pacific Islands*, A Report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research 1, Washington, D.C. (USA)

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

Pielke, J., R., J. Gratz, C. Landsea, D. Collins, M. Saunders, and R. Musulin 2008: Normalized Hurricane Damage in the United States: 1900–2005. *Nat. Hazards Rev.* **9**(1): 29–42, doi: 10.1061/(ASCE)1527-6988(2008)9:1(29)

Pinto, J. G., N. Bellenbaum, M. K. Karremann, and P. M. Della-Marta 2013: Serial clustering of extratropical cyclones over the North Atlantic and Europe under recent and future climate conditions. *J. Geophys. Res.-Atmos.* **118**(22): 12476–12485, doi: 10.1002/2013JD020564

Pinto, J. G., E. L. Frohlich, G. C. Leckebusch, and U. Ulbrich 2007: Changing European storm loss potentials under modified climate conditions according to ensemble simulations of the ECHAM5/MPI-OM1 GCM. *Nat. Hazard Earth Sys.* **7**(1): 165–175

Pinto, J. G., M. K. Karremann, K. Born, P. M. Della-Marta, and M. Klawa 2012: Loss potentials associated with European windstorms under future climate conditions. *Climate Res.* **54**(1): 1–20, doi: 10.3354/cr01111

Pinto, J. G., C. P. Neuhaus, G. C. Leckebusch, M. Reyers, and M. Kerschgens 2010: Estimation of wind storm impacts over Western Germany under future climate conditions using a statistical-dynamical downscaling approach. *Tellus Series A-dynamic Meteorology and Oceanography* **62**(2): 188–201, doi: 10.1111/j.1600-0870.2009.00424.x

Pohlmann, H., W. A. Müller, K. Kulkarni, M. Kameswarrao, D. Matei, F. Vamborg, C. Kadow, S. Illing, and J. Marotzke 2013a: Improved forecast skill in the tropics in the new MiKlip decadal climate predictions. *Geophys. Res. Lett.* **40**: 5798–5802, doi: doi:10.1002/2013GL058

Pohlmann, H., D. M. Smith, M. A. Balmaseda, N. S. Keenlyside, S. Masina, D. Matei, W. A. Müller, and P. Rogel 2013b: Predictability of the mid-latitude Atlantic merid-

ional overturning circulation in a multi-model system. *Climate Dyn.* **41**(3-4): 775–785, doi: 10.1007/s00382-013-1663-6

Polkova, I., A. Köhl, and D. Stammer 2014: Impact of initialization procedures on the predictive skill of a coupled ocean-atmosphere model. *Climate Dyn.* **42**(11–12): 3151–3169, doi: 10.1007/s00382-013-1969-4

Powell, M. D. and T. A. Reinhold 2007: Tropical Cyclone Destructive Potential by Integrated Kinetic Energy. *Bull. Amer. Meteor. Soc.* **88 (4)**: 513–526

Pryor, S. C., J. T. Schoof, and R. J. Barthelmie 2005: Empirical downscaling of wind speed probability distributions. *J. Geophys. Res.-Atmos.* **110**(D19): D19109, doi: 10.1029/2005JD005899

Raible, C. C. 2007: On the relation between extremes of midlatitude cyclones and the atmospheric circulation using ERA40. *Geophys. Res. Lett.* **34**(7): L07703, doi: 10.1029/2006GL029084

Raible, C. C., M. Yoshimori, T. F. Stocker, and C. Casty 2007: Extreme midlatitude cyclones and their implications for precipitation and wind speed extremes in simulations of the Maunder Minimum versus present day conditions. *Climate Dyn.* **28**(4): 409–423

Ramsay, H. A., S. J. Camargo, and D. Kim 2012: Cluster analysis of tropical cyclone tracks in the Southern Hemisphere. *Climate Dyn.* **39**(3-4): 897–917, doi: 10.1007/s00382-011-1225-8

Rappaport, E. N., J. L. Franklin, L. A. Avila, S. R. Baig, J. L. Beven, II, E. S. Blake, C. A. Burr, J.-G. Jiing, C. A. Juckins, R. D. Knabb, C. W. Landsea, M. Mainelli, M. Mayfield, C. J. McAdie, R. J. Pasch, C. Sisko, S. R. Stewart, and A. N. Tribble 2009: Advances and Challenges at the National Hurricane Center. *Wea. Forecasting* **24**(2): 395–419, doi: 10.1175/2008WAF2222128.1

Ray, S. and B. S. Giese 2012: Historical changes in El Niño and La Niña characteristics in an ocean reanalysis. *J. Geophys. Res.-Oceans* **117**: C11007, doi: 10.1029/2012JC008031

Renggli, D. 2011: *Seasonal predictability of wintertime windstorm climate over the North Atlantic and Europe.* Ph.D. thesis, Department of Earth Sciences, Freie Universität Berlin, Berlin, Germany

Renggli, D., G. C. Leckebusch, U. Ulbrich, and E. Faust 2014: Potential sources of wintertime windstorm predictability over the North Atlantic and Europe on seasonal time scales. *Int. J. Climatol.* Currently under revision

Renggli, D., G. C. Leckebusch, U. Ulbrich, S. N. Gleixner, and E. Faust 2011: The Skill of Seasonal Ensemble Prediction Systems to Forecast Wintertime Windstorm Frequency over the North Atlantic and Europe. *Mon. Wea. Rev.* **139**(9): 3052–3068, doi: 10.1175/2011MWR3518.1

Reyers, M., J. G. Pinto, and J. Moemken 2015: Statistical-dynamical downscaling for wind energy potentials: evaluation and applications to decadal hindcasts and climate change projections. *Int. J. Climatol.* **35**(2): 229–244, doi: 10.1002/joc.3975

Roberts, J. F., A. J. Champion, L. C. Dawkins, K. I. Hodges, L. C. Shaffrey, D. B. Stephenson, M. A. Stringer, H. E. Thornton, and B. D. Youngman 2014: The XWS open access catalogue of extreme European windstorms from 1979 to 2012. *Nat. Hazard Earth Sys.* **14**(9): 2487–2501, doi: 10.5194/nhess-14-2487-2014

Robson, J., R. Sutton, and D. Smith 2014: Decadal predictions of the cooling and freshening of the North Atlantic in the 1960s and the role of ocean circulation. *Climate Dyn.* **42**(9-10): 2353–2365, doi: 10.1007/s00382-014-2115-7

Robson, J. I., R. T. Sutton, and D. M. Smith 2012: Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s. *Geophys. Res. Lett.* **39**: L19713, doi: 10.1029/2012GL053370

Rockel, B. and K. Woth 2007: Extremes of near-surface wind speed over Europe and their future changes as estimated from an ensemble of RCM simulations. *Climatic Change* **81**: 267–280, doi: 10.1007/s10584-006-9227-y

Rudeva, I. and S. K. Gulev 2007: Climatology of Cyclone Size Characteristics and Their Changes during the Cyclone Life Cycle. *Mon. Wea. Rev.* **135**: 2568–2587

Sakuth, O. 2011: Objektive Identifikation des schadenträchtigen Windfeldes tropischer Zyklonen. Bachelor's thesis, Institute of Meteorology – Dept. of Earth Sciences – Freie Universität Berlin, Berlin, Germany

Salameh, T., P. Drobinski, M. Vrac, and P. Naveau 2009: Statistical downscaling of near-surface wind over complex terrain in southern France. *Meteor. Atmos. Phys.* **103**(1-4): 253–265, doi: 10.1007/s00703-008-0330-7

Scaife, A. A., M. Athanassiadou, M. Andrews, A. Arribas, M. Baldwin, N. Dunstone, J. Knight, C. MacLachlan, E. Manzini, W. A. Müller, H. Pohlmann, D. Smith, T. Stockdale, and A. Williams 2014: Predictability of the quasi-biennial oscillation and its northern winter teleconnection on seasonal to decadal timescales. *Geophys. Res. Lett.* **41**: 1752–1758, doi: 10.1002/2013GL059160

Schenkel, B. A. and R. E. Hart 2012: An Examination of Tropical Cyclone Position, Intensity, and Intensity Life Cycle within Atmospheric Reanalysis Datasets. *J. Climate* **25**(10): 3453–3475, doi: 10.1175/2011JCLI4208.1

Schneidereit, A., R. Blender, and K. Fraedrich 2010: A radius-depth model for mid-latitude cyclones in reanalysis data and simulations. *Quart. J. Roy. Meteor. Soc.* **136**(646): 50–60, doi: 10.1002/qj.523

Schreck, C. J., III, K. R. Knapp, and J. P. Kossin 2014: The Impact of Best Track Discrepancies on Global Tropical Cyclone Climatologies using IBTrACS. *Mon. Wea. Rev.* **142**(10): 3881–3899, doi: 10.1175/MWR-D-14-00021.1

Schulz, J.-P. 2008: Revision of the Turbulent Gust Diagnostics in the COSMO Model. *COSMO Newsletter* **8**: 17–22

Schwierz, C., P. Koellner-Heck, E. Z. Mutter, D. N. Bresch, P.-L. Vidale, M. Wild, and C. Schaer 2010: Modelling European winter wind storm losses in current and future climate. *Climatic Change* **101**(3-4): 485–514, doi: {10.1007/s10584-009-9712-1}

Serreze, M. C. 1995: Climatological Aspects of Cyclone Development and Decay In the Arctic. *Atmos.-Ocean* **33**(1): 1–23

Simmonds, I., K. Keay, and J. A. T. Bye 2012: Identification and Climatology of Southern Hemisphere Mobile Fronts in a Modern Reanalysis. *J. Climate* **25**(6): 1945–1962, doi: 10.1175/JCLI-D-11-00100.1

Sinclair, M. R. 1994: An Objective Cyclone Climatology For the Southern-hemisphere. *Mon. Wea. Rev.* **122**(10): 2239–2256, doi: 10.1175/1520-0493(1994)122⟨2239: AOCCFT⟩2.0.CO;2

Smith, D. M., R. Eade, N. J. Dunstone, D. Fereday, J. M. Murphy, H. Pohlmann, and A. A. Scaife 2010: Skilful multi-year predictions of Atlantic hurricane frequency. *Nat. Geosci.* **3**(12): 846–849, doi: 10.1038/NGEO1004

Smith, D. M., R. Eade, and H. Pohlmann 2013: A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate Dyn.* **41**(11-12): 3325–3338, doi: 10.1007/s00382-013-1683-2

Smith, R. B. 1979: The influence of mountains on the atmosphere. *Adv. Geophys.* **21**: 87–230

Smith, R. B. 1985: On Severe Downslope Winds. *J. Atmos. Sci.* **42**(23): 2597–2603, doi: 10.1175/1520-0469(1985)042⟨2597:OSDW⟩2.0.CO;2

Solomon, A., L. Goddard, A. Kumar, J. Carton, C. Deser, I. Fukumori, A. M. Greene, G. Hegerl, B. Kirtman, Y. Kushnir, M. Newman, D. Smith, D. Vimont, T. Delworth, G. A. Meehl, and T. Stockdale 2011: Distinguishing the Roles of Natural and Anthropogenically Forced Decadal Climate Variability. *Bull. Amer. Meteor. Soc.* **92**(2): 141–156, doi: 10.1175/2010BAMS2962.1

Steppeler, J., G. Doms, and G. Adrian 2002: Das Lokal-Modell LM. *promet* **27**: 123–128

Stevens, B., M. Giorgetta, M. Esch, T. Mauritsen, T. Crueger, S. Rast, M. Salzmann, H. Schmidt, J. Bader, K. Block, R. Brokopf, I. Fast, S. Kinne, L. Kornblueh, U. Lohmann, R. Pincus, T. Reichler, and E. Roeckner 2013: Atmospheric component of the MPI-M Earth System Model: ECHAM6. *J. Adv. Model. Earth Syst.* **5**(2): 146–172, doi: 10.1002/jame.20015

Taylor, K. E., R. J. Stouffer, and G. A. Meehl 2012: An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Meteor. Soc.* **93**(4): 485–498, doi: 10.1175/ BAMS-D-11-00094.1

Tilinina, N., S. K. Gulev, I. Rudeva, and P. Koltermann 2013: Comparing Cyclone Life Cycle Characteristics and Their Interannual Variability in Different Reanalyses. *J. Climate* **26**(17): 6419–6438, doi: 10.1175/JCLI-D-12-00777.1

Tsai, H.-C., K.-C. Lu, R. L. Elsberry, M.-M. Lu, and C.-H. Sui 2011: Tropical Cyclone-like Vortices Detection in the NCEP 16-Day Ensemble System over the Western North Pacific in 2008: Application and Forecast Evaluation. *Wea. Forecasting* **26**(1): 77–93, doi: 10.1175/2010WAF2222415.1

Ulbrich, U., G. C. Leckebusch, J. Grieger, M. Schuster, M. Akperov, M. Y. Bardin, Y. Feng, S. Gulev, M. Inatsu, K. Keay, S. F. Kew, M. L. R. Liberato, P. Lionello, I. I. Mokhov, U. Neu, J. G. Pinto, C. C. Raible, M. Reale, I. Rudeva, I. Simmonds, N. D. Tilinina, I. F. Trigo, S. Ulbrich, X. L. Wang, and H. Wernli 2013: Are Greenhouse Gas Signals of Northern Hemisphere winter extra-tropical cyclone activity dependent on the identification and tracking algorithm? *Meteor. Z.* **22**(1): 61–68, doi: 10.1127/0941-2948/2013/0420

Ulbrich, U., G. C. Leckebusch, and J. Pinto 2009: Extra-tropical cyclones in the present and future climate: a review. *Theor. Appl. Climatol.* **96**: 117–131

Ulbrich, U., J. G. Pinto, H. Kupfer, G. C. Leckebusch, T. Spangehl, and M. Reyers 2008: Changing northern hemisphere storm tracks in an ensemble of IPCC climate change simulations. *J. Climate* **21**(8): 1669–1679, doi: 10.1175/2007JCLI1992.1

Uppala, S., P. Kållberg, A. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J. Gibson, J. Haseler, A. Hernandez, G. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. Allan, E. Andersson, K. Arpe, M. Balmaseda, A. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Holm, B. Hoskins, L. Isaksen, P. Janssen, R. Jenne, A. McNally, J.-F. Mahfouf, J.-J. Morcrette, N. Rayner, R. Saunders, P. Simon, A. Sterl, K. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, , and J. Woollen 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.* **131**: 2961–3012

van Oldenborgh, G. J., F. J. Doblas-Reyes, B. Wouters, and W. Hazeleger 2012: Decadal prediction skill in a multi-model ensemble. *Climate Dyn.* **38**(7-8): 1263–1280, doi: 10.1007/s00382-012-1313-4

Vecchi, G. A., T. Delworth, R. Gudgel, S. Kapnick, A. Rosati, A. T. Wittenberg, F. Zeng, W. Anderson, V. Balaji, K. Dixon, L. Jia, H.-S. Kim, L. Krishnamurthy, R. Msadek, W. F. Stern, S. D. Underwood, G. Villarini, X. Yang, and S. Zhang 2014: On the

Seasonal Forecasting of Regional Tropical Cyclone Activity. *J. Climate* **27**(21): 7994–8016, doi: 10.1175/JCLI-D-14-00158.1

Vecchi, G. A., R. Msadek, W. Anderson, Y. S. Chang, T. Delworth, K. Dixon, R. Gudgel, A. Rosati, B. Stern, G. Villarini, A. Wittenberg, X. Yang, F. R. Zeng, R. Zhang, and S. Q. Zhang 2013: Multiyear Predictions of North Atlantic Hurricane Frequency: Promise and Limitations. *J. Climate* **26**(15): 5337–5357, doi: 10.1175/JCLI-D-12-00464.1

Velden, C., J. Daniels, D. Stettner, D. Santek, J. Key, J. Dunion, K. Holmlund, G. Dengel, W. Bresky, and P. Menzel 2005: Recent innovations in deriving tropospheric winds from meteorological satellites. *Bull. Amer. Meteor. Soc.* **86**(2): 205–+, doi: 10.1175/BAMS-86-2-205

Vitart, F., M. R. Huddleston, M. Deque, D. Peake, T. N. Palmer, T. N. Stockdale, M. K. Davey, S. Ineson, and A. Weisheimer 2007: Dynamically-based seasonal forecasts of Atlantic tropical storm activity issued in June by EUROSIP. *Geophys. Res. Lett.* **34**(16): L16815, doi: 10.1029/2007GL030740

Vitart, F., F. Prates, A. Bonet, and C. Sahin 2012: New tropical cyclone products on the web. *ECMWF Newsletter* **130**: 17–23. Published online at `http://www.ecmwf.int/en/about/news-centre/media-resources`. Retrieved 21 Nov. 2014

Vitolo, R., D. B. Stephenson, I. M. Cook, and K. Mitchell-Wallace 2009: Serial clustering of intense European storms. *Meteor. Z.* **18**(4): 411–424, doi: 10.1127/0941-2948/2009/0393

Voigt, M. 2012: *Statistisches Downscaling von extrmen Windböen mit künstlichen Neuronalen Netzen.* Master's thesis, Institute of Meteorology – Dept. of Earth Sciences – Freie Universität Berlin, Berlin, Germany

Wang, X. L., Y. Feng, G. Compo, V. Swail, F. Zwiers, R. Allan, and P. Sardeshmukh 2013: Trends and low frequency variability of extra-tropical cyclone activity in the ensemble of twentieth century reanalysis. *Climate Dyn.* **40**(11–12): 2775–2800, doi: 10.1007/s00382-012-1450-9

Webster, P. J., G. J. Holland, J. A. Curry, and H. R. Chang 2005: Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science* **309**(5742): 1844–1846, doi: 10.1126/science.1116448

Welker, C. and E. Faust 2013: Tropical cyclone-related socio-economic losses in the western North Pacific region. *Nat. Hazard Earth Sys.* **13**(1): 115–124, doi: 10.5194/nhess-13-115-2013

Welker, C. and O. Martius 2014: Decadal-scale variability in hazardous winds in northern Switzerland since end of the 19th century. *Atmos. Sci. Let.* **15**(2): 86–91, doi: 10.1002/asl2.467

Wernli, H., S. Dirren, M. A. Liniger, and M. Zillig 2002: Dynamical aspects of the life cycle of the winter storm 'Lothar' (24-26 December 1999). *Quart. J. Roy. Meteor. Soc.* **128**(580): 405–429

Wernli, H. and C. Schwierz 2006: Surface cyclones in the ERA-40 dataset (1958-2001). Part I: Novel identification method and global climatology. *J. Atmos. Sci.* **63**(10): 2486–2507

Wieringa, J. 1986: Roughness-dependent Geographical Interpolation of Surface Windspeed Averages. *Quart. J. Roy. Meteor. Soc.* **112**(473): 867–889, doi: 10.1002/qj.49711247316

Wieringa, J. 1993: Representative Roughness Parameters For Homogeneous Terrain. *Bound.-Layer Meteor.* **63**(4): 323–363, doi: 10.1007/BF00705357

Wilks, D. S. 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd edition, Elsevier, Amsterdam

Yao, Y. H., W. Perrie, W. Q. Zhang, and J. Jiang 2008: Characteristics of atmosphere-ocean interactions along North Atlantic extratropical storm tracks. *J. Geophys. Res.-Atmos.* **113**(D14): D14124

Yeager, S., A. Karspeck, G. Danabasoglu, J. Tribbia, and H. Teng 2012: A Decadal Prediction Case Study: Late Twentieth-Century North Atlantic Ocean Heat Content. *J. Climate* **25**(15): 5173–5189, doi: 10.1175/JCLI-D-11-00595.1

Zappa, G., L. Shaffrey, and K. Hodges 2014: Can Polar Lows be Objectively Identified and Tracked in the ECMWF Operational Analysis and the ERA-Interim Reanalysis? *Mon. Wea. Rev.* **142**(8): 2596–2608, doi: 10.1175/MWR-D-14-00064.1

Zappa, G., L. C. Shaffrey, and K. I. Hodges 2013: The Ability of CMIP5 Models to Simulate North Atlantic Extratropical Cyclones. *J. Climate* **26**(15): 5379–5396, doi: 10.1175/JCLI-D-12-00501.1

Zinner, T., H. Mannstein, and A. Tafferner 2008: Cb-TRAM: Tracking and monitoring severe convection from onset over rapid development to mature phase using multi-channel Meteosat-8 SEVIRI data. *Meteor. Atmos. Phys.* **101**(3-4): 191–210, doi: 10.1007/s00703-008-0290-y

# A. Appendix

## A.1. Skill scores of individual hindcast systems

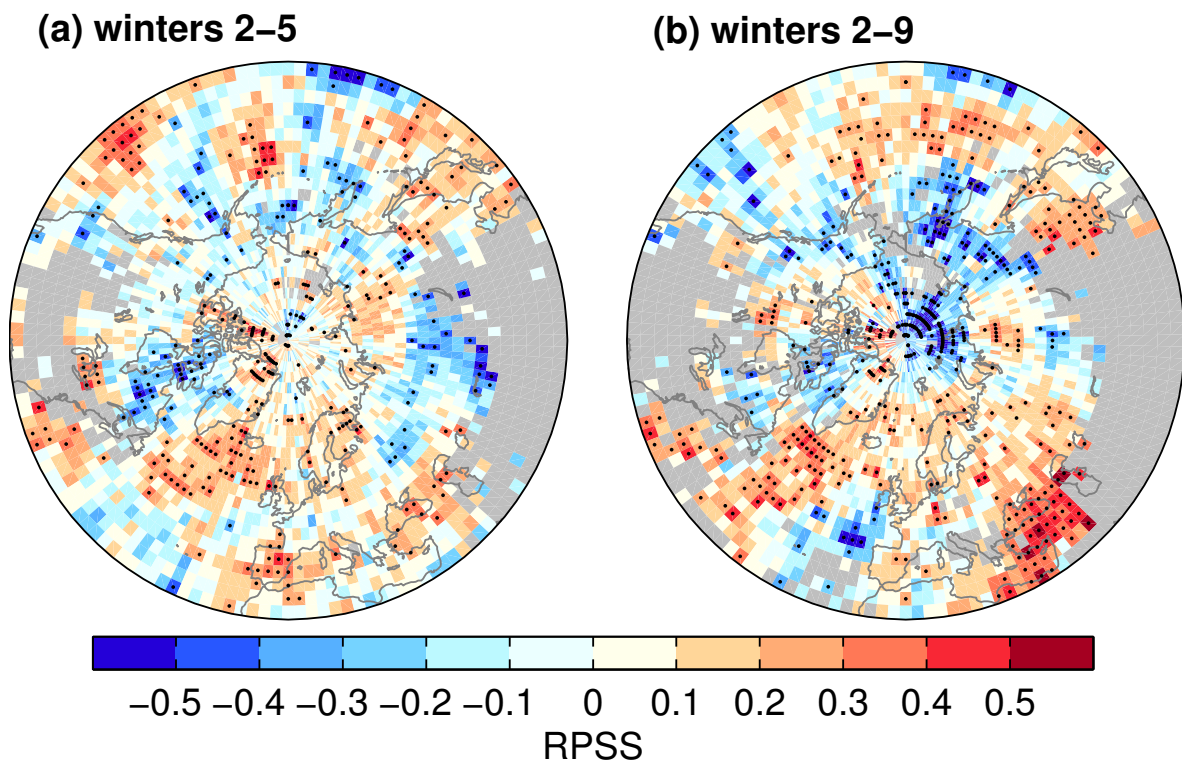**(a) winters 2–5**　　　　　**(b) winters 2–9**



Figure A.1.: **RPSS of *baseline0* over uninitialized simulations** regarding the average winter storm frequency (number of tracks per ONDJFM in the vicinity of 1000km) for (a) hindcast winters 2–5 and (b) hindcast winters 2–9 based on ERA-reanalyses as observational reference; significant skill scores ($\alpha < 5\%$) as black dots, areas of strong inconsistencies between ERA40 and ERA-Int. are masked out (grey)
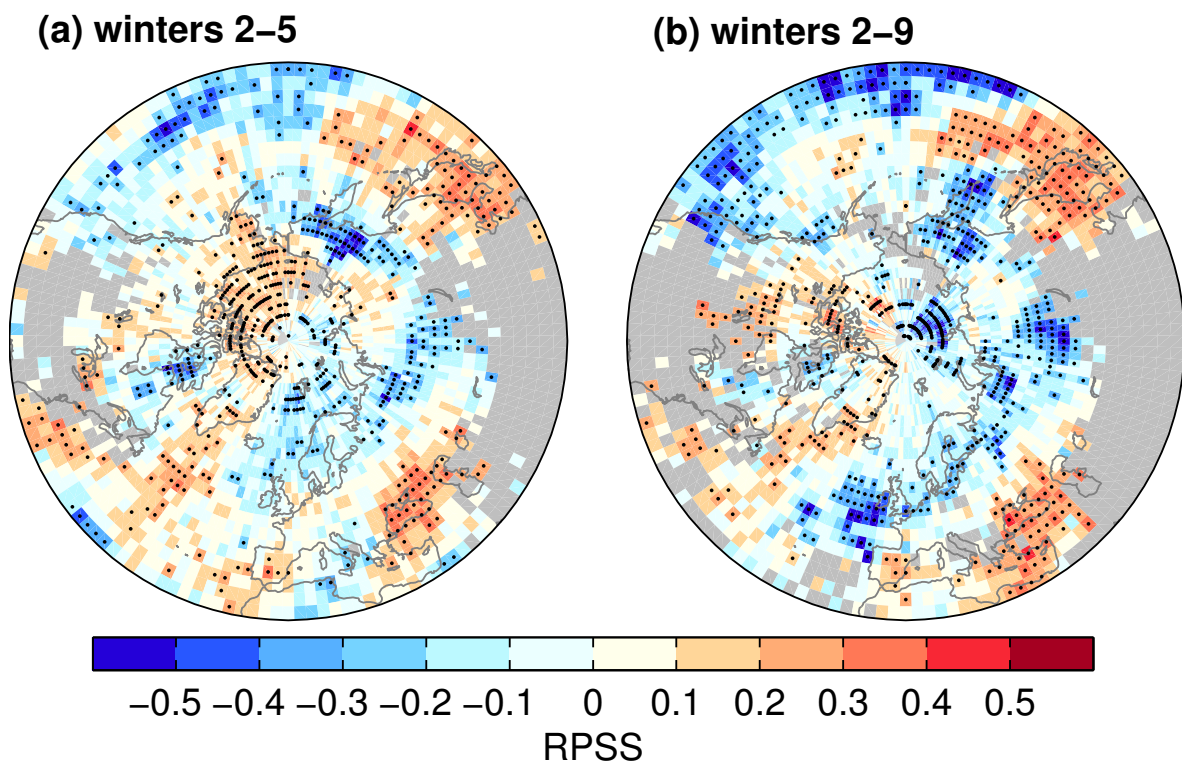
Figure A.2.: **RPSS of *baseline1* over uninitialized simulations** regarding the average winter storm frequency (number of tracks per ONDJFM in the vicinity of 1000km) for (a) hindcast winters 2–5 and (b) hindcast winters 2–9 based on ERA-reanalyses as observational reference; significant skill scores ($\alpha < 5\%$) as black dots, areas of strong inconsistencies between ERA40 and ERA-Int. are masked out (grey)
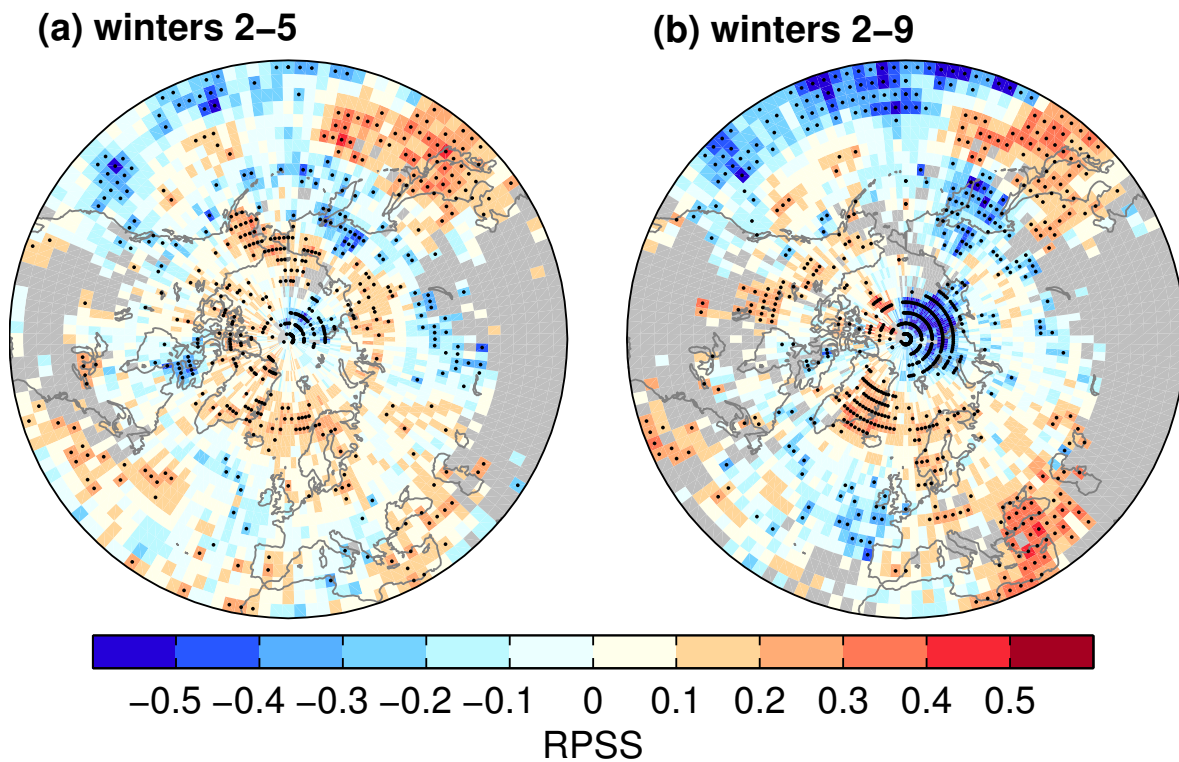
**(a) winters 2–5**  **(b) winters 2–9**



Figure A.3.: **RPSS of *ORAff* over uninitialized simulations** regarding the average winter storm frequency (number of tracks per ONDJFM in the vicinity of 1000km) for (a) hindcast winters 2–5 and (b) hindcast winters 2–9 based on ERA-reanalyses as observational reference; significant skill scores ($\alpha < 5\%$) as black dots, areas of strong inconsistencies between ERA40 and ERA-Int. are masked out (grey)

**(a) winters 2–5**　　　**(b) winters 2–9**

RPSS

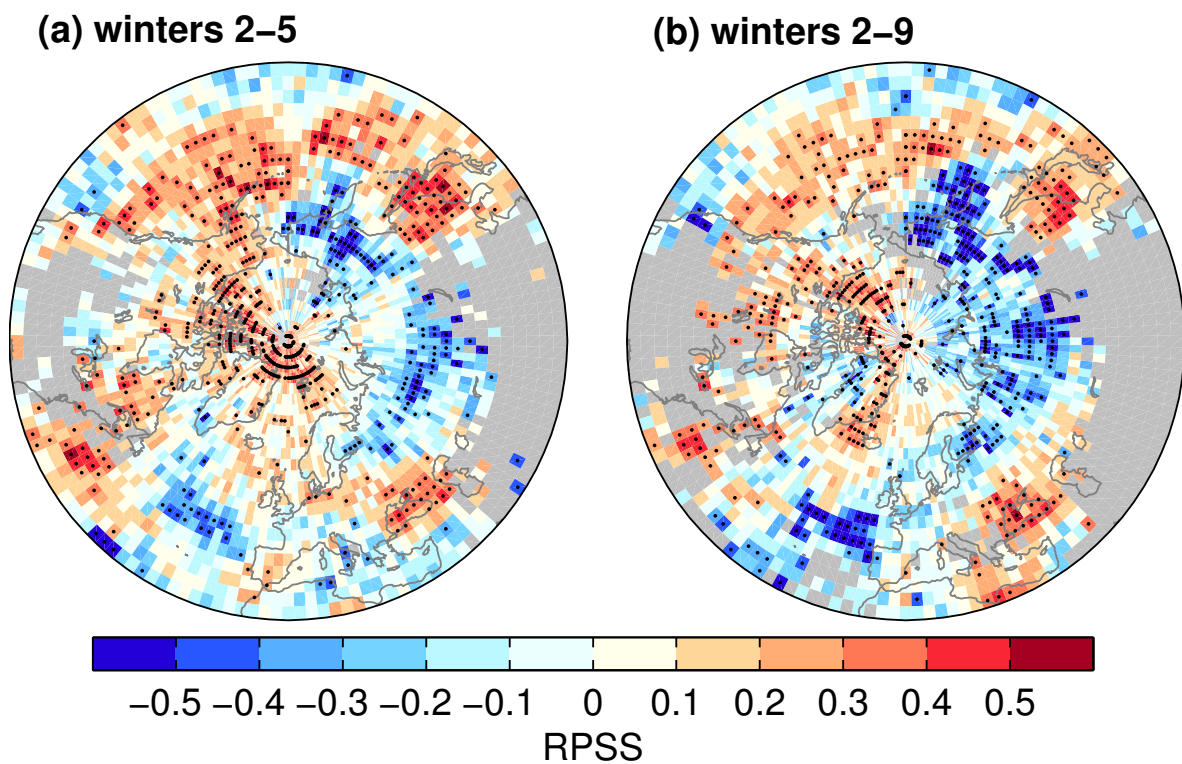−0.5 −0.4 −0.3 −0.2 −0.1　0　0.1　0.2　0.3　0.4　0.5

Figure A.4.: **RPSS of *GECCOano* over uninitialized simulations** regarding the average winter storm frequency (number of tracks per ONDJFM in the vicinity of 1000km) for (a) hindcast winters 2–5 and (b) hindcast winters 2–9 based on ERA-reanalyses as observational reference; significant skill scores ($\alpha < 5\%$) as black dots, areas of strong inconsistencies between ERA40 and ERA-Int. are masked out (grey)
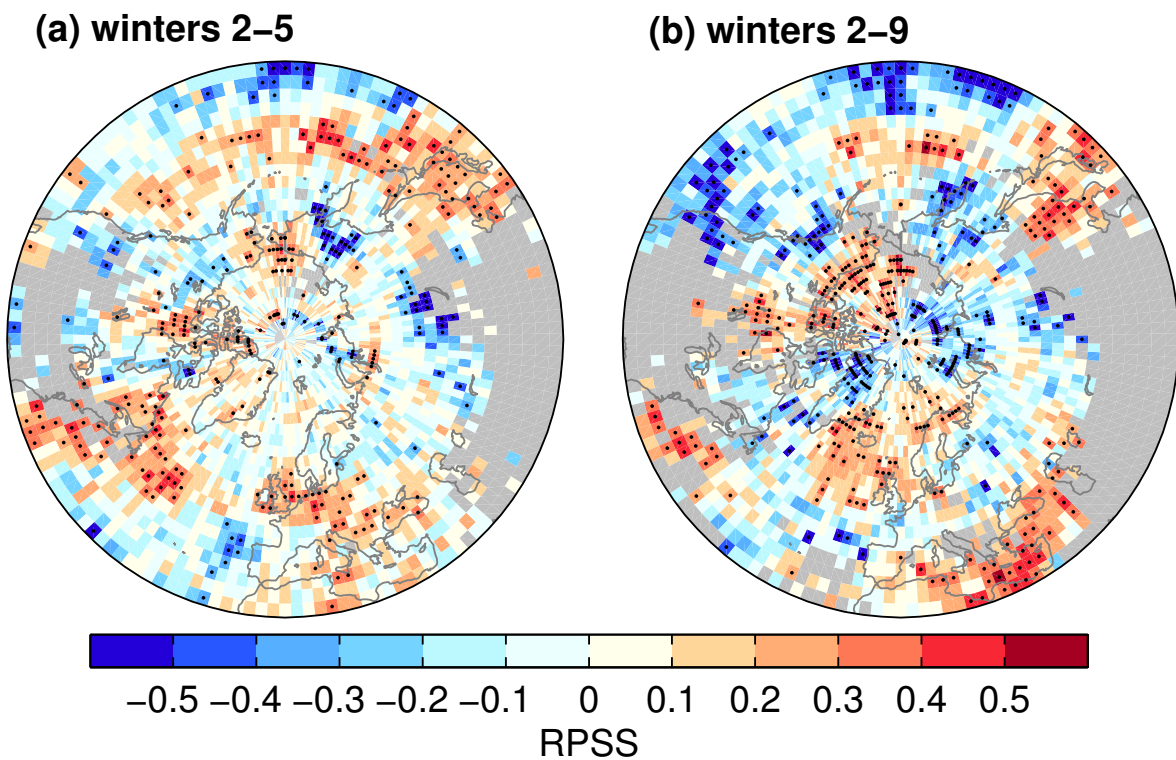
Figure A.5.: **RPSS of *GECCOff* over uninitialized simulations** regarding the average winter storm frequency (number of tracks per ONDJFM in the vicinity of 1000km) for (a) hindcast winters 2–5 and (b) hindcast winters 2–9 based on ERA-reanalyses as observational reference; significant skill scores ($\alpha < 5\%$) as black dots, areas of strong inconsistencies between ERA40 and ERA-Int. are masked out (grey)

## A.2. Peer-reviews with respect to Ch. 3

As stated at the beginning of Ch. 3, this Chapter is almost identical to a manuscript which has been submitted to *Meteorologische Zeitschrift* on 01 Aug. 2014. An editorial decision regarding this study has been made on 10 Dec. 2014, stating that the manuscript may be accepted after review comments have been taken care of without any second review round being recommended to find a final decision. In order to demonstrate this status and for informational purposes, the decision letter from the editor Frank Kaspar is to be found below (in italics). After revising the manuscript according to the Reviewers comments listed here, the paper is now published as Kruschke et al. (2015).

---

*We have reached a decision regarding your submission to Meteorologische Zeitschrift, "Probabilistic evaluation of decadal predictions of Northern Hemisphere winter storms".*

*Our decision is: Revisions required*

*( ) English language editing is required for your paper*

*(X) English language of your paper is fine*

*If English language editing of your paper is required please consult a native language colleague to improve the style of your paper. If after this review round and after final acceptance of your submission an additional English language editing is still required, language editing costs of Euro 10,– per printed page (plus taxes) will be charged.*

*Please carefully revise your manuscript within 6 weeks and follow the reviewers comments. Please comment on all reviewers comments point by point. After revision please upload the revised manuscript version under the Editors Decision tab as Authors version and inform the handling editor about the upload by clicking the notify editor email symbol.*

*Please store all submission related files in one compressed archive and make sure that the maximum file size does not exceed 25 mB per upload. Please wait until the revised uploaded file is visible on the submission page.*

*Frank Kaspar (Deutscher Wetterdienst)*

*Reviewer A:*

*1. General Vote*

*The reviewer recommends to accept the manuscripts after minor changes with mandatory revisions.*

*2. Justification for publication*

*The manuscript presents several novelties with high relevance to the scientific community. The focus discusses core topics of the MIKLIP project. It is perfectly suited to be part of a corresponding special issue.*

*3. Referees synopsis*

*The manuscript analysis the capabilities of decadal forecasting systems in predicting winter storms over the Northern Hemisphere. The study is an extension to a study by the same authors published recently in Tellus A (Kruschke et al, 2014). It extends this earlier study in several dimensions. The manuscript focuses on more severe storms than the previous work. Also, the manuscript analyses additional prediction data sets of the same model making it possible to compare between different initialization strategies, a major issue in the field of decadal forecasts. A further innovation is the novel form of a bias correction being presented.*

*4. Main review points*

*The parameter analyzed in this manuscript is the count of winter storms per grid point. The maximum climatological count of events is 30, most regions have lower counts though, individual years even more so. In some regions, events must even be considered as extreme events. Count data implies special statistical treatment. Least square regression, such as used for the bias correction, is usually not appropriate for count data. How do skill scores behave with count data? Is the bootstrapping for estimating significance appropriate? How do you treat zero values? All these issues are of much stronger relevance than in the predecessor study Kruschke et al (2014), since more rare events are considered. It is strongly advised that the authors analyze this potential implications and make appropriate changes to the methodology.*

*The study would benefit substantially, if the results on winter storms could presented along with results on more general quantities such as bias and skill of average and vari-*

*ance of geopotential fields. Such a comparison will allow a much better interpretation of the results in terms of physical processes and model deficiencies. please add a corresponding figure. Please discuss some studies analyzing the capability of the model, i.e. ECHAM6 in this configuration, in use in representing extratropical dynamics. The language of the requires some improvements. Terms like "basically", "hence", "however" are overused and not necessary in many places. Long sentences should be split. The method to identify winter storms is based on local wind speed values exceeding a percentile (98%) of the local climatology. Is there a risk, that the method underestimates (overestimates) the number of events in regions prone to strong (weak) winds? The atmospheric process under investigation is not linked to relative exceedances, but rather absolute wind speeds. It is not the scope of this study to extend this existing method. Nevertheless the manuscript benefits in discussing this issue based in Section 2.2 and eventually in the conclusions.*

*5. Minor review points*

*Abstract: mention the previous study Kruschke et al (2014)*

*Abstract: make a clear distinction between the cyclones analyzed by Kruschke et al (2014) and the storms discussed in this study.*

*L61-62: this sentence is out of context. please rephrase.*

*L76: please provide, if available, also numbers for other regions prone to winter storms, such as Northern America.*

*L81-83: This sentence is confusing. The reader might interpret transient runs as being initialized at the same time as the decadal predictions, but just with arbitrary data. However, a transient run is just one simulation covering the full period. please rephrase.*

*L127: the introduction is not clear enough about the difference between winter storms and cyclones. As this difference is crucial for the disctinction to Kruschke et al (2014), it is recommended to elaborate on this in the introduction.*

*L157: give additionally approximate values in km*

*L170: "The four other" instead of "The remaining four"*

*L187-L189: language. please split up this sentence.*

*L202: typo: "done by computing"*

*L204: language: "outperform the uninitialized"*

*L215: are daily data used for surface wind? 6hourly?*

*L237: some Figures indicate grey with different definition. Please make sure, that grey shading is clearly defined for all figures.*

*L262: Is it ensured that significance is overestimated due to the low frequency variability (ie. serial correlation) to be expected in multi-annual averages and, even more so, running averages for the transient runs?*

*L271-273: language: please rephrase.*

*L277: the limited sample could also result in non-zero trends.*

*L298: Based on the discussion, please consider to swap 2(a) and 2(b).*

*L305: might the weaker trends be due to the use of the ensemble average?*

*L321contd.: please consider also the limited sample of the observations hampering a trend estimate.*

*L399: "b) corrects for a long-term trend.."*

*L399: "c) eliminates a potential.."*

*L421: replace "state" with "assume"*

*L434-436: move to Section 2. Refer to it as terciles afterwards, eg.*

*L445: there is no "second" following the "first"*

*L472: "applied method here"*

*L499-508: move to Section 4. As the approach is a combination of Kharin et al (2012) and Gangsto et al (2013), it makes sense to mention both of these studies as "inspiration" here and at L559, as you do in Section 4.*

*L543-546: rephrase sentence, break it up into several sentences.*

*L567: the recently published study by Fuckar et al al might be mentioned here: Fučkar, N. S., Volpi, D., Guemas, V., Doblas-Reyes, F. J. (2014). A posteriori adjustment of near-term climate predictions: Accounting for the drift dependence on the initial conditions. Geophysical Research Letters, 41(14), 5200–5207. doi:10.1002/2014GL060815*

*L574: replace "systematic..deformations" with a phrase such as: "systematic biases in*

*the location of the storm track are.. ".*

*Tables 1+2: legend: indicate meaning of parenthesis*

---

*Reviewer B:*

*I would recommend that this paper be accepted after minor reviews. My main issues are that they may be overfitting their data and that initialized and unitialized forecasts are treated differently. Both are serious issues and must be dealt with before I can recommend this paper for publication. When addressed, this paper is timely and presents world-leading research on decadal prediction of climate impacts.*

*Kruschke et al have evaluated the winter storm statistics in several versions of the MiK-ilp decadal prediction systems. They discuss thoroughly their methods and suggest a particular hindcast correction method. None of the decadal prediction systems can be distinguished from the others and they all provide little improvement on uninitialized forecasts.*

*1. The first issue is the third-order polynomial function used for bias corrections (eq 4.1). First of all on line 379 an untrue claim is made that it is more parsimonious than the ICPO procedure. Are the authors saying that if their hindcasts were only five years long then eq 4.1 would be less parsimonious than the ICPO procedure? I don't think you can compare a method that applies to each lead time separately to a method that incorporates all lead times. The ICPO method is later called "non-parametric", which should be a clue that it is far more parsimonious than eq 4.1 (if you insist on a comparison). If the authors wanted a parsimonious method, then that of the cited Kharin et al (2012) only takes three parameters in its exponential form (their eq 5). My main worry however, is for overfitting. Using eight (!) parameters you can get a pretty good fit to your hindcasts that will not necessarily improve your forecasts, due to overfitting. I would recommend that the authors address the issue of overfitting both in section 4 and in section 6.*

*2. The second issue is that the uninitialized forecasts are treated differently to the initialized hindcasts (lines 398-410). This section is unclear, so forgive me if I have misunderstood (and please make this clearer in the paper). It appears that the initialized hindcasts have eq 4.1 applied to them and then the equation between lines 405 + 406, whereas the uninitialized forecasts only have the latter equation applied to them. It is not clear why the hindcasts need a second correction (but this is clearly stated on line 410).*

*The main issue is however, that both hindcast sets (initialized and unitialized) need to be treated the same in order to be comparable.*

*Minor Comments*

*Line 85: "include the unforced component" - it is worth pointing out that initializing a climate model will also include the imprint of previous external forcing (or forced component) that exists in the observations, potentially correcting incorrect responses to this forcing.*

*Line 116: "certain meteorological events" - this is cryptic, what do you mean by this sentence? You should probably include a reference to decadal prediction studies of tropical cyclones as several centres are doing them now, eg. Smith et al (2010)*

*Line 159: Initializing annual 1961-2000 will give you 40 hindcasts, not 41 as stated.*

*Line 236: What is alpha? Is it like a p-value? Make this clearer, as the latter is much more common in climate literature.*

*Line 266: For full-field initialization you have additionally the problem of the model drifting towards its own climatology away from observational space (possibly unrelated to external forcing)*

*Line 327: "we conclude" - you cannot! Multi-decadal variability, such as aerosols, will not average out. You also have volcanoes, which give a different signature again. But both of these are (should be) in the uninitialized models and might make up the bulk of the trend by aliasing. You have not presented enough evidence to conclude the reason for any major difference between the model and re-analysis.*

*Line 379: As mentioned above, this model is not more parsimonious. Also please address the overfitting issue.*

*Line 381: "exemplarily" is not real word.*

*Line 395: Is the model tuned for a 1990s climate? The biases are smallest for the yellow colours, which could be due to the model being developed to mimic the climate of this era.*

*Line 400: For a fair comparison, the transient simulations should be treated exactly the same way as the hindcasts. Why not use eq 4.1? Let the term for drift fall out naturally rather than remove it "by definition".*

*Line 410: Why do the initialized forecasts need corrections applied when eq 4.1 has already been applied?*

*Line 435: I am unfamiliar with what (1/3, 2/3 and 1) means. Do you mean terciles? Does category "1" mean full range of variability covered by climatology and only if the model falls outside of this variability will it have failed?*

*Line 458: What are you measuring over North Africa? Are these really extra-tropical storms in the sense we expect? Perhaps the 98th percentile is not very extreme here.*

*Line 471: "This matches our expectation" - it might be worth reminding the reader here that your method involves tracking wind extremes, the sentence makes a lot more sense when that is clear (rather than just obliquely referred to).*

*Line 496: "hardly significant" - this is not good scientific language. Explain what you mean or quantify what you mean.*

*Lines 499-508: This feels like you are repeating yourself. Refer back to section 4 to make it clearer that it is an application of that work.*

*Line 511: Why did you only test ORAff, why not a grand ensemble of all available hindcasts? That would surely give you the most general result.*

*Line 531: "These are the entrance regions..." - why do you believe that these regions have been improved, what is the mechanism? Aren't these the same regions described on line 496 as "hardly significant"?*

*Line 553: You are not considering events that are more extreme, just more events (you should remove the word "extreme" here).*

*Line 560: "This leads to better skill assessments" - I don't think you can say that. Only if you were to compare your different skill assessments to the actual skill in the forecasts could you conclude this. This might be a good place to discuss overfitting as well.*

*Line 563: The sentence starting here and going all the way to line 566 is a little too long to be easily read.*

*Doug M. Smith, Rosie Eade, Nick J. Dunstone, David Fereday, James M. Murphy, Holger Pohlmann and Adam A. Scaife (2010): Skilful multi-year predictions of Atlantic hurricane frequency. Nature Geoscience 3, 846-849. doi:10.1038/ngeo1004*

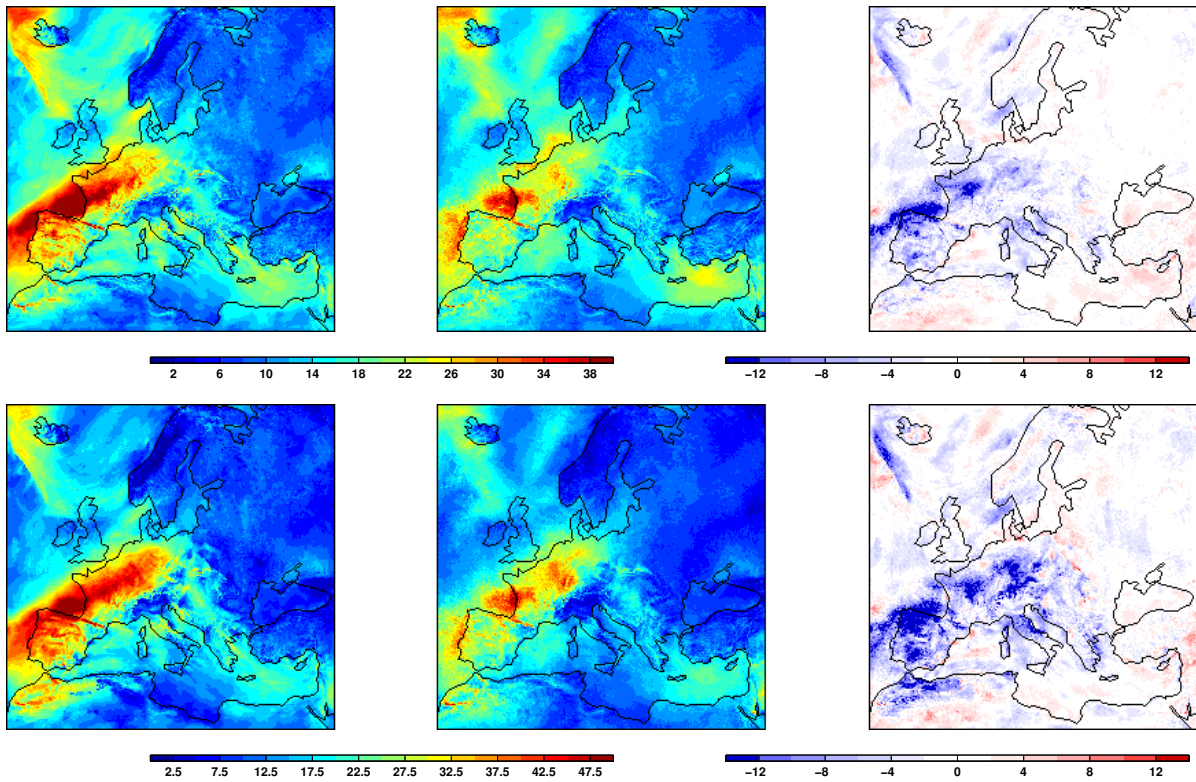## A.3. Statistical downscaling result for winter storm "Xynthia"



Figure A.6.: Example for bad performance of statistical downscaling in case of "unusual" storm events: Footprint of winter storm "Xynthia" (27./28.02.2010) according to dynamical (left) and statistical downscaling (center) as well as the difference between the two (right) with respect to the operational DWD gust estimation (Schulz, 2008, top) (top) and the approach of Brasseur (2001, bottom) (bottom) in m/s

# Danksagung

Das Erstellen dieser Dissertation im Verlauf der letzten vier Jahre wäre ohne die Unterstützung von vielen Seiten nicht möglich gewesen. Zuallererst gilt großer Dank meinem Doktorvater Prof. Dr. Uwe Ulbrich. Als studentische Hilfskraft erfuhr ich in seiner Arbeitsgruppe den Einstieg in die Wissenschaft und durfte seitdem vieles erleben und erlernen. Eine Bachelor-, eine Masterarbeit, mehrere wissenschaftliche Veröffentlichungen und nun diese Dissertation sind die schriftlichen Ergebnisse dieser spannenden, manchmal auch anstrengenden, aber immer lehrreichen Zeit. Mein Dank gilt gleichermaßen PD Dr. Gregor C. Leckebusch und Dr. Henning Rust, die auf ganz unterschiedliche und doch ähnliche Art und Weise Begeisterung aufbringen und wecken, wenn man mit ihnen neue Erkenntnisse, Probleme oder Dinge des Alltags wie Fußballspiele oder Sitzblockaden diskutiert. Danke euch dreien für die tolle Zeit in der AG *Clidia*!

Ich bedanke mich auch bei den verschiedenen Projektpartnern für die gute Zusammenarbeit, vor allem den Kollegen der Munich RE, Dr. Thomas Hofherr, Dr. Eberhard Faust und Peter Miesen, sowie den Kollegen vom MPI-M, Dr. Wolfgang A. Müller und Dr. Holger Pohlmann. Ermöglicht wurden die vorliegenden Arbeiten durch finanzielle Förderung seitens der Munich RE und des Bundesministeriums für Bildung und Forschung über das Forschungsprogramm *MiKlip* (FKZ: 01LP1104A, 01LP1144A, 01LP1160A).

Der gesamten Arbeitgruppe *Clidia* und einer Vielzahl von Institutskollegen danke ich für eure Unterstützung, Anregungen, und Kritik. Hervorheben möchte ich dabei die „verschlissenen" SHKs, Max, Michael, Bruno und Igor, die typische Mensa-Runde sowie meine Freunde vom „MESS(i)E"-Team Robert und Philip, Mareike, Christopher, Martin, Tobias und meinen geschätzten Bürokollegen und Leidensgenossen Jens.

Nicht fehlen darf mein ehemaliger Kollege und (weiterhin) guter Freund Daniel: Mach es fertig bevor es dich fertig macht!

Am Ende bleibt die Familie: Mama, Papa, Oma, Ina und Basti, danke euch... für alles! Silas, du weißt es (noch) nicht, aber du lässt mich die Welt mit anderen Augen sehen, du bist ein Geschenk! Und natürlich, das Beste zum Schluß, Felicitas! Danke für deine Geduld und Toleranz, für deinen Glauben an mich und die Kraft die du mir damit jeden Tag spendest. Du bist mein Sonnenschein!