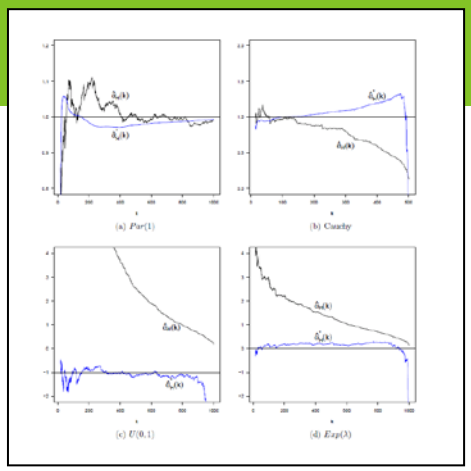
Wegener Zentrum für Klima und Globalen Wandel
Karl-Franzens-Universität Graz

Wissenschaftlicher Bericht Nr. 44-2011



Sturmschaden Österreich: Modellierung des Jahresschadenbedarfes mithilfe einer Nachbarschaftsrelationsmethode

Catharina Altenhuber

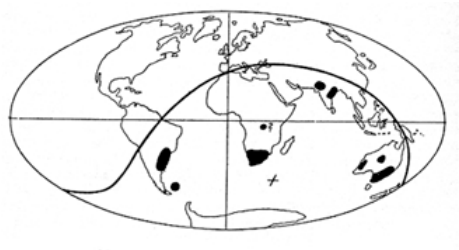
September 2011

Wegener Center
www.wegcenter.at

UNI GRAZ

Das **Wegener Zentrum für Klima und Globalen Wandel** vereint als interdisziplinäres und international orientiertes Forschungszentrum die Kompetenzen der Karl-Franzens-Universität Graz im Forschungsbereich "Klimawandel, Umweltwandel und Globaler Wandel". Forschungsgruppen und ForscherInnen aus Bereichen wie Geo- und Klimaphysik, Meteorologie, Volkswirtschaftslehre, Geographie und Regionalforschung arbeiten in unmittelbarer Campus-Nähe unter einem Dach zusammen. Gleichzeitig werden mit vielen KooperationspartnerInnen am Standort, in Österreich und international enge Verbindungen gepflegt. Das Forschungsinteresse erstreckt sich dabei von der Beobachtung, Analyse, Modellierung und Vorhersage des Klima- und Umweltwandels über die Klimafolgenforschung bis hin zur Analyse der Rolle des Menschen als Mitverursacher, Mitbetroffener und Mitgestalter dieses Wandels. (mehr Informationen unter www.wegcenter.at)

Der vorliegende Bericht wurde im Rahmen einer im März 2011 fertiggestellten Diplomarbeit erarbeitet.



Alfred Wegener (1880-1930), Namensgeber des Wegener Zentrums und Gründungsinhaber des Geophysik-Lehrstuhls der Universität Graz (1924-1930), war bei seinen Arbeiten zur Geophysik, Meteorologie und Klimatologie ein brillianter, interdisziplinär denkender und arbeitender Wissenschaftler, seiner Zeit weit voraus. Die Art seiner bahnbrechenden Forschungen zur Kontinentaldrift ist großes Vorbild — seine Skizze zu Zusammenhängen der Kontinente aus Spuren einer Eiszeit vor etwa 300 Millionen Jahren als Logo-Vorbild ist daher steter Ansporn für ebenso mutige wissenschaftliche Wege:
Wege entstehen, indem wir sie gehen (Leitwort des Wegener Center).

Wegener Center Verlag • Graz, Austria
© 2011 Alle Rechte vorbehalten.
Auszugsweise Verwendung einzelner Bilder, Tabellen oder Textteile bei klarer und korrekter Zitierung dieses Berichts als Quelle für nicht-kommerzielle Zwecke gestattet. Verlagskontakt bei allen weitergehenden Interessen: wegcenter@uni-graz.at.

ISBN 978-3-9503112-1-1

September 2011

Kontakt: Catharina Altenhuber
cathi_a@sbox.tugraz.at

Wegener Zentrum für Klima und Globalen Wandel
Karl-Franzens-Universität Graz
Leechgasse 25
8010 Graz, Austria
www.wegcenter.at

Catharina ALTENHUBER

Sturmschaden Österreich:
Modellierung des Jahresschadenbedarfes mithilfe einer Nachbarschaftsrelationsmethode

MASTERARBEIT

zur Erlangung des akademischen Grades einer/s
Diplom-Ingenieur/in

Masterstudium Finanz- und Versicherungsmathematik

Technische Universität Graz

Betreuer/in:
Univ.-Prof. Dipl.-Ing. Dr.techn. Hansjörg ALBRECHER

Institut für Mathematik B

Graz, im September 2011

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am ....
(Unterschrift)

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitely marked all material which has been quotes either literally or by content from the used sources.

....
date (signature)

Kurzfassung

In Europa wird der Großteil der durch Naturkatastrophen verursachten Versicherungsschäden durch Sturmereignisse verursacht. Zur statistischen Analyse dieser eher seltenen Ereignisse wird häufig Extremwerttheorie verwendet. In einem Teil dieser Arbeit wird deshalb ein Überblick über diese mathematischen Methoden gegeben. Allerdings müssen die Stürme, welche große Gesamtschäden verursachen, nicht unbedingt extreme Windgeschwindigkeiten besitzen, sondern erstrecken sich meist über eine weite Fläche. Aus diesem Grund werden in dieser Arbeit die verbreiteten Ansätze in der Modellierung von Abhängigkeiten, die Copula-Funktionen, beschrieben. Des Weiteren wird versucht das Distanzmatrix-Modell nach Url, welches im Jahr 2009 bereits von der Forschungsgesellschaft Joanneum Research zur Darstellung der Abhängigkeitsstruktur extremer Hochwasser verwendet wurde, auf Sturmereignisse anzupassen bzw. zu erweitern. Dieses Modell bzw. seine Erweiterungen werden auf die Daten von Versicherungsschäden in Österreich der Jahre 1998-2009, welche eine Grenze von 2 Mio. Euro überschreiten, angewendet. Diese Ergebnisse werden anschließend mit jenen der aktuellen Studie zu Sturmschäden der Joanneum Research verglichen.

Abstract

In Europe losses due to storm events make up the major part of insurance losses which are caused by natural catastrophes. Usually extreme value theory is used to describe such rare events. One part of work summarized these mathematical methods. Often the storm events leading to extraordinary insurance losses are not characterized by strong wind speeds but by affecting a large area. Therefore this work describes Copula functions; these are a helpful tool for modeling dependence structures. In addition the „Distanzmatrix-Modell“ (distance matrix model) by Url, which was used by the research cooperation Joanneum Research in the year 2009 to build a model for insurance losses cause by extreme floods, is adapted to storm losses. Then the resulting models are applied to the loss data of the years 1998-2009 which cumulative insurance loss is larger than 2 Mio. Euro. Finally these results are compared to the current study of storm events by Joanneum Research.

Inhaltsverzeichnis

1 Einleitung **1**
1.1 Literaturüberblick 1

2 Sturmschaden in Österreich **6**
2.1 Schadensdaten 6
2.2 Meteorologische Grundlagen 10
2.2.1 Großflächige Stürme 11
2.2.2 Regionale Winde 11
2.2.3 Böigkeit und Düsenwirkung 12
2.2.4 Windgeschwindigkeit 13
2.2.5 Windschäden 13

3 Mathematische Modellierung der Jährlichkeiten **16**
3.1 Jährlichkeiten 16
3.1.1 Wiederkehrperiode einer Folge von Zufallsvariablen 16
3.1.2 Markierte Punktprozesse 18
3.1.3 Jährlichkeiten im markierten Punktprozess 20
3.2 Extremwerttheorie 21
3.2.1 Grenzwertverteilung der Maxima 22
3.2.2 Überschussverteilung 29
3.2.3 Parameter-Schätzung der Generalisierten Pareto-Verteilung 31

4 Modellierung der Abhängigkeiten: Copulas **38**
4.1 Grundlegende Definitionen und Eigenschaften 38
4.2 Archimedische Copula 41
4.3 Tail-Abhängigkeit 43
4.4 Kendalls τ 44
4.5 Der hierarchische Simulationsalgorithmus für Copulas 48

5 Sturmschadenmodelle **51**
5.1 Trend und Saisonalität der Daten 51
5.2 Schadensgradverteilung 53
5.3 Nachbarschaftsrelationen 57
5.4 Distanzmatrixmethode 62
5.4.1 Hochwassermodell nach Url - Überblick 62
5.4.2 Modul 1 63

5.4.3 Modul 2 64
5.4.4 Verteilung der Abweichungen 65
5.4.5 Copula-Modell der Nachbarn 68
5.5 Alternatives Copula-Modell 70

6 Ergebnisse **73**
6.1 Ergebnisse für den Schaden eines Sturmereignisses 73
6.2 Extremwertanalyse der Simulationsergebnisse 78
6.3 Ergebnisse für den Jahresgesamtschaden 81
6.4 Fazit 84

Abbildungsverzeichnis

2.1 Gesamtschäden der Sturmereignisse 7
2.2 Log-Lineares Poisson-Modell der Sturmanzahl 9
2.3 Graphische Überprüfung der Residuen des log-linearen Poisson-Modells . . 10
(a) Residuenplot des Modells 10
(b) Q-Q-Plot der Residuen 10

3.1 Return Period der Zufallsvariablen X_n zum Level x 17
3.2 Markierter Punktprozess ($T \sim \text{Exp}(\lambda)$, $X \sim \mathcal{LN}$) 18
3.3 Der Prozess M_n der Stichprobenmaxima 22
3.4 Generalisierte Extremwertverteilung ($\mu = 0,\ \sigma = 1$) 29
(a) Verteilungsfunktion 29
(b) Dichtefunktion 29
3.5 Generalisierte Pareto Verteilung ($\mu = 0,\ \sigma = 1$) 31
(a) Verteilungsfunktion 31
(b) Dichtefunktion 31
3.6 Q-Q-Plots 32
(a) $U(0,1)$ 32
(b) $Par(1)$ 32
3.7 Log-Log-Plot der Pareto-Verteilung (schwarz) und der Cauchy-Verteilung (grau) 33
3.8 Hillschätzer 35
(a) $Par(1)$ 35
(b) Cauchy 35
(c) $U(0,1)$ 35
(d) $Exp(\lambda)$ 35
3.9 Mean-Excess-Plots 37
(a) $Par(2)$ - 1000 Simulationen 37
(b) $Par(2)$ - je 100 Simulationen 37
(c) $LN(0,1)$ 37

4.1 Scatterplots von gleichverteilten Zufallsvariablen mit verschiedenen Copula-Funktionen 43
(a) Clayton-Copula ($\theta = 4$) 43
(b) Gumbel-Copula ($\theta = 4$) 43
(c) Frank-Copula ($\theta = 8$) 43

5.1 Gesamtschaden mit linearen Trendfunktionen . . . 52
(a) Gesamtschaden . . . 52
(b) normierter Gesamtschaden . . . 52
5.2 Empirische Verteilungsfunktion der Jährlichkeiten (grün: theoretische Verteilungsfunktion) . . . 56
5.3 Histogramm und Log-Log-Plot der Jährlichkeiten (grün: theoretische Verteilungsfunktion) . . . 57
5.4 Kyrill . . . 58
5.5 Paula . . . 58
5.6 Emma . . . 59
5.7 Wolfgang . . . 59
5.8 Einteilung Österreichs in verschiedene Regionen . . . 60
5.9 Grafische Überprüfung der Anpassung der Pareto(2)-Verteing an die Abweichungen der Jährlichkeiten . . . 66
(a) ECDF . . . 66
(b) Histogramm . . . 66
(c) Log-Log-Plot . . . 66
5.10 Boxplot der Kendalls τ für benachbarte PLZ-2-Steller . . . 68
5.11 Boxplot der Kendalls τ der Regionen . . . 71
5.12 Mittelpunkte der Regionen . . . 71

6.1 Graphische Methoden für Modell 1 . . . 79
(a) Hill-Plot . . . 79
(b) Maximum-Likelihood . . . 79
(c) Mean-Excess-Plot . . . 79
(d) Mean-Excess-Plot (linker Rand) . . . 79
6.2 Graphische Methoden für Modell 2 . . . 80
(a) Hill-Plot . . . 80
(b) Maximum-Likelihood . . . 80
(c) Mean-Excess-Plot . . . 80
(d) Mean-Excess-Plot (linker Rand) . . . 80
6.3 Graphische Methoden für Modell 3 . . . 82
(a) Hill-Plot . . . 82
(b) Maximum-Likelihood . . . 82
(c) Mean-Excess-Plot . . . 82
(d) Mean-Excess-Plot (linker Rand) . . . 82

Tabellenverzeichnis

2.1 Sturmschäden 1998-2009 6
2.2 Anzahl der Sturmereignisse/-einzeltage pro Jahr 8
2.3 Sturmereignisse/-einzeltage pro Monat 10
2.4 Ausschnitte aus der Beaufort-Skala nach [22], S. 401-402 14

5.1 p-Werte der Lognormalverteilung 54
5.2 Nullanteile der PLZ-2-Steller 55
5.3 Regionen 61
5.4 Anzahl der PLZ-2-Steller in den Regionen 61
5.5 Nachbarschafts-Distanzen der verschiedenen Regionen 62
5.6 Anzahl der Nachbarn der einzelnen PLZ-2-Steller 67

6.1 Parameter der Lognormalverteilung 74
6.2 Quantile des Gesamtschadens der gegebenen Daten 76
6.3 Quantile des Gesamtschadens eines Sturmereignisses für die verschiedenen Modelle aus Kapitel 5 76
6.4 Quantile der Gesamtschäden mit beschränkten Schadensgraden 76
6.5 Quantile der Gesamtschäden eines Sturmeregnisses mit einer einfachen Abhängigkeitsstruktur zwischen den PLZ-2-Stellern 77
6.6 Quantile der Gesamtschäden eines Sturmeregnisses bei Verwendung einer Clayton-Copula 78
6.7 Quantile des Jahresgesamtschadens für die verschiedenen Modelle aus Kapitel 5 83
6.8 Quantile des Jahresgesamtschadens bei einfachen Abhängigkeitsstrukturen zwischen den PLZ-2-Stellern 84

1 Einleitung

Extreme Sturmereignisse bzw. der dadurch verursachte Schaden stellen große Herausforderungen an die Versicherungswirtschaft. So verursachte allein der großflächige Sturm Kyrill im Jänner 2007 insgesamt rund 6097 Mio. USD Versicherungsschaden über ganz Europa. Damit machten in diesem Jahr Sturmereignisse mit über 60% den Großteil des durch Naturkatastrophen hervorgerufenen Schadens aus (vgl. [52]).
Die Ursachen des hohen Schadens, der durch Kyrill aufgetreten ist, liegen vor allem in dessen großer geografischer Ausbreitung. Deshalb ist es notwendig, die örtliche Abhängigkeitsstruktur von Schadensdaten in die Modellierung von Sturmschäden einfließen zu lassen.

1.1 Literaturüberblick

In den letzten Jahren wurden aus diversen Gründen eine Vielzahl an wissenschaftlichen Publikationen über extreme Sturmereignisse veröffentlicht. Zum einen hat die scheinbare Zunahme von heftigen Naturkatastrophen in den letzten Dekaden zur Frage geführt, ob und wie sich der Klimawandel und insbesondere die Erhöhung der Treibhausgaskonzentration in der Erdatmosphäre auf extreme Sturmereignisse auswirkt. Mit diesen Überlegungen geht auch die Nachfrage nach erneuerbarer Energie, wie Windenergie, einher. Auf der anderen Seite hat auch die EU-Verordnung Solvency II strengere Eigenkapitalanforderungen an Versicherungsunternehmen gestellt. Somit müssen genaue Modelle zur Beschreibung des erwarteten Schadens bzw. des PMLs („probable maximum loss“ – wahrscheinlicher maximaler Schaden) getroffen werden.

Um die Auswirkungen des Klimawandels auf Sturmereignisse in Europa zu bestimmen, werden meist Simulationen allgemeiner Zirkulations-Modelle, d.h. Modelle für die Entstehung und den Verlauf von Hoch- und Tiefdruckgebieten, für verschiedene Szenarien der Zunnahme von Treibhausgasen betrachtet.
Aufgrund derartiger Analysemethoden rechnen u.a. Knippertz et. al. (siehe [27]) und Leckebusch et. al. (siehe [30]) mit einer Verlagerung von Tiefdruckgebieten während der Wintermonate über Europa weiter in Richtung Norden, d.h. Nordrussland, Hudson Bay, Grönland. Diese Verschiebungen sind aber teilweise modellabhängig, sodass auch Westeuropa und das westliche Zentraleuropa davon betroffen sein können. Eine Folge ist das Auftreten von extremen Windgeschwindigkeiten (vgl. auch Abschnitt 2.2). In diesen Arbeiten werden jene Windgeschwindigkeiten als extrem bezeichnet, die das 99%-Quantil der jeweiligs betrach-

teten Rasterzelle überschreiten. Dabei ist vor allem mit einer Häufung im Norden Europas zu rechnen.
Donat et. al. betrachten in [12] 9 verschiedene Klima-Simulationen um die Anzahl und Dauer von Winterstürmen zu analysieren. Ein Sturmtag liegt in dieser Arbeit dann vor, wenn entweder in mindestens 25% des betrachteten Gebietes das 98%-Quantil der täglichen maximalen Windgeschwindigkeit überschritten wird oder wenn der Gale-Index G einen Wert von mindestens 35 beträgt[1]. Dabei ist

$$G = \sqrt{F^2 + Z^2/2},$$

wobei F die Stärke des sogenannten geostrophischen Windes und Z die Wirbelstärke misst. Der geostrophische Wind ist jene Luftbewegung, die nur von der Corrioliskraft und den Druckunterschieden zwischen Hoch- und Tiefdruckgebieten abhängt. Für die auf diese Weise definierten Sturmtage wird eine Zunahme aufgrund der westlichen Strömungen in den Wintermonaten erwartet.

Allerdings hat die Topografie des Geländes einen großen Einfluss auf die tatsächlichen Windgeschwindigkeiten. Deshalb ist die Erstellung von Windkarten für einzelne Länder äußerst komplex. Für die Schweiz wird eine solche Einteilung von Etienne, Lehmann und Goyette durchgeführt (siehe [18]). Anhand von 3-stündlichen Messungen der Windgeschwindigkeit in 70 Stationen seit 1981 kann die Schweiz in 7 Kategorien unterteilt werden, welche u.a. von der Geländehöhe wie auch der umliegenden Topogafie abhängen. Dabei wurden auch Parameter wie Krümmungen des Geländes und Ausrichtungen von Gebirgszügen betrachtet.
In Österreich wird im Moment von der ZAMG[2] ein ähnlicher Bericht namens BEAUFORT[3] entwickelt, der sowohl die hochaufgelöste Topografie als auch Rauigkeitsfaktoren berücksichtigt (siehe [56]).
Weiters wird zur Zeit ein detaillierter Windatlas für Österreich entwickelt. An diesem Projekt sind der Verein Energiewerkstatt, die Forschungseinheit Studio iSpace, die Firma Meteotest und das Wegener Zentrum für Klima und Globalen Wandel beteiligt (siehe [2]). Dieser Atlas soll mögliche Standorte für Windparks zur Erzeugung von Windenergie in Österreich aufzeigen.
Die Problematik der richtigen Positionen von Windkraftanlagen ist nicht unabhängig von der Sturmschaden-Modellierung, da Windkraftwerke durch ihre Höhe und die Länge der Rotorblätter starken, durch Wind verursachten Kräften ausgesetzt sind. Während auf der einen Seite ein konstanter, böenloser Wind zur Stromerzeugung benötigt wird, darf auf der anderen Seite der Standort keinen extremen Windgeschwindigkeiten ausgeliefert sein. Die Methodik zur Modellierung dieser Windgeschwindigkeiten ist grundsätzlich dieselbe wie jene im Zusammenhang mit Sturmschäden, für eine Zusammenfassung der aktuell verwendeten Methoden siehe [41].

[1] Der englische Ausdruck „gale“ wird häufig als „Sturm“ oder „Starkwind“ ins Deutsche übersetzt. In der englischen Version der Beaufort-Skala wird der Grad 8 als „gale“ bezeichnet (siehe 2.4).
[2] Zentralanstalt für Meteorologie und Geodynamik
[3] Bessere Wind Energie Ausnutzung - Verortung der Windgefährdung

Ein weit verbreiteter Ansatz zur Beschreibung der Verteilungsfunktion extremer Windgeschwindigkeiten ist, dass ihr Quadrat einer Gumbel-Verteilung folgt. Simiu et. al. schlagen in [50] eine Weibull-Verteilung vor (siehe Kapitel 3.2.1). Im Allgemeinen sollte die Frage nach der passenden Verteilung immer in Hinblick auf die gegebenen Daten geklärt werden.

Zur Modellierung der Versicherungsschäden können verschiedene Ansätze gewählt werden. Hier kann entweder ein Modell zur Beschreibung des entstandenen Schadens an einzelnen Gebäuden, welche anschließend aggregiert werden, verwendet werden, oder es wird direkt eine Verteilung an den Gesamtschaden angepasst.
In [23] wird versucht, den Schaden an Einzelgebäuden mithilfe diverser Verteilungsfunktionen darzustellen. So wird angenommen, dass der Anteil der Reparaturkosten am Gebäudewert eine steigende Funktion g der Windgeschwindigkeit ist. Dabei tritt ein Schaden erst ab einer gewissen Geschwindigkeit v_{crit} auf und ab einer Geschwindigkeit v_{tot} entsteht ein Totalschaden. Dabei sind diese beiden Grenzgeschwindigkeiten sowie die Funktion unbekannt und hängen vom Zustand des jeweils betrachteten Gebäudes ab. Durch die Wahl von passenden Verteilungsfunktionen für v_{crit} und v_{tot} bzw. der Wahl von g als quadratische oder kubische Funktion einer normierten Windgeschwindigkeiten kann in dieser Arbeit der entstandene Schaden aus vier vorliegenden Sturmereignissen durch dieses Modell angenähert werden.
Weiters wird auch in [14] der durch Wind verursachte Schaden an Einzelgebäuden betrachtet. Diese Arbeit widmet sich der Einstufung von Tornadoereignissen[4], welche im allgemeinen nicht so große Gesamtschäden wie die in dieser Arbeit betrachteten hervorrufen. Außerdem treten in Deutschland nur 2-5 Tornados pro Jahr auf, in Österreich wurden maximal 6 pro Jahr gemeldet. Diese Zahlen sind aber stark abhängig von der Bevölkerungszahl und werden durch bessere Aufklärung im Laufe der Zeit noch steigen. Zur Meldung dieser Wetterereignisse wurde das Netzwerk Tordach gegründet (siehe [28]). Dennoch kann die verwendete „T-Skala“ mit den darin angegebenen Schadensätzen eine mögliche Höchstgrenze für Versicherungsschäden bieten. Die verschiedenen Klassen dieser Skala beschreiben Tornados ab einer Windgeschwindigkeit von 76km/h, welche einem Sturm nach der üblicherweise verwendeten Beaufort-Skala (siehe 2.2) entsprechen. Für eine Windgeschwindigkeit bis zu ca. 300km/h wird hier für Massivbauten ein Schadensanteil von 30% des Neuwertes angegeben, an Leichtbauten wie z.B. Lagerhallen und Scheunen können Schadensgrade von bis zu 90% entstehen. Der Schadensanteil von Massivbauten liegt aber bis zu einer Windgeschwindigkeit von 200km/h unter 1%.

Ein Modell zur Beschreibung des Jahresgesamtschadens findet sich unter anderem in [26]. Dafür werden die täglichen Maximalgeschwindigkeiten für Windböen aus 90 Messstationen in Deutschland betrachtet. Die Analyse dieser Daten und jener der jährlichen Gesamtschäden zeigt, dass diese Windgeschwindigkeiten normalisiert werden müssen, da

[4]Das Naturphänomen „Tornado“ wird im Deutschen häufig auch Windhose oder Trombe genannt. Mit diesen Begriffen werden vertikale Luftwirbel bezeichnet, welche von einer Gewitterwolke bis zum Boden reichen. Obwohl der Begriff Tornado eher mit den USA in Verbindung gebracht wird, können solche Wirbel auch in Europa auftreten.

ansonsten diesselbe Windgeschwindigkeit unterschiedliche Schäden verursachen kann. Danach kann der Verlust der Versicherungen als kubische Funktion der Windgeschwindigkeit dargestellt werden.
Eine weitere Arbeit, welche den entstandenen Schaden als Funktion der Windgeschwindigkeit ausdrückt, wurden von Dorland et. al. in [13] durchgeführt. Hier wird der logarithmierte Schaden in einer Region i bzgl. des Sturmereignisses t als gewichtete Summe der Logarithmen der Anzahl der versicherten Objekte, der Fläche der Region und der maximalen Böengeschwindigkeit dargestellt. In dieser Studie wird dieses Modell zwar nur für 5 Stürme betrachtet, durch die hochaufgelösten Versicherungs- und Winddaten, allerdings ohne Berücksichtigung topografischer Effekte, kann jedoch eine gute Anpassung erzielt werden. Interessant sind vor allem die Zukunftsprognosen. Es stellt sich heraus, dass eine Änderung im Versicherungs-Portfolio relativ wenig Einfluss auf den Gesamtschaden hat, während eine Zunahme der Windstärke von 2% einen Anstieg des Versicherungsverlustes von 50% bedeutet.

Auch die Anwendung von Extremwertverteilungen bietet eine mögliche Beschreibung von Schadensdaten. In [32] wird vorgeschlagen, multivariate Extremwertverteilungen zur Darstellung von Schäden aus jenen Versicherungssparten zu verwenden, die von demselben Naturereignis betroffen sind.
Eine ausführliche Analyse der Versicherungsdaten der schwedischen Versicherungsgruppe Länsförsäkringar wurde von Rootzén und Tajvidi in [45] und [46] bzw. von Rootzén und Brodin in [3] durchgeführt. In diesen Arbeiten wird neben loglinearer Regression der Schäden bzgl. der Windgeschwindigkeit auch die Generalisierte Pareto-Verteilung für die Beschreibung des Gesamtschadens der einzelnen Sturmereignisse verwendet. Diese Verteilungsfunktion wird dazu verwendet, Überschüsse einer gewissen Schranke zu beschreiben (siehe 3.2.2) angepasst. In [46] wird der Skalenparameter als Funktion des Winddrucks dargestellt. Dennoch können durch die meteorologischen Daten die enstandenen Schäden nicht vollständig beschrieben werden. Nach einer Bereinungung des Trends bzgl. Inflation und Portfolio-Änderungen wird in [3] auch versucht den Skalen- und Form-Parameter der Generalisierten Pareto-Verteilung als Funktion der Zeit zu schätzen. Dabei kann allerdings kein signifikanter Einfluss eines weiteren zeitlichen Trends gefunden werden. Die Ausnahme bildet hier die zusätzliche Betrachtung des außergewöhnlich starken Orkans Gudrun. Dadurch kam es zu einer Steigerung des Formparameters von $0,99$ in den Jahren von 1982-2004 auf 1.21.

In dieser Arbeit wird eine solche Analyse der Daten durch das von Url entwickelte Distanzmatrixmodell (siehe [53]) durchgeführt. Diese Methode wurde schon im Jahr 2009 von der Forschungsgesellschaft Joanneum Research verwendet, um im Zuge ihrer Studien zum Klimawandel [39] die räumliche Abhängigkeitsstruktur von extremen Hochwassern zu modellieren. Die mit dieser Methode erzielten Ergebnisse wurden mit den anderen Ergebnissen der Studie sowie mit einem alternativen Copula-Modell verglichen. Dabei stellte sich heraus, dass die Simulationsergebnisse aus dem Distanzmatrixmodell ähnliche Größenordnungen für Hochwasserschäden hervorbrachten wie die anderen Modelle.

Die grundlegende Annahme des Distanzmatrixmodells ist, dass sich die Hochwasser-Jährlichkeit in benachbarten Gebieten nur um eine zufällige multiplikative Konstante unterscheidet. Durch eine Monte-Carlo-Simulation kann anhand des Modells für jedes Gebiet in Österreich die entsprechende Jährlichkeit generiert werden. In den betroffenen Gebieten kann anschließend die Anzahl der überfluteten Gebäude sowie das Ausmaß des Schadens simuliert werden.
Für Sturmschäden können allerdings die betroffenen Gebäude in Abhängigkeit der Jährlichkeit der Windgeschwindigkeit nicht ermittelt werden. Deshalb wird das Distanzmatrixmodell direkt auf die Jährlichkeit der Sturmschäden angewendet. Da kein Flussnetz vorliegt, welches die Definition einer Nachbarschaftsrelation induziert, wird aufgrund von meteorologischen Überlegungen und der topografischen Struktur der einzelnen Gebiete eine Einteilung Österreichs in verschiedene Nachbarschaften vorgenommen. Weiters werden neben den multiplikativen Abweichungen auch Copula-Funktionen zur Beschreibung der Abhängigkeitsstruktur zwischen zwei benachbarten Gebieten verwendet.

In Kapitel 2 wird ein Überblick über die verwendeten Daten sowie über die meteorologischen Grundlagen zur Wind-Entwicklung gegeben. Kapitel 3 widmet sich der Modellierung von Jährlichkeiten durch die Betrachtung extremer (bzw. seltener) Ereignisse und in Kapitel 4 wird das Konzept der Copula-Funktionen beschrieben. Die letzten beiden Kapitel erklären, wie die Modelle aus der Hochwassermodellierung auf Sturmschäden angepasst wurden und welche Ergebnisse die veränderten Modelle liefern.

2 Sturmschaden in Österreich

Bevor in den nächsten Kapiteln auf die mathematische Modellierung von extremen Wettereignissen bzw. Abhängigkeitsstrukturen eingegangen wird, sollen hier die zugrundeliegenden Daten der Analyse vorgestellt werden. Außerdem wird in Abschnitt 2.2 ein kurzer Überblick über die meteorologischen Grundlagen zur Wind-Entstehung gegeben.

2.1 Schadensdaten

In den letzten Jahren wurde den Sturmschäden in Österreich durch die Öffentlichkeit, aber auch durch Versicherungsunternehmen immer mehr Aufmerksamkeit geschenkt. Allein die größten Sturmereignisse der letzten Jahre, wie Kyrill (18. bis 20. Jänner 2007), Paula (26. bis 28. Jänner 2008) oder Emma (1. bis 3. März 2008), verursachten Versicherungsschäden an Wohngebäuden in der Höhe von ca. 160 Mio. Euro, 64 Mio. Euro bzw. 110 Mio. Euro.

Nr.	Datum	Bezeichnung	Schaden	Nr.	Datum	Bezeichnung	Schaden
1	07.07.1998		2,85	20	18.11.2004		13,93
2	31.07.1998		3,35	21	30.05.2005		2,59
3	04.02.1999	Lara	4,23	22	01.01.2006		7,73
4	01.06.1999		12,83	23	27.06.2006		10,38
5	26.12.1999	Lothar	21,41	24	29.08.2006		0,11
6	17.01.2000		8,41	25	11.01.2007	Franz	4,37
7	04.07.2000		25,57	26	18.01.2007	Kyrill	160,30
8	07.07.2000		3,18	27	28.01.2007	Olli	9,10
9	07.07.2001		5,24	28	20.06.2007		19,59
10	23.06.2002		5,22	29	26.01.2008	Paula	63,31
11	10.07.2002		2,15	30	01.03.2008	Emma	110,65
12	11.08.2002		5,26	31	23.06.2008		13,20
13	27.10.2002	Jeanette	13,66	32	12.07.2008		3,71
14	16.11.2002		9,25	33	14.08.2008		4,45
15	01.01.2003	Calvann	3,83	34	22.08.2008		8,62
16	13.05.2003		2,42	35	26.05.2009	Felix	4,44
17	28.08.2003		3,89	36	16.06.2009		2,84
18	12.01.2004	Gerda/Hanne	3,85	37	06.07.2009		4,67
19	12.08.2004		3,56	38	23.07.2009	Wolfgang	115,30

Tabelle 2.1: Sturmschäden 1998-2009

Zur Modellierung solch extremer Sturmschäden werden Schadensdaten jener Sturmereignisse zwischen Juli 1998 bis Juli 2009 betrachtet, deren Gesamtschaden in ganz Österreich die Grenze von 3 Mio. Euro überschreitet (siehe Tabelle 2.1). Dieser Datensatz beinhaltet den Schaden für die einzelnen Sturmtage in jedem PLZ-2-Steller sowie die dessen jährliche Gesamtversicherungssumme. Ein PLZ-2-Steller ist die Zusammenfassung jener Gebiete, deren ersten beiden Zahlen der Postleitzahl gleich sind. Das Gebiet, dessen Postleitzahlen mit der Nummer 13 beginnen (PLZ-2-Steller 4), weist dabei ein außergewöhnliches Verhalten auf. Hier liegen nur Schäden an 6 Sturmtagen vor, während durchschnittlich an 60 der 90 Tage ein positiver Schaden gegeben ist. Außerdem beträgt der höchste eingetretene Schaden ca. 14‰ der versicherten Summe in diesem PLZ-2-Steller. Dieser Wert ist 10-mal so hoch wie der im Durchschnitt eingetretene Anteil in allen PLZ-2-Stellern. Bei diesem PLZ-2-Steller handelt es sich um das Gebiet um den Flughafen Wien-Schwechat. Die Anzahl der Gebäude in diesem PLZ-2-Steller ist relativ gering und führt somit zu hohen Promillewerten. Deshalb wurde dieser PLZ-2-Steller aus den weiteren Berechnungen ausgenommen.

Die Schadensdaten wurden vom VVÖ bereits bzgl. dem Verbraucherpreisindex inflations-

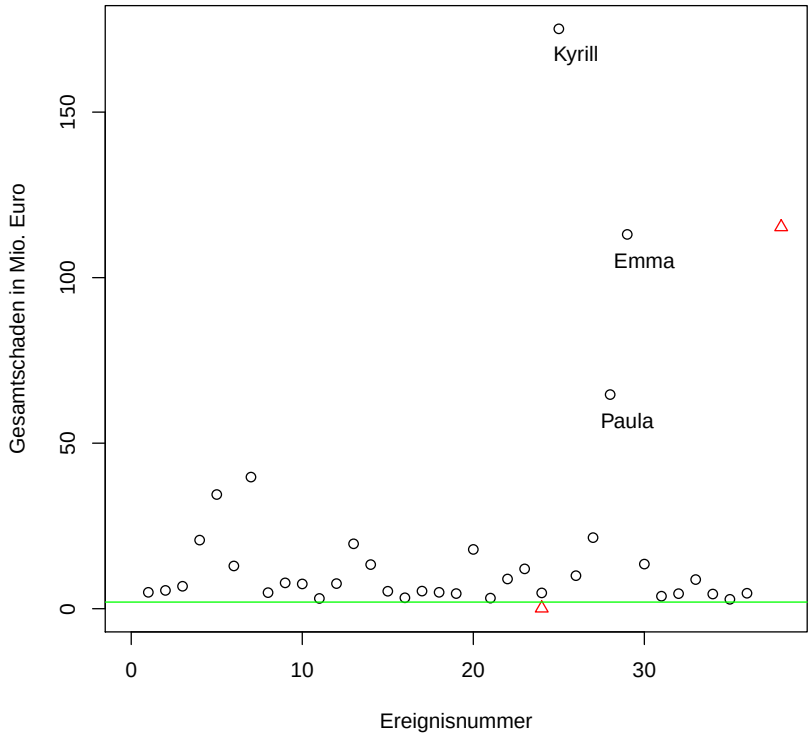


Abbildung 2.1: Gesamtschäden der Sturmereignisse

bereinigt. Die Grenze von 3 Mio. Euro bezieht sich allerdings nur auf den Gesamtschaden

an Industrie- und Wohngebäuden. In dieser Arbeit soll aber ein Modell für den Schaden an Wohngebäuden beschrieben werden, deshalb kann der Gesamtschaden unter dieser Schranke liegen. Die Schadensdaten sind für jeden der 77 PLZ-2-Steller Österreichs an jedem Tag des Sturmereignisses gegeben. Somit liegen für jeden PLZ-2-Steller die Versicherungsschäden an 90 Einzeltagen bzw. bei 38 Sturmereignissen vor. Zur Ermittlung der Sturmereignisse werden aufeinanderfolgende Tage zu unabhängigen Ereignissen zusammengefasst, indem die Schäden an den einzelnen Tagen aufsummiert wurden. Diese Sturmereignisse sowie ihre Gesamtschäden sind in Tabelle 2.1 aufgelistet. Wie man erkennen kann, bewegt sich der Schaden von Sturm Nr. 24 weit unter der Grenze von 3 Mio. Euro. Bei diesem Schaden liegen vor allem Versicherungsschäden an Industriegebäuden vor, deshalb wurde er aus der weiteren Analyse genommen. Auch jenes Sturmereignis mit dem zweitgrößten Gesamtschaden, „Wolfgang“, passt wegen seiner niedrigen Windgeschwindigkeiten nicht zu den übrigen Sturmdaten. Hier sind die Schadensdaten vermutlich durch Überflutungsschäden verunreinigt. Somit kann auch dieser Sturm nicht in die Berechnungen miteinbezogen werden.
Die Höhen der Gesamtschäden der einzelnen Sturmereignisse sind auch grafisch in Abbildung 2.1 dargestellt. Die ausgenommenen Stürme sind als Dreiecke eingezeichnet.
Die Dauer der einzelnen Stürme beträgt zwischen einem und 7 Tagen (7-Tage-Sturm im Jänner 2006), im Mittel dauert ein Sturm ca. 2,4 Tage.

In Tabelle 2.2 ist die Anzahl der Sturmereignisse bzw. der Sturmtage pro Jahr dargestellt. Anscheinend hat die Anzahl der Sturmeinzeltage pro Jahr zugenommen.

1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
2/3	3/7	3/7	1/1	5/10	3/7	3/9	1/1	3/11	4/11	6/15	4/5

Tabelle 2.2: Anzahl der Sturmereignisse/-einzeltage pro Jahr

Um zu überprüfen, ob auch die Häufigkeit extremer Sturmereignisse in den letzten Jahren zugenommen hat, wird die Anzahl n_i der Sturmereignisse als log-lineares Poisson-Modell dargestellt. Dabei ist n_i Poisson-verteilt mit Parameter μ_i, dessen Logarithmus einem linearen Regressionsmodell der vergangenen Jahre entspricht:

$$n_i \sim Poi(\mu_i) \qquad \text{mit} \quad \log(\mu_i) = \alpha + \beta i,\ \alpha,\ \beta \in \mathbb{R}.$$

Dem Jahr der ersten Beobachtung, 1998, wird hier der Wert $i = 0$ zugewiesen. In Abbildung 2.2 ist die Kurve dieses Regressionsmodells eingezeichnet. Die Schätzung der Parameter liefert eine Konstanten α von 0,8421 und eine Steigung β von 0.0445. D.h., dass die erwartete Anzahl im Jahr 1998 laut diesem Modell

$$\mathbb{E}[n_0] = \exp\{\alpha\} = 2,321208$$

beträgt und während der Dauer der Beobachtungen auf beinahe 4 extreme Sturmereignisse pro Jahr angestiegen ist:

$$\mathbb{E}[n_1 1] = \exp\{\alpha + 11 \cdot \beta\} = 3,787076.$$

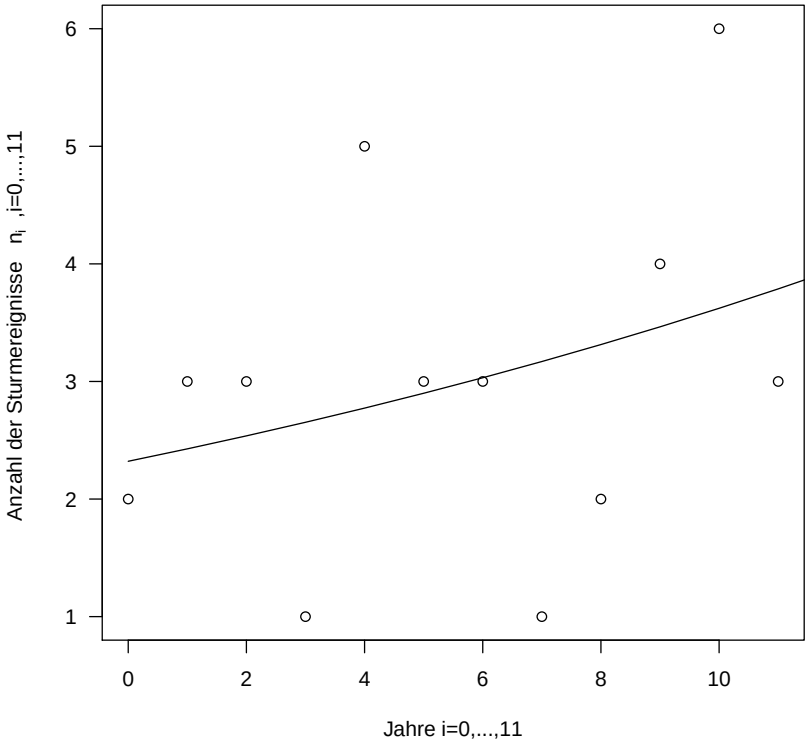
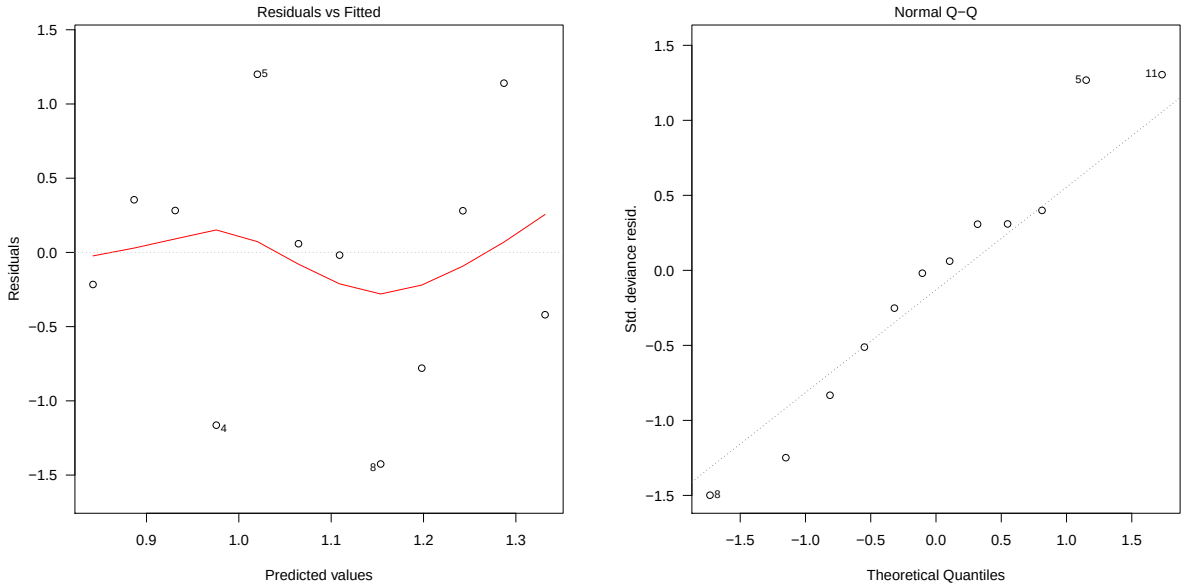


Abbildung 2.2: Log-Lineares Poisson-Modell der Sturmanzahl

Würde sich dieser Trend durchsetzen, rechnet das Modell in den Jahren 2010 und 2011 mit 3,96 bzw. 4,14 extremen Sturmereignissen.
Die Standardabweichung dieses Modells liegt mit 8,0893 unter der Anzahl der Freiheitsgrade, also würde auch ein konstanter Mittelwert eine gute Anpassung liefern. Der Wert der Pearson-Statistik, wenn die vergangenen Jahre als erklärende Variable hinzugenommen werden, ist mit 7,2458 nicht weit von der Standardabweichung entfernt. Somit ist die Zeit als beschreibende Variable nicht unbedingt nötig. In Abbildung 2.3(a) sind die Pearson-Residuen des Modells dargestellt, welche einer Standardnormalverteilung folgen sollten. Starke Ausreißer gibt es vor allem in den Jahren 5 (2002) und 8 (2005) mit 5 bzw. einem Sturmereignis. Der Q-Q-Plot der Residuen scheint das Modell nicht zu verwerfen, aber auch hier führen die beiden Jahre zu Abweichungen von der Geraden.

Ein Zusammenhang zwischen den Schäden und der Dauer eines Sturmes bzw. die Aufteilung der Schäden auf die Einzeltage eines Ereignisses konnte anhand dieser Daten nicht gefunden werden. Vermutlich ist dies auf eine Messungenauigkeit (z.B. wenn Schäden für den falschen Sturmtag gemeldet wurden) zurückzuführen. Möglicherweise könnte eine solche Analyse durchgeführt werden, wenn auch Windgeschwindigkeiten in das Modell einfließen. Dazu müssten aber die Versicherungsdaten in höherer räumlicher Auflösung gegeben sein.

Da insgesamt nur 36 unabhängige Ereignisse in den Beobachtungszeitraum fallen, wurde



(a) Residuenplot des Modells

(b) Q-Q-Plot der Residuen

Abbildung 2.3: Graphische Überprüfung der Residuen des log-linearen Poisson-Modells

davon abgesehen, diese in Winter- und Sommerstürme zu unterteilen (vgl. Tabelle 2.3). Diese Unterteilung wäre prinzipiell sinnvoll, weil die Schäden der Sommerstürme oft durch Hagelschäden verunreinigt sein können. Dies könnte auch auf den Sturm vom 23. Juli 2009 zutreffen, der aus den Daten ausgenommen wurde.

1	2	3	4	5	6	7	8	9	10	11	12
8/30	1/2	1/3	0/0	3/4	6/19	8/9	5/11	0/0	1/2	2/5	1/1

Tabelle 2.3: Sturmereignisse/-einzeltage pro Monat

Da die täglichen Schadenshöhen in den einzelnen PLZ-2-Stellern möglicherweise durch zu späte oder zeitlich ungenaue Meldungen verunreinigt sind und um zeitliche Unabhängigkeit der Schadensdaten zu erzielen, wurden die einzelnen Sturmtage zu Sturmereignissen zusammengefasst. Dabei gehören zwei Tage zum selben Ereignis, wenn für beide Tage Schadensdaten existieren und die Tage aufeinander folgen bzw. wenn für alle dazwischenliegenden Tage ebenfalls Schadensdaten gegeben sind.

2.2 Meteorologische Grundlagen

Um die Nachbarschaften aus dem Distanzmatrixmodell an Sturmschäden anzupassen, muss bekannt sein, wie Stürme und auch Sturmschäden entstehen. Dieser Abschnitt soll einen

kurzen Überblick über die Eigenschaften von Wind geben und auch die Frage klären, wodurch sich Wind und Sturm unterscheiden. Diese Einteilung ist letztendlich auch für Versicherungen interessant. Für eine genauere Beschreibung der meteorologischen Grundlagen siehe [22] und [49].

2.2.1 Großflächige Stürme

Winde entstehen bekanntlich durch hohe Druckunterschiede in der Atmosphäre. Die Luft aus den Gebieten mit hohem Luftdruck (Hochdruckgebiet) fließt in Gebiete mit niedrigem Luftdruck (Tiefdruckgebiet). Dabei wird zwischen thermischen und dynamischen Tief- bzw. Hochdruckgebieten unterschieden.

1. **Dynamische Tief- und Hochdruckgebiete**
   Die Sonnenstrahlung trifft die Erde durch ihre Kugelform in Äquatornähe in einem steileren Winkel als an den Polen. Somit wirkt sie am Äquator gebündelt, während sie an den Polen auf eine größere Fläche verteilt ist. Dadurch entsteht eine unterschiedliche Erwärmung der Lufttemperatur, woduch sich warme („tropische“) und kalte („polare“) Luftmassen bilden. Diese treffen in der sogenannten „Frontalzone“aufeinander. In dieser Zone sollen die Temperaturunterschiede ausgeglichen werden, in Folge entstehen starke Winde in der oberen Troposphäre, die sogenannten „Jetstreams“oder „Strahlströme“.
   In Bodennähe werden Winde durch eine andere Ursache erzeugt. Die Frontalzone ist ein Band, welches parallel zum Äquator verläuft. Wenn die Temperaturunterschiede allerdings sehr groß sind, kann der Strahlstrom eine wellenartige Bewegung annehmen. Dadurch werden die transportierten Luftmassen abgebremst bzw. beschleunigt und es entsteht ein Luftüberschuss bzw. -defizit entlang der Frontalzone. Um diese Luftdruckunterschiede auszugleichen, dehnt sich ein Hochdruckgebiet nach oben und unten aus, während durch ein Tiefdruckgebiet die bodennahe Luft in die Höhe gesaugt wird.

2. **Thermische Tief- und Hochdruckgebiete**
   Durch die erhitzte Erdoberfläche wird allerdings auch Energie an die bodennahen Luftschichten abgegeben. Durch den Aufstieg der warmen Luft kommt es zu einem Unterdruck in Bodennähe. Dieser Unterdruck wird durch Strömungen ausgeglichen, es entsteht Wind.

In Europa verursachen vor allem das Azorenhoch über dem Nordatlantik bzw. das Islandtief große Druckunterschiede, welche in Westwinden resultieren können.

2.2.2 Regionale Winde

Bei Windsystemen, welche regional und meist auch zeitlich stark begrenzt sind, spricht man von regionalen Windsystemen. Ein bekanntes Beispiel ist das Land- und Seewindsystem, das vor allem in Küstennähe auftritt. Im Laufe eines Tages erwärmt sich die Luft

über dem Festland mehr als jene über See, während nachts dieser Effekt umgekehrt wird. Dadurch entsteht eine Zirkulation, die tagsüber Luft in Bodennähe in Richtung Festland bringt und in größerer Höhe die am Festland erwärmte und somit aufgestiegene Luft in Richtung Meer transportiert. Nachts verläuft die Zirkulation natürlich in entgegengesetzter Richtung.
Ein regionales Windsystem, welches mehr Einfluss auf Österreich hat, ist das Hangwindsystem. Hier erwärmt sich bei Schönwetter ein südlich ausgerichteter Gebirgshang, sodass die bodennahe Luft entlang des Hanges in die Höhe steigt. Die Luft in einiger Entfernung wird jedoch nicht so stark erwärmt. Somit entsteht in Hangnähe ein thermisches Hoch. Der daraus resultierende Wind wird „Hangaufwind“genannt. Auch dieser Effekt dreht sich in der Nacht um und es kommt zu einem „Hangabwind“.
Durch die Hangauf- bzw. abwinde können noch weitere Winde in verzweigten Gebirgssystemen hervorgerufen werden. Auf Grund der Luftströmung entlang der Hänge und etwa in Talmitte kommt es hier tagsüber zu einem Talwind, nachts zu einem Bergwind. Nicht zu verwechseln ist dieser Effekt mit dem sogenannten Föhn. Diese Erscheinung tritt dann auf, wenn eine Luftströmung auf ein großes Hindernis, wie z.B. ein Gebirge stößt. Dabei kommt es zu Turbulenzen der Leeseite (windabgewandte Seite), welche, wenn sie in den Alpen auftreten, Föhn genannt werden.
Neben diesen regionalen Systemen kann noch zwischen weiteren, wie z.B. dem Stadt–Umland–Windsystem oder dem Wirbelwindsystem, unterschieden werden. Diese haben allerdings auf die Modellierung der österreichischen Sturmschadensdaten in dieser Arbeit keinen großen Einfluss und werden deshalb nicht näher beschrieben (siehe stattdessen [49] oder [22]). Die oben genannten regionalen Windsysteme sind im Allgemeinen von keinen hohen Windgeschwindigkeiten geprägt. Allerdings können sie durch größerflächige Winde verstärkt werden.

2.2.3 Böigkeit und Düsenwirkung

Ein Effekt von Winden, der große Schäden verursachen kann, ist die Schwankung der Windgeschwindigkeit und -richtung. Diese Faktoren können sich für kurze Zeit stark ändern und man spricht dann von der Böigkeit eines Windes. Die Ursachen für Böen können unterschiedlich sein, so kann z.B. eine raue Bodenoberfläche zur Entstehung von Wirbeln führen. Starke Böen kommen natürlich dann auf, wenn die Höhenunterschiede der Oberflächen groß sind, wie z.B. an Hausdächern, Bäumen oder auch Gebirgen. Weiters können Böen auch dadurch hervorgerufen werden, dass die Geschwindigkeit des Windes mit der Höhe zunimmt. Absinkende Luftmassen bewegen sich dann in Bodennähe schneller als die umliegende Luft, die deshalb kurz beschleunigt wird.
Ein Problem, welches sich aus der Böigkeit des Windes ergibt, ist die schlechte Messbarkeit der Windgeschwindigkeit während einer Böe. Da diese örtlich und zeitlich stark begrenzt sind, kann man nicht aus der Windgeschwindigkeit eines Punktes einer Region auf die Geschwindigkeiten im übrigen Teil des Gebietes schließen.
Eine weitere Verstärkung der Windgeschwindigkeit resultiert aus der Düsenwirkung. Dieser

Effekt entsteht dann, wenn sich ein Tal plötzlich verengt. Somit werden die durchziehenden Luftströmungen aufgehalten und der Luftdruck erhöht. Dasselbe kann passieren, wenn die Windströmung durch ein Hindernis seitlich abgelenkt wird. Weiters kann Wind in Straßenzügen kanalisiert werden, in diesem Fall sind aber die Ausmaße der Verstärkung nicht so groß. Das bekannteste Beispiel eines Windes, bei dem ein solcher Effekt auftritt, ist der Fallwind Mistral im Rhônetal.

2.2.4 Windgeschwindigkeit

Wie in den vorherigen Punkten zur Wind-Enstehung erklärt wurde, handelt es sich bei Wind um die Bewegung von Luftteilchen. Diese kann man mithilfe eines Vektors, der die Richtung der Bewegung beschreibt, darstellen. Die Länge dieses Vektors ist der zurückgelegte Weg des Luftteilchens und kann, normiert auf eine gewisse Zeiteinheit wie Minuten oder Stunden, als Windgeschwindigkeit interpretiert werden.
Eine weitverbreitete Einteilung der Windgeschwindigkeiten bietet die sogenannte Beaufort-Skala. Sie ist nach Admiral Francis Beaufort benannt, der die Skala als Methode verwendete, Windgeschwindigkeiten nach ihren Auswirkungen auf den Seegang zu klassifizieren. Später wurde die Beaufort-Skala auch für das Festland erweitert. Diese Erweiterung ist in Tabelle 2.4 beschrieben.

2.2.5 Windschäden

Das Auftreffen des Windes auf feststehende Körper, wie z.B. Häuser oder Bäume, kann eine Druckeinwirkung hervorrufen. Die Höhe dieses Druckes ist abhängig von der Oberfläche des betroffenen Körpers. So schreibt [22], dass sich Bäume unter hohem Druck stromlinienförmig verformen können und somit einem geringeren Druck ausgeliefert sind. Dennoch können diese bei Sturm entwurzelt oder abgeknickt werden, was aber vor allem auf die Böenwirkung des Windes zurückzuführen ist. Dabei versetzen die Böen des Windes das Objekt in Schwingung. Diese können, wenn sie mit der Frequenz der Böen übereinstimmen, noch verstärkt werden. Besonders wichtig ist diese Art von Windschäden für die Forst- und Landwirtschaft. In weiterer Folge ist sie auch für die Sturmschadensversicherung von Privatgebäuden relevant, da die von einem Sturm umgeknickten Bäume auch auf Häuser fallen und diese somit beschädigen können.
Eine weitere Schadensform tritt durch die Sogeinwirkung von Winden auf. Dabei verringert sich der Druck an jenen Flächen, die nicht dem direkten Wind ausgesetzt sind, linear zum Quadrat der Windgeschwindigkeit (vgl. Bernoullische Gleichung). Durch diese Kräfte können teilweise ganze Hausdächer abgedeckt werden.
Die Einwirkung von Wind auf Gebäude wird auch als Windlast bezeichnet. Auch sie ist abhängig von der Oberfläche, welche dem Wind ausgesetzt ist. So entsteht auf der windzugewandten Seite (Frontalseite oder Luvseite) des Gebäudes Druck, auf der windabgewandten Seite (Leeseite) sowie an großen Teilen des Daches führt allerdings die Sogwirkung zu Schäden an Gebäuden. Dabei sind die einwirkenden Kräfte von der Windrichtung abhängig, da diese festlegt, auf welchen Teilen des Gebäudes Druck- bzw. Sogwirkungen entstehen.

Beaufort-Grad	Bezeichnung	Auswirkung des Windes im Binnenlande	Geschwindigkeit km/h
0	still	Windstille, Rauch steigt gerade empor	1
1	leiser Zug	Windrichtung angezeigt nur durch Zug des Rauches, aber nicht durch Windfahne.	1-5
2	leichte Brise	Wind am Gesicht fühlbar, Windfahne bewegt sich.	6-11
3	schwache Brise	Blätter und dünne Zweige bewegen sich, Wind streckt einen Wimpel.	12-19
4	mäßige Brise	Hebt Staub und loses Papier, bewegt Zweige und dünnere Äste.	20-28
5	frische Brise	Kleine Laubbäume beginnen zu schwanken. Schaumköpfe bilden sich auf Seen.	29-38
6	starker Wind	Starke Äste in Bewegung, Pfeifen in Telegraphen-Leitungen, Regenschirme schwierig zu benutzen.	39-49
7	steifer Wind	Ganze Bäume in Bewegung, fühlbare Hemmung beim Gehen gegen den Wind.	50-61
8	stürmischer Wind	Bricht Zweige von den Bäumen, erschwert erheblich das Gehen im Freien.	62-74
9	Sturm	Kleinere Schäden an Häusern (Rauchhauben und Dachziegel werden abgeworfen).	75-88
10	schwerer Sturm	Entwurzelt Bäume, bedeutende Schäden an Häusern.	89-102
11	orkanartiger Sturm	Verbreitete Sturmschäden (sehr selten im Binnenland).	103-117
12	Orkan	Schwerste Verwüstungen.	118-133

Tabelle 2.4: Ausschnitte aus der Beaufort-Skala nach [22], S. 401-402

Des Weiteren spielen auch hier Faktoren, wie z.B. die Topographie und Rauigkeit der Umgebung, eine bedeutende Rolle. Auf eine genaue Beschreibung dieser Kräfte soll in Rahmen dieser Arbeit verzichtet werden, siehe dazu z.B. [25].

Allerdings muss nicht jeder durch Wind verursachte Schaden ein Versicherungsfall sein. Viele Versicherungen bezeichnen einen Schaden erst als Sturmschaden, wenn eine gewisse Windgeschwindigkeit vorliegt. So definiert z.B. der Versicherungsverband Österreichs in seinen Musterbedingungen für Sturmschadenversicherungen die versicherten Gefahren ([54], Artikel 1.1) folgendermaßen:

> „**Sturm;** Sturm ist eine wetterbedingte Luftbewegung, deren Geschwindigkeit am Versicherungsort mehr als 60 Kilometer je Stunde beträgt.
> Für die Feststellung der Geschwindigkeit ist im Einzelfall die Auskunft der Zentralanstalt für Meteorologie und Geodynamik maßgebend.“

3 Mathematische Modellierung der Jährlichkeiten

3.1 Jährlichkeiten

Der Begriff der Jährlichkeit (return period) wird vor allem in der Hydrologie verwendet, um die Wahrscheinlichkeit extremer Hochwasser zu beschreiben. Betrachtet man ein solches Hochwasser mit einem bestimmten Pegelstand s, so definiert man dessen Jährlichkeit als Kehrwert der Wahrscheinlichkeit, dass ein Hochwasser mit demselben oder höherem Pegelstand in einer Beobachtungsperiode auftritt. Somit hat z.B. ein 100-jähriges Hochwasser eine Auftrittswahrscheinlichkeit von 1/100 pro Jahr.
Die Jährlichkeit eines Hochwassers kann relativ gut durch Messungen des Pegelstandes des zugrundeliegenden Flusses bestimmt werden. Auch die entstehenden Schäden können anhand der Jährlichkeiten des Pegelstandes beschrieben werden. So kann mit einer gewissen Wahrscheinlichkeit zumindest das überschwemmte Gebiet durch geografische und hydrologische Überlegungen ausgewiesen werden.
Da ein Sturm aber sehr variabel in Richtung und Geschwindigkeit ist, können durch Wind erzeugte Schäden nicht einfach an der Jährlichkeit der Windgeschwindigkeit gemessen werden. Des Weiteren können gemittelte Windgeschwindigkeiten aufgrund der Böigkeit des Windes nicht als Grundlage zur Bestimmung der entstandenen Schäden verwendet werden.

3.1.1 Wiederkehrperiode einer Folge von Zufallsvariablen

Eine formale Einführung der „*return periods*“ bietet [17], die hier kurz präsentiert werden soll. Die Bezeichnung „Jährlichkeit“ ist bei dieser Herangehensweise noch nicht angebracht, steht aber in direkter Relation.
Sei $(X_n)_n,\ n = 1, \dots, N$, eine Folge von unabhängigen Zufallsvariablen mit stetiger Verteilungsfunktion F und $u \in \mathbb{R}$ eine bestimmte Schranke. Die sogenannten Überschreitungszeitpunkte zum Level u („*exceedence times*“) bezeichnet jene Indizes $\tau_1, \dots, \tau_m$, an denen die Schranke u überschritten wird, d.h.

$$\tau_1 = \inf\{n \geq 1 | X_n \geq u\} \qquad \text{und} \qquad \tau_l = \inf\{n > \tau_{l-1} | X_n \geq u\}, \qquad l = 2, \dots, m$$

Bemerkung. Da natürlich zu jedem Zeitpunkt n bekannt ist, ob der Prozess (X_n) die Schranke u überschreitet, handelt es sich hier um Stoppzeiten.

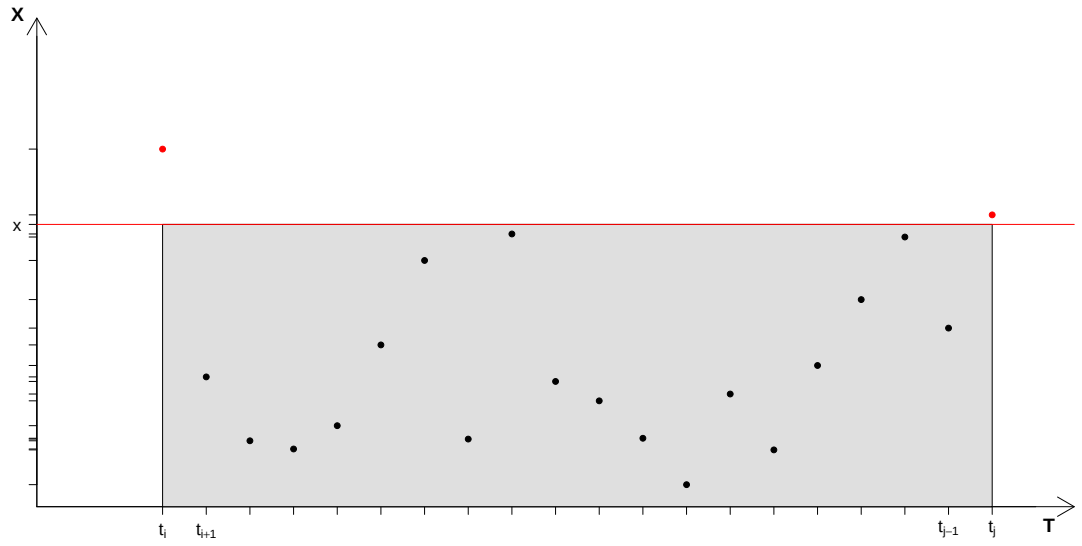


Abbildung 3.1: Return Period der Zufallsvariablen X_n zum Level x

Um die Jährlichkeiten wie in dem oben genannten Beispiel zu erhalten, ist man natürlich an der Anzahl der Indizes n_l zwischen zwei Überschreitungszeitpunkten τ_l und τ_{l+1}, d.h.

$$n_1 = \tau_1, \qquad \text{und} \qquad n_l = \tau_l - \tau_{l-1}, \qquad l = 2, \dots, m,$$

interessiert. Durch die iid-Eigenschaft der X_n, $n = 1, \dots, N$ sind auch $n_1, \dots, n_m$ unabhängig und identisch verteilt. Weiters gilt für die Zeit der ersten Überschreitung n_1, dass

$$\begin{aligned}
\mathbb{P}[n_1 = k] &= \mathbb{P}[X_1 < u, \dots, X_{k-1} < u \text{ und } X_k \geq u] = (F(u))^{k-1}(1 - F(u)) \\
\mathbb{P}[n_1 \leq k] &= \sum_{j=1}^{k} \mathbb{P}[X_1 < u, \dots, X_{l-1} < u \text{ und } X_l \geq u] \\
&= \sum_{j=1}^{k} (F(u))^{l-1}(1 - F(u)) = (1 - F(u))\frac{(F(u))^k - 1}{F(u) - 1} \\
&= 1 - (F(u))^k.
\end{aligned}$$

Somit folgen die Zeiten zwischen den Überschreitungen einer geometrischen Verteilung mit Erfolgswahrscheinlichkeit $p = 1 - F(u)$. Der Erwartungswert dieser Verteilung ist gegeben als

$$r(u) = \mathbb{E}[n_1] = \cdots = \mathbb{E}[n_m] = \frac{1}{p} = \frac{1}{1 - F(u)}. \tag{3.1}$$

Definition 3.1.1: Wiederkehrperiode („*return period*“)
Den Erwartungswert $r(u)$ aus (3.1) bezeichnet man als Wiederkehrperiode („return period“) der Schranke u.

3.1.2 Markierte Punktprozesse

Obwohl die Indizes der Überschreitungen Zeiten genannt wurden, enthält diese Beschreibung der Jährlichkeiten noch keine Modellierung der zufälligen Zeitpunkte, an denen extreme (Wetter-)Ereignisse eintreten. Um auch diesen stochastischen Faktor in die bisherigen Überlegungen aufzunehmen, betrachtet man nun die Folge von Zufallsvariablen $(X_n)_{1\leq n\leq N}$, welche zu den Zeitpunkten $(t_n)_{1\leq n}$ auftreten.

Auch hier möchte man die Anzahl der Punkte X_{t_i} zählen, welche zwischen zwei Überschreitungszeitpunkten τ_l und τ_{l+1} liegen. Dazu wird häufig ein markierter Punktprozess verwendet, der die aufgetretenen Realisationen (X_i, t_i) in einem bestimmten Rechteck bzgl. der Prozesse $(X_i)_i$ und $(t_i)_i$ zählt (vgl. Abbbildung 3.2). Eine bekannte Anwendung dieser Prozesse ist z.B. das Cramér-Lundberg-Modell. In diesem Modell soll die Frage beantwortet werden, ob und wann der Ruinzeitpunkt einer Versicherung eintritt, d.h. dass eingetretenen Forderungen an eine Versicherung deren Kapital überschreiten (vgl. dazu [9], Abschnitt 6.4).

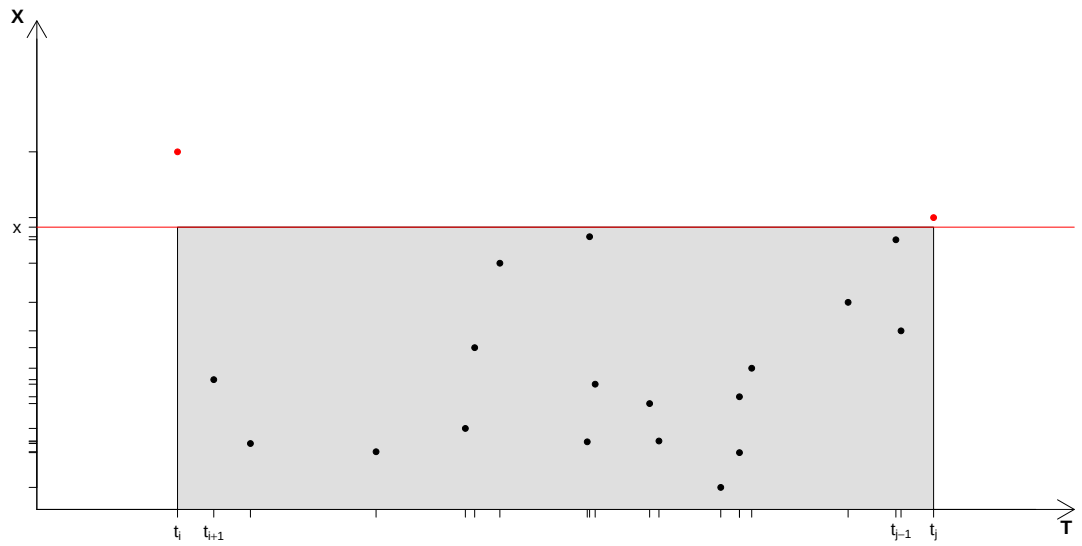


Abbildung 3.2: Markierter Punktprozess ($T \sim \text{Exp}(\lambda)$, $X \sim \mathcal{LN}$)

Die Theorie der Punktprozesse kann in [9], [10] und [29] nachgelesen werden. Hier sollen allerdings nur die wichtigsten Definitionen von Punktprozessen wiedergegeben werden. Bevor der markierte Punktprozess definiert werden kann, muss die einfachere Variante, der Punktprozess, , definiert werden (vgl. [10]).

Definition 3.1.2: Punktprozess
Sei T ein vollständiger, teilbarer metrischer Raum mit Borel'scher σ-Algebra $\mathcal{T}$. Bezeichne $\mathcal{M}_T$ den Raum der endlichen Maße auf $\mathcal{T}$ und $\mathcal{N}_T$ die Familie der Maße $N \in \mathcal{M}_T$, welche nur Werte in $\mathbb{N}_0$ annehmen (Zählmaße). Dann ist ein Punktprozess N eine messbare Abbildung vom Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{N}_T, \mathcal{B}(\mathcal{N}_T))$.

Definition 3.1.3: Träger eines Zählmaßes, Atome , einfaches Zählmaße
Sei N ein Zählmaß auf T. Dann bezeichnet man K_N, mit $K_N = \{t \in T : N(\{t\}) > 0\}$, als Träger des Zählmaßes N und die Elemente $\{t\}$ als Atome des Zählmaßes N.
Ein Zählmaß heißt einfach, wenn $N(\{t\}) = 1 \quad \forall\, \{t\} \in K_N$.

Diese Definition des Punktprozesses ist sehr theoretisch. Grundsätzlich können Punktprozesse auch als eine Folge von Zufallsvariablen oder zufälligen Intervallen dargestellt werden. Die Darstellung als Zählprozess wird durch folgenden Satz vereinfacht:

Satz 1:
Jedes Zählmaß $N \in N_T$ verfügt über einen höchstens abzählbar unendlichen Träger und für jede Borelmenge $B \in \mathcal{T}$ gilt:

$$N(B) = \sum_{t \in K_N} N(\{t\}) \mathbb{1}_{\{t \in B\}} \tag{3.2}$$

Beweis. siehe [29], S. 30.

Damit kann für die Folge von Zufallsvariablen aus Punkt 3.1.1 ein sogenannter „point process of exceedances“dargestellt werden:

$$N([u, \infty)) = \sum_{i=1}^{n} \mathbb{1}_{\{X_i \geq u\}} \tag{3.3}$$

Dieser Prozess ist allerdings nicht zur Messung der Zeit zwischen zwei Überschreitungen gedacht. Eher ist man an den Ereignissen $\{N([u, \infty)) = 0\}$ (vgl. Abschnitt 3.2.1) und $\{N([u, \infty)) \leq k\}$ (vgl. Abschnitt 3.2.2) interessiert. Für eine Diskussion dieser Prozesse, für (X_n) sowohl unabhängig als auch stationär, siehe [17] oder [42].

Definition 3.1.4: einfacher Punktprozess
Sei $\mathcal{N}_T^+$ die Familie der einfachen Zählmaße in $\mathcal{N}_T$. Dann heißt ein Punktprozess N einfach, wenn $N \in \mathcal{N}_T^+$ $f.s.$

Definition 3.1.5: *marked point process* (Markierter Punktprozess)
Der markierte Punktprozess N mit Positionen (*locations*) t im vollständigen, teilbaren metrischen Raum $\mathcal{T}$ und Markierungen (*marks*) k im vollständigen, teilbaren metrischen Raum $\mathcal{K}$ ist ein einfacher Punktprozess $\{(t_i, k_i)\}$ auf den Produktraum $\mathcal{B}(\mathcal{T} \times \mathcal{K})$ mit der Eigenschaft, dass der zugrundeliegende Prozess N_g von t selbst ein einfacher Punktprozess ist.

In der Definition des markierten Punktprozesses soll mit der Bezeichnung der t_i als „Positionen“ hervorgehoben werden, dass es sich nicht um Zeitpunkte handeln muss, sondern dass damit beliebige Zufallsvariablen und sogar multivariate Zufallsvektoren gemeint sein können. Deshalb finden markierte Punktprozesse häufig Anwendung in räumlichen Analysemethoden. So verwenden z.B. Holden et. al. in [24] einen markierten Punktprozess zur Analyse von Erdbeben. In deren Arbeit beschreiben die Positionen des verwendeten

markierten Punktprozesses die Zeitpunkte t und die geografische Position, $y = (y_1, y_2, y_3)$, bestehend aus Longitude, Latitude und Tiefe, des Erdbebens und die Magnituden M werden als Markierungen behandelt.
Weitere Anwendungen des markierten Punktprozesses zur Beschreibung extremer Wetterereignisse finden sich auch in [55] und [5]. In letzterem wird sogar eine doppelte Markierung verwendet, eine um die Dauer einer Flut zu messen und die andere zur Beschreibung der Abstände zwischen zwei Hochwassern.

3.1.3 Jährlichkeiten im markierten Punktprozess

Nun kann mithilfe der markierten Punktprozesse eine allgemeine Beschreibung der Jährlichkeit gefunden werden. Dazu beschreibt man ein Ereignis zum Zeitpunkt t_n als Realisation der Zufallsvariablen X_n (in dieser Arbeit wird das i-te Sturmereignis durch die Schadenhöhe X_i dargestellt). Wie in Punkt 3.1.1 ist man an jenen X_n interessiert, welche eine Schranke u überschreiten. Bezüglich dieser Schranke werden die Markierungen k_n des Prozesses,

$$k_n = \begin{cases} 1 & X_n \geq u \\ 0 & X_n < u. \end{cases}$$

festgelegt. Außerdem seien $\tau_1, \ldots, \tau_m$ die Zeitpunkte, an jenen $k_n = 1$ gilt.
Wenn verlangt ist, dass sowohl X als auch t unabhängig und identisch verteilt sind, liegt natürlich ein stationärer markierter Punktprozess vor. Wie in 3.1.1 gilt somit auch hier

$$((X_{\tau_i+1}, \tau_i + 1), (X_{\tau_i+2}, \tau_i + 2), \ldots, (X_{\tau_{i+1}}, \tau_{i+1})) \sim \cdots \sim ((X_1, t_1), (X_2, t_2), \ldots, X_{\tau_1}, \tau_1,) \tag{3.4}$$

Dabei gilt, dass jeder einfache Punktprozess mit stationären und unabhängigen Zuwächsen ein Poisson-Prozess ist.

Definition 3.1.6: Poisson-Prozess
Ein stochastischer Prozess N heißt Poisson-Prozess mit Intensität λ, wenn

1. $N(0) = 0$ f.s.
2. $N((t_1, t_2]) \sim Poi(\lambda(t_2 - t_1)) \qquad \forall t_1 < t_2 \in \mathbb{R}_0^+$
3. $N((t_1, t_2])$ und $N((t_3, t_4])$ sind unabhängig $\forall t_1 < t_2 < t_3 < t_4 \in \mathbb{R}_0^+$.

Somit ist die Dauer der Zeit zwischen τ_i und τ_{i+1} ein zusammengesetzter Poisson-Prozess mit Intensität $\lambda = \mathbb{E}[\tau_1]$. Dessen Erwartungswert kann einfach über die bedingten Wahrscheinlichkeiten berechnet werden.

$$\begin{aligned}\tilde{r}(u) &= \mathbb{E}[\sum_{k=\tau_{m-1}+1}^{\tau_m} t_k] \overset{(3.4)}{=} \cdots = \mathbb{E}[\sum_{k=\tau_1+1}^{\tau_2} t_k] \overset{(3.4)}{=} \mathbb{E}[\sum_{k=1}^{\tau_1} t_k] \\ &= \mathbb{E}[\tau_1]\mathbb{E}[t_1] = \mathbb{E}\left[\sum_{k=1}^{n} t_k \middle| \tau_1 = n\right] = \mathbb{E}[t_1] \sum_{n=0}^{\infty} n\mathbb{P}[\tau_1 = n] \\ &= \lambda\mathbb{E}[t_1]\end{aligned}$$

Da X und t unabhängig sind, kann der Erwartungswert der ersten Stoppzeit, λ, auf die Wiederkehrperiode $r(u)$ zurückgeführt werden. Somit gilt

$$\tilde{r}(u) = \mathbb{E}[\tau_1]\mathbb{E}[t_1] = \mathbb{E}[t_1]\frac{1}{1 - F(u)} \tag{3.5}$$

Sind die Zeitpunkte t_i auf ein Jahr normiert, kann damit die Jährlichkeit definiert werden:

Definition 3.1.7: Jährlichkeit
Die Jährlichkeit der Schranke u, J_u, ist gegeben als die erwartete Anzahl von Jahren, bis ein Ereignis X die Schranke u wiederereicht. D.h. $J_u = \tilde{r}(u)$.

Ist auch der Punktprozess der Markierungen ein Poisson-Prozess, dann folgt die Zeit zwischen den Markierungen des Prozesses einer Exponentialverteilung mit Parameter λ_1. In diesem Fall ist der Erwartungswert von t_1 als $1/\lambda_1$ gegeben.

In [47], Abschnitt 1.4, wird diese Definition der Jährlichkeiten verwendet, um das Risiko von verschiedenen Naturgefahren abzuschätzen. Dort wird auch die Möglichkeit besprochen, nicht die Zeit der Überschreitungen, sondern der Unterschreitungen zu messen. Dieser Ansatz kann z.B. zur Modellierung der Jährlichkeiten von Dürreperioden verwendet werden.

Bemerkung. In dieser Arbeit sind die X_i die Schäden bzw. Schadensgrade (vgl. 5.1) der Sturmereignisse zu den Zeitpunkten i, $i = 1, \ldots, 36$. In diesem Fall bezeichnet $L(u)$ den ersten Zeitpunkt, an dem der Schaden eine bestimmte Schranke u überschreitet.

3.2 Extremwerttheorie

Häufig ist man aber nicht nur an der Jährlichkeit einer bestimmten Schranke interessiert, sondern an der Verteilung der Zufallsvariablen, die diese Schranke überschreiten. Um Aussagen über die Verteilungsfunktion jener X zu erhalten, welche die Schranke u überschreiten, müssen Annahmen über das Verhalten der Maxima einer Stichprobe getroffen werden. Dieses Gebiet der Stochastik wird Extremwerttheorie genannt.
Eine direkte Anwendung auf den Versicherungsverlust durch Feuerschäden findet sich in [33]. Hier werden sowohl eine untere als auch eine obere Schranke festgelegt um den Preis

eines sogenannten „loss layers“ von Versicherungen zu berechnen.
Im Folgenden soll ein kurzer Überblick über die Grenzverteilungen von extremen Zufallsvariablen gegeben werden. Eine detaillierte Beschreibung findet sich z.B. in [17] und [43], verschiedene Anwendungsmöglichkeiten und statistische Analysen können [6] bzw. [42] nachgelesen werden.

3.2.1 Grenzwertverteilung der Maxima

Im ersten Schritt soll die Verteilung des Maximums, M_n, einer n-elementigen Stichprobe aus unabhängigen und identisch verteilten Zufallsvariablen $(X_n)_n$,

$$M_n = \max_{1 \leq i \leq n} \{X_1 \ldots, X_n\},$$

betrachtet werden. In Abbildung 3.3 ist eine Simulation von 30 unabhängigen und lognormalverteilten Zufallsvariablen abgebildet. In dieser Realisation gilt

$$M_1 = 1,\ M_2 = M_3 = \ldots M_{16} = 2,\ M_{17} = \ldots = M_{21} = 17.$$

Nun stellt sich die Frage, ob noch weitere Stichprobenmaxima auftreten und unter welchen Voraussetzungen man etwas über ihre Verteilung sagen kann.

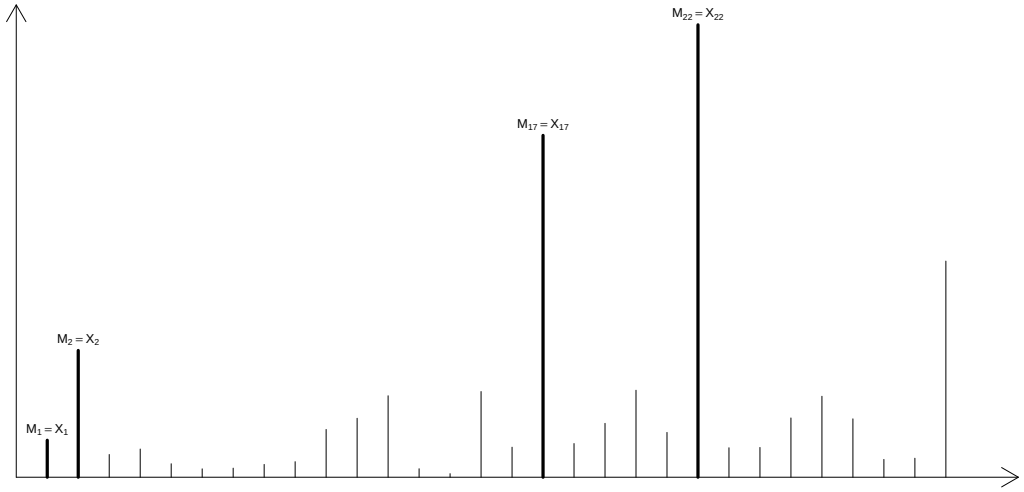


Abbildung 3.3: Der Prozess M_n der Stichprobenmaxima

Die Verteilung der Stichprobenmaxima von unabhängigen Zufallsvariablen ist zwar einfach zu berechnen, aber leider nicht sehr aussagekräftig. So gilt für die Verteilung des Maximums M_n :

$$\mathbb{P}[M_n \leq x] = \mathbb{P}[X_1 \leq x, \ldots, X_n \leq x] = (F(x))^n = \begin{cases} 1 & F(x) = 1 \\ 0 & F(x) < 1 \end{cases} \tag{3.6}$$

Die obige Verteilung kann auch als

$$\mathbb{P}[M_n \leq x] = \begin{cases} 1 & x \geq x_F \\ 0 & x < x_F \end{cases}$$

dargestellt werden, wobei x_F den rechten Endpunkt der Verteilung F bezeichnet.

Definition 3.2.1: rechter Endpunkt einer Verteilung
Sei F eine Verteilungsfunktion. Dann bezeichnet man mit x_F,

$$x_F = \sup\{x \in \mathbb{R} | F(x) < 1\}$$

den rechten Endpunkt der Verteilung.

Eine Verteilungsfunktion F mit der Eigenschaft aus (3.6) gehört zur Klasse der sogenannten „degenerierten“Verteilungen.

Definition 3.2.2: degenerierte Verteilung
Eine Verteilung mit Verteilungsfunktion F heißt degeneriert, wenn

$$\exists x_F \in \mathbb{R}: \qquad G(x) = \begin{cases} 1 & x \geq x_F \\ 0 & x < x_F \end{cases}$$

Gesucht werden also Folgen von normierenden Konstanten, $(a_n)_n \in \mathbb{R}$ und $(b_n)_n > 0$, sodass die Grenzverteilung der normierten Maximuma nicht mehr degeneriert ist. Außerdem ist man an der Grenzverteilung des normierten Stichprobenmaximums interessiert, d.h. man sucht eine Verteilungsfunktion H, die Folgendes erfüllt.

$$\mathbb{P}\left[\frac{M_n - a_n}{b_n} \leq x\right] = \mathbb{P}[M_n \leq \underbrace{b_n x + a_n}_{u_n}] = F^n(u_n) \to H(x)$$

Definition 3.2.3: Maximaler Anziehungsbereich (MDA)
Sei X eine Zufallsvariable. Dann ist X aus dem maximalen Anziehungsbereich einer Verteilungsfunkion H, $X \in MDA(H)$, wenn Folgen von Konstanten $(a_n)_n \in \mathbb{R}$ und $(b_n)_n > 0$ existieren, sodass

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{M_n - a_n}{b_n}\right] = H(x)$$

Definition 3.2.4: Tail einer Verteilung
Sei F eine Verteilungsfunktion. Die Überschreitungswahrscheinlichkeit

$$\mathbb{P}[X > x] = 1 - F(x) =: \bar{F}(x) \qquad \forall x \in \mathbb{R}$$

heißt Tail der Verteilungsfunktion.

Nicht jede Verteilung muss im maximalen Anziehungsbereich einer Grenzverteilung H liegen. Ein mögliches Kriterium, ob normierende Konstantenfolgen gefunden werden können, liefert das folgende Lemma:

Lemma 2: Poisson-Approximation
Sei $\tau \in [0, \infty]$ und $(u_n)_n$ eine Folge reeler Zahlen. Dann gilt

$$n\bar{F}(u_n) \overset{n\to\infty}{\longrightarrow} \tau \qquad \Longleftrightarrow \qquad \mathbb{P}\left[M_n \leq u_n\right] \overset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\tau}.$$

Beweis.

1. Sei $\tau \in [0, \infty)$:
   a) „$\Rightarrow$“: Sei τ der Grenzwert von $n\bar{F}(u_n)$ für $n \to \infty$. Dann gilt:

$$\mathbb{P}[M_n \leq u_n] = F^n(u_n) = \Big(1 - \frac{1}{n}\underbrace{(n\bar{F}(u_n))}_{\to\tau}\Big)^n \longrightarrow \mathrm{e}^{-\tau}$$

   b) „$\Leftarrow$“: Sei nun $\mathbb{P}\left[M_n \leq u_n\right] \overset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\tau}$ erfüllt. Damit muss auch $\bar{F} \to 0$ gelten. Ann: $\bar{F} \nrightarrow 0$, d.h. es existiert eine Teilfolge n_k, sodass $\bar{F}(u_{n_k}) \geq m > 0$. Damit folgt aber, dass

$$\mathbb{P}[M_{n_k} < u_{n_k}] = \big(1 - \bar{F}(u_{n_k}\big)^{n_k} \leq (\underbrace{1-m}_{<1})^{n_k} \overset{k\to\infty}{\longrightarrow} 0.$$

   Somit muss $\bar{F}$ gegen 0 streben. Da für ein kleines x die Funktion $log(1-x) \sim x$ ist, gilt

$$-\tau = \lim_{n\to\infty} \log\left(\mathbb{P}[M_n < u_n]\right) = \lim_{n\to\infty} n \log\big(1 - \bar{F}(u_n)\big) = -\lim_{n\to\infty} n\bar{F}(u_n).$$

2. Sei $\tau = \infty$:
   a) „$\Rightarrow$“: Sei $\lim\limits_{n\to\infty} n\bar{F}(u_n) = \infty$, aber $\lim\limits_{n\to\infty} \mathbb{P}[M_n < u_n] > 0$. Dann existiert eine Teilfolge n_k, sodass

$$\lim_{k\to\infty} \mathbb{P}[M_{n_k} \leq u_{n_k}] = \mathrm{e}^{-\upsilon} \text{ mit } \upsilon < \infty \overset{Fall1b}{\Longrightarrow} \lim_{k\to\infty} n_k\bar{F}(u_{n_k}) = \upsilon < \infty.$$

   Das ist aber ein Widerspruch zur Annahme, dass $\lim\limits_{n\to\infty} n\bar{F}(u_n) = \infty$.

   b) „$\Leftarrow$“: Ann: $\mathbb{P}[M_n \leq u_n] \to 0$, aber

$$\lim_{n\to\infty} n\bar{F}(u_n) \neq \infty \to \exists(n_k)_k : \lim_{k\to\infty} n_k\bar{F}(u_{n_k}) = \upsilon < \infty$$

   Nun würde aber laut Fall 1a gelten, dass $\mathbb{P}[M_{n_k} \leq u_{n_k}] \to e^{\upsilon} < \infty$.

Durch das obige Lemma kann nun direkt die Grenzverteilung H mit dem Tail der ursprünglichen Verteilungsfunktion F in Verbindung gebracht werden:

Satz 3:
Seien F und H Verteilungsfunktionen. Dann ist F aus $MDA(H)$ mit $b_n > 0$, $a_n \in \mathbb{R}$, genau dann, wenn

$$\lim_{n\to\infty} n\bar{F}(b_n x + a_n) = -\log(H(x))$$

Definition 3.2.5: vom selben Typ
Seien X und Y zwei Zufallsvariablen mit Verteilungsfunktionen F_X und F_Y. Dann gehören X und Y (bzw. F_X und F_Y) zum selben Typ, wenn Konstanten $\sigma > 0$ und $\mu \in \mathbb{R}$ existieren, sodass

$$F_X(z) = F_Y(\mu + \sigma z) \quad \Longleftrightarrow \quad X \overset{d}{\sim} \frac{Y - \mu}{\sigma}.$$

Durch den Satz von Fisher-Tippet können die möglichen Grenzverteilungen H in verschiedene Klassen von Verteilungsfunktionen eingeteilt werden.

Satz 4: Fisher-Tippett
Seien $(X_n)_n$ eine Folge von Zufallsvariablen mit $X_n \overset{iid}{\sim} F$ und M_n das Stichprobenmaximum. Ist $F \in MDA(H)$, dann gehört H zum Typ einer der folgenden Verteilungen:

1. **Frechet**

$$\Phi_\alpha(x) = \begin{cases} \exp\{-x^{-\alpha}\} & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad \alpha > 0$$

2. **Gumbel**

$$\Lambda(x) = \exp\left\{-\mathrm{e}^{-x}\right\} \qquad x \in \mathbb{R}$$

3. **Weibull**

$$\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & x < 0 \\ 1 & x \geq 0 \end{cases} \qquad \alpha > 0.$$

Dabei macht es große Unterschiede, zu welcher Klasse die Grenzverteilung des Maximums gehört. Eine mögliche Einteilung liefert Satz 5. Die genaue Herleitung dieser Bedingungen kann in [17] nachgelesen werden. Dort werden auch die sogenannten Von-Mises-Funktionen behandelt. Anhand des Grenzverhaltens der Dichte bzw. der Verteilung solcher Funktionen kann bestimmt werden, zu welchem Anziehungsgebiet eine Verteilung gehört. In diesem Satz ist auch eine mögliche Wahl der normierenden Konstanten (a_n) und (b_n) beschrieben.

Definition 3.2.6: regulär variierend, langsam variierend
Eine messbare Funktion $F : (0, \infty) \to (0, \infty)$ heißt regulär variierend mit Index $\alpha \in \mathbb{R}$, $F \in$

RV_α , wenn

$$\lim_{t\to\infty} \frac{F(xt)}{F(t)} = x^\alpha. \qquad \forall x > 0$$

Ist $\alpha = 0$, spricht man von einer langsam variierenden Funktion.

Bemerkung. Im Zusammenhang mit regulärer Variation tritt auch häufig der Begriff der „heavy tailed“ Verteilung auf. Die Definition eines „heavy tails“ ist in der Literatur nicht eindeutig. Oft wird eine Verteilung dann als „heavy tailed“ bezeichnet, wenn

$$\lim_{t\to\infty} \frac{\bar{F}(t)}{\exp\{-xt\}} = \infty \qquad \forall x > 0.$$

In diesem Fall sind die Momente der Verteilung nicht endlich. Andere Definitionen verlangen, dass $\bar{F} \sim x^{-\alpha}$ mit $0 < \alpha < 2$ oder auch nur $\alpha > 0$ für $x \to \infty$ gilt. Wichtig ist vor allem die Eigenschaft, dass in diesen Definition die Momente der Verteilung mit einer höheren Ordnung als α nicht exisiteren. Das hat vor allem Auswirkungen auf die Versicherbarkeit von Risiken, da nur jene mit einem endlichen ersten Moment als versicherbar gelten.

Satz 5: Charakterisierung der Anziehungsgebiete
Sei F eine Verteilungsfunktion. Dann gilt:

1. $F \in MDA(\Phi_\alpha), \alpha > 0 \iff x_R = \infty$ und $\bar{F}(x) \in RV_{-\alpha}$.
   In diesem Fall können die Folgen (a_n) und (b_n) folgendermaßen gewählt werden:

$$b_n = F^{[-1]}\left(1 - \frac{1}{n}\right) \quad \text{und} \quad a_n = 0.$$

2. $F \in MDA(\Psi_\alpha), \alpha > 0 \iff x_F < \infty$ und $1 - F\left(x_f - \frac{1}{x}\right) \in RV_{-\alpha}$.
   Dabei sind

$$b_n = x_F - F^{[-1]}\left(1 - \frac{1}{n}\right) \quad \text{und} \quad a_n = x_F.$$

   eine mögliche Wahl der Folgen von normierenden Konstanten.

3. $F \in MDA(\Lambda) \iff$ eine positive, messbare Funktion $a(x)$ existiert, für die gilt

$$\lim_{x\to x_F} \frac{\bar{F}(x + ta(x))}{\bar{F}(x)} = \mathrm{e}^{-t} \qquad \forall t \in \mathbb{R}.$$

   Dann können $a,\ a_n,\ b_n$ folgendermaßen gewählt werden:

$$a(x) = \int_x^{x_F} \frac{\bar{F}(t)}{\bar{F}(x)} dt = \mathbb{E}[X - x | X > x] \tag{3.7}$$

$$a_n = F^{[-1]}\left(1 - \frac{1}{n}\right), \qquad b_n = a(a_n).$$

Bemerkung. Die Funktion $a(x)$ aus (3.7) ist durch die Darstellung der Verteilungsfunktion als Von-Mises-Funktion gegeben (siehe [17]) und wird dort „auxiliary function“ genannt. Diese mögliche Wahl als Erwartungswert wird als „mean-excess-function“ bezeichnet.

Beispiel 1 (Pareto-Verteilung). Der Tail der Pareto-Verteilung ist $\bar{F}(x) = \left(\frac{m}{x}\right)^{\alpha}$ mit $m, \alpha > 0$. Es gilt $\bar{F}(x) \in RV_{-\alpha},\ \alpha > 0$, denn:

$$\lim_{t\to\infty} \frac{\bar{F}(xt)}{\bar{F}(t)} = \lim_{t\to\infty} \frac{\left(\frac{1}{x}\right)^{\alpha}\left(\frac{m}{t}\right)^{\alpha}}{\left(\frac{m}{t}\right)^{\alpha}} = x^{-\alpha}.$$

Damit gehört die Pareto-Verteilung zum Anziehungsgebiet der Frechet-Verteilung und die normierenden Folgen sind:

$$a_n = 0 \quad \text{und} \quad b_n = F^{[-1]}\left(1 - \frac{1}{n}\right) = m\left(1 - 1 + \frac{1}{n}\right)^{-\frac{1}{\alpha}} = mn^{\frac{1}{\alpha}}$$

Weitere Beispiele der Frechet-Verteilung: Cauchy, Loggamma.

Beispiel 2 (Gleichverteilung). Sei F die Verteilungsfunktion einer Gleichverteilung auf $[0, 1]$, d.h. $x_F = 1$. Dann gilt $\bar{F} = -x$ und

$$\lim_{t\to\infty} \frac{\bar{F}\left(x_F - \frac{1}{xt}\right)}{\bar{F}\left(x_F - \frac{1}{x}\right)} = \lim_{t\to\infty} \frac{1 - (xt)^{-1}}{1 - (x)^{-1}} = \lim_{t\to\infty} \frac{1}{x}\frac{x - \frac{1}{t}}{1 - \frac{1}{t}} = 1 \ \Rightarrow\ F \in MDA(\Psi_1).$$

Die normierenden Folgen sind dann:

$$a_n = x_F = 1 \quad \text{und} \quad b_n = 1 - F^{[-1]}\left(1 - \frac{1}{n}\right) = 1 - 1 + \frac{1}{n} = \frac{1}{n}$$

Bemerkung. Ein weiteres Beispiel aus dem Weibull-Anziehungsbereich ist die Beta-Verteilung. Häufig tritt die Weibull-Verteilung jedoch dann auf, wenn man eine Grenzverteilung für das Minimum einer Stichprobe sucht. Die Aussagen aus diesem Abschnitt sind prinzipiell leicht umlegbar, denn es gilt

$$\mathbb{P}[\min\{X_1, \ldots, X_n\} \le x] = \mathbb{P}[-\max\{-X_1, \ldots, -X_n\} \le x].$$

Dort ist sie die Grenzverteilung von z.B. Gamma, Exponential, Pareto.

Beispiel 3 (Exponential-Verteilung). Für $X \sim Exp(\lambda)$, mit $F(x) = 1 - \exp\{-\lambda x\}$ und $x_F = \infty$, ist die Funktion $a(x)$ aus Satz 5

$$a(x) = \int_x^{x_f} \frac{\bar{F}(t)}{\bar{F}(x)} = \int_x^{\infty} \frac{\mathrm{e}^{-\lambda t}}{\mathrm{e}^{-\lambda x}} dx = -\frac{1}{\lambda}\left(\frac{0 - \mathrm{e}^{-\lambda x}}{\mathrm{e}^{-\lambda x}}\right) = \frac{1}{\lambda}.$$

Somit ist $F \in MDA(\Lambda)$, denn

$$\lim_{x\to x_F} \frac{\bar{F}(x + a(x)t)}{\bar{F}(x)} = \lim_{x\to\infty} \frac{\exp\{-\lambda(x + t\lambda)\}}{\exp\{-\lambda x\}} = \mathrm{e}^{-t} \qquad \forall t \in \mathbb{R}$$

mit normierenden Konstantenfolgen

$$a_n = F^{[-1]}\left(1 - \frac{1}{n}\right) = -\frac{\log\left(1 - 1 + \frac{1}{n}\right)}{\lambda} = \frac{\log(n)}{\lambda} \quad \text{und} \quad b_n = a(a_n) = \frac{1}{\lambda}$$

Weitere Beispiele: Normal, Lognormal, Gamma

Häufig werden die Verteilungen aus Satz 4 zu einer Verteilungsfunktion, der generalisierten Extremwertverteilung, zusammengefasst:

Definition 3.2.7: Generalisierte Extremwertverteilung
Die Verteilungsfunktion H_γ,

$$H_\gamma(x) = \begin{cases} \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\} & \gamma \neq 0 \\ \exp\left\{-\exp\left\{-x\right\}\right\} & \gamma = 0 \end{cases},$$

mit $1 + \gamma x > 0$, d.h.

1. $\gamma > 0 \iff x \in (-\infty, -\frac{1}{\gamma})$
2. $\gamma = 0 \iff x \in \mathbb{R}$
3. $\gamma < 0 \iff x \in (-\frac{1}{\gamma}, \infty)$,

wird als Generalisierte Extremwertverteilung bezeichnet. Oftmals wird durch Einführung eines Lagen- und Skalenparameters, μ und σ, die Verteilungsfunktion folgendermaßen dargestellt:

$$H_{\gamma,\mu,\sigma} = H_\gamma\left(\frac{x-\mu}{\sigma}\right) = \begin{cases} \exp\left\{-(1+\gamma\frac{(x-\mu)}{\sigma})^{-1/\gamma}\right\} & \gamma \neq 0 \\ \exp\left\{-\exp\left\{-\frac{(x-\mu)}{\sigma}\right\}\right\} & \gamma = 0, \end{cases}$$

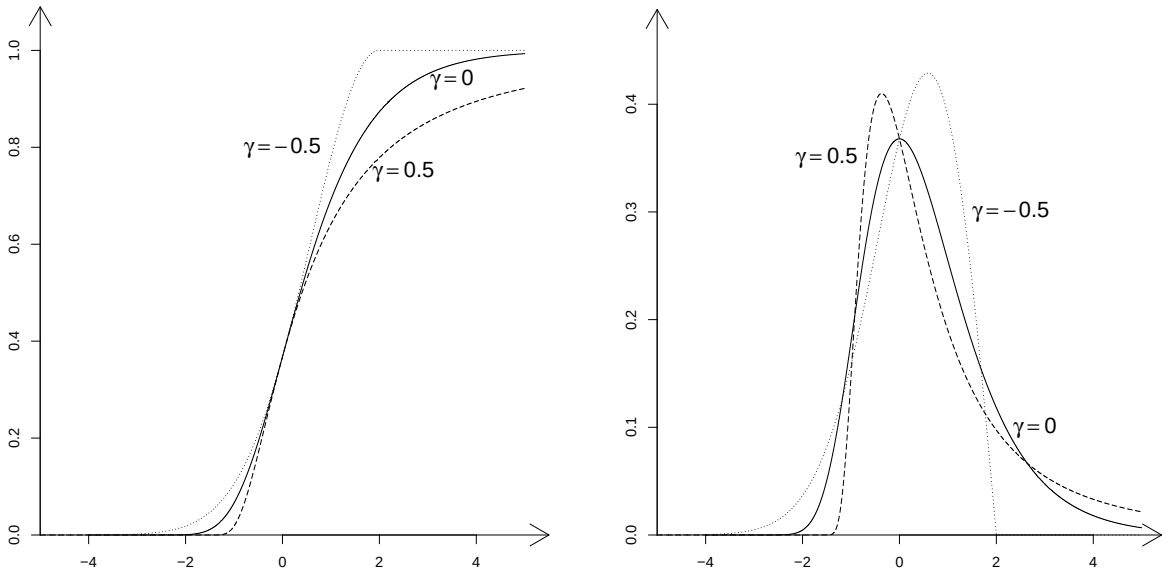
für $1 + \gamma(x-\mu)/\sigma > 0$, also

1. $\gamma > 0 \implies x \in (-\infty, \mu - \frac{\sigma}{\gamma})$
2. $\gamma = 0 \implies x \in \mathbb{R}$
3. $\gamma < 0 \implies x \in (\mu - \frac{\sigma}{\gamma}, \infty)$.

In Abbildung 3.4 sind die Verteilungs- und die Dichtefunktion der Generalisierten Extremwertverteilung für verschiedene Formparameter γ dargestellt. Am rechten Rand kann man das heavy-tail Verhalten der Verteilungsfunktion für $\gamma > 0$ gut erkennen. Auch der rechte Endpunkt der Generalisierten Extremwertverteilung im Fall $\gamma < 0$ ist deutlich zu sehen.

Bemerkung. Die Grenzverteilungen aus Satz 4 hängen folgendermaßen mit der Generalisierten Extremwertverteilung zusammen:

1. $F \in MDA(H_\gamma)$ mit $\gamma > 0 \iff x \in MDA(\Phi_\alpha)$ mit $\alpha = \frac{1}{\gamma} > 0$.



(a) Verteilungsfunktion

(b) Dichtefunktion

Abbildung 3.4: Generalisierte Extremwertverteilung ($\mu = 0,\ \sigma = 1$)

2. $F \in MDA(H_\gamma)$ mit $\gamma = 0 \iff x \in MDA(\Lambda)$.

3. $F \in MDA(H_\gamma)$ mit $\gamma < 0 \iff x \in MDA(\Psi_\alpha)$ mit $\alpha = -\frac{1}{\gamma} > 0$.

Satz 6: Charakterisierung des Anziehungsgebietes von H_γ
Sei F eine Verteilungsfunktion mit rechtem Endpunkt x_F. Dann ist $F \in MDA(H_\gamma)$ dann und nur dann, wenn eine messbare Funktion $a > 0$ existiert, sodass

$$\lim_{u \to x_F} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \begin{cases} (1+\gamma x)^{-\frac{1}{\gamma}} & \gamma \neq 0 \\ \mathrm{e}^{-x} & \gamma = 0. \end{cases} \qquad \text{für } 1 + \gamma x > 0$$

Beweis. siehe [17].

3.2.2 Überschussverteilung

Um die Verteilung jener Zufallsvariablen zu erhalten, die eine gewisse Schranke u überschreiten, betrachtet man die Bedingung aus Satz 6 genauer:

$$\frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \frac{\mathbb{P}[X > u + a(u)x]}{\mathbb{P}[X > u]} = \mathbb{P}[X > u + a(u)x | x > u] = \mathbb{P}\left[\frac{X - u}{a(u)} > x | X > u\right] \tag{3.8}$$

Somit ist durch den Grenzwert in Satz 6 schon eine Grenzverteilung für normierte Zufallsvariablen, welche eine Schranke u überschreiten, gegeben. Diese wird als Generalisierte Pareto-Verteilung bezeichnet.

Definition 3.2.8: Generalisierte Pareto Verteilung
Die Verteilungsfunktion der Generalisierten Pareto Verteilung (generalized Pareto distribution, GPD), G_γ, ist definiert als

$$G_\gamma(x) = \begin{cases} 1 - (1+\gamma x)^{-1/\gamma} & \gamma \neq 0 \\ 1 - \exp\{-x\} & \gamma = 0, \end{cases}$$

mit $x \geq 0$ und $1 + \gamma x > 0$, d.h.

1. $\gamma > 0 \implies x \in [0, \infty)$
2. $\gamma = 0 \implies x \in [0, \infty)$
3. $\gamma < 0 \implies x \in [0, -\frac{1}{\gamma}]$.

Auch hier ist es wieder möglich, Lage- und Skalenparamter $\mu \in \mathbb{R}$ und $\sigma > 0$ einzuführen, sodass die Verteilungsfunktion $G_{\gamma,\mu,\sigma}(x) = G_\gamma((x-\mu)/\sigma)$ gegeben ist durch

$$G_{\gamma,\mu,\sigma}(x) = \begin{cases} 1 - \left(1 + \frac{\gamma(x-\mu)}{\sigma}\right)^{-1/\gamma} & \gamma \neq 0 \\ 1 - \exp\left\{-\frac{x-\mu}{\sigma}\right\} & \gamma = 0, \end{cases}$$

wobei $(x-\mu)/\sigma \geq 0$ und $1 + \gamma(x-\mu)/\sigma \geq 0$, d.h.

1. $\gamma > 0 \implies x \in [\mu, \infty)$
2. $\gamma = 0 \implies x \in [\mu, \infty)$
3. $\gamma < 0 \implies x \in [\mu, \mu - \frac{\sigma}{\gamma}]$.

Die Verteilungs- und Dichtefunktion der Generalisierten Pareto-Verteilung sind in Abbildung 3.5 dargestellt.

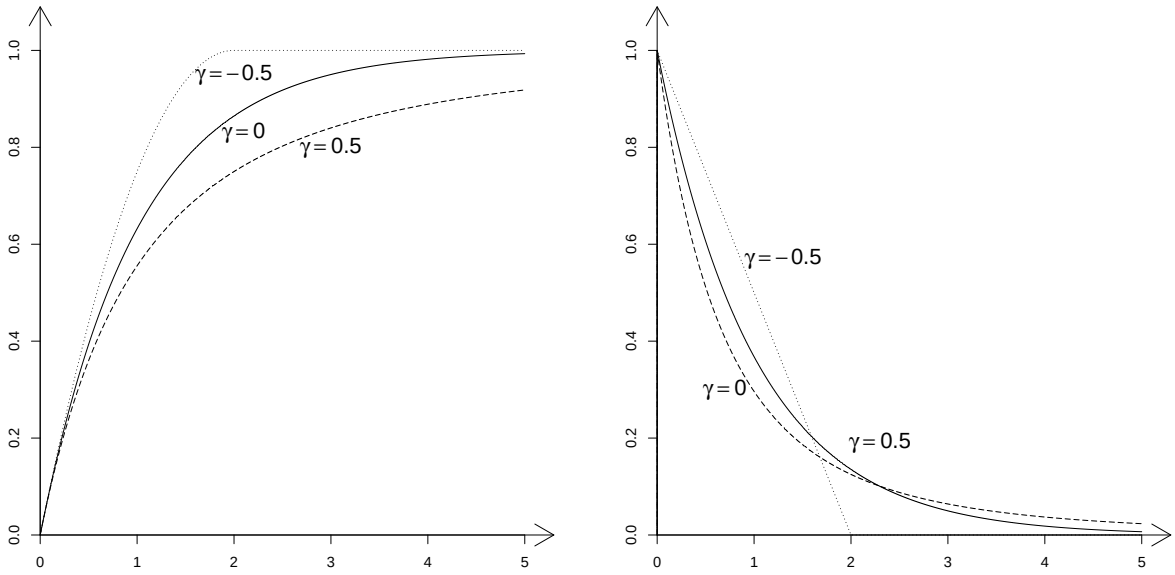
Eine der wichtigsten Eigenschaften der Generalisierten Pareto- Verteilung ist durch den Satz von Pickands-Balkema-de Haan gegeben. Dazu muss aber zuerst die sogenannte Excess-Funktion definiert werden.

Definition 3.2.9: Excess-Funktion

$$F_u(x) = \mathbb{P}[X - u \leq x | X > u] = \frac{F(u+x) - F(u)}{\bar{F}(u)}$$

Satz 7: Pickands-Balkema-de Haan
Sei F eine Verteilung mit rechtem Endpunkt x_F. Dann ist $F \in MDA(H_\gamma)$ für $\gamma \in \mathbb{R}$, genau dann, wenn für eine geeignete Funktion $b > 0$ gilt:

$$\lim_{u \to x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\gamma,0,b(u)}(x)| = 0$$



(a) Verteilungsfunktion

(b) Dichtefunktion

Abbildung 3.5: Generalisierte Pareto Verteilung ($\mu = 0,\ \sigma = 1$)

3.2.3 Parameter-Schätzung der Generalisierten Pareto-Verteilung

Im Folgenden sei $(X_{1,n}, \ldots, X_{n,n})$ mit $X_{1,n} \leq \cdots \leq X_{n,n}$ die sogenannte Ordnungsstatistik der Stichprobe $x_1, \ldots, x_n$, d.h. die geordnete Stichprobe.
Zur Ermittlung der Parameter einer Grenzverteilung stehen diverse grafische Mittel und Schätzmethoden zur Verfügung. Einige dieser Verfahren sollen hier kurz vorgestellt werden.

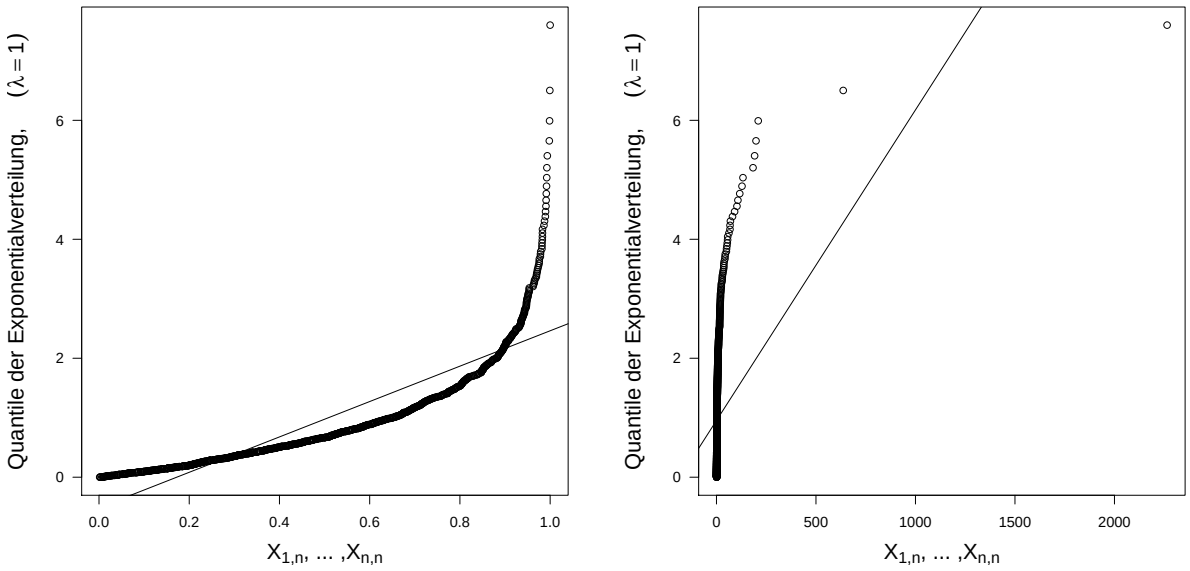
Der Formparameter $\gamma = \frac{1}{\alpha}$ und die Schranke u

Vor einer Parameterschätzung muss geklärt werden, welchem Typ von Extremwertverteilung die Stichprobe angehört. Darüber kann ein Q-Q-Plot Auskunft geben. In Abbildung 3.6(a) wird die Ordnungsstatistik einer simulierten Stichprobe von gleichverteilten Zufallsvariablen gegen die Quantile der Exponentialverteilung mit Parameter $\lambda = 1$ aufgetragen. Durch die konvexe Form der Q-Q-Line kann man erkennen, dass der Tail der Stichprobe leichter ist als jener der Exponentialverteilung. Der Q-Q-Plot in Abbildung 3.6(b) ist hingegen konkav, womit es sich hier um eine heavy-tailed Verteilung, in diesem Fall um eine $Par(1)$-Verteilung, handelt.

Für Stichproben, die aus einer heavy-tailed Verteilung stammen bzw. Pareto-ähnlich sind, kann ein weiterer Wahrscheinlichkeitsplot zu Bestimmung des Tail-Index α herangezogen werden. Für $F \in MDA(\Phi_\alpha)$ mit $\alpha > 0$ und ausreichend große x gilt (vgl. Satz 5):

$$\bar{F}(x) \approx x^{-\alpha} \quad \Longleftrightarrow \quad \log(1 - F(x)) \approx -\alpha \log(x) \tag{3.9}$$

Setzt man für $F(x_i)$ die empirische Verteilungsfunktion (ECDF) $F_n(x_i) = i/n$ ein, sollte



(a) $U(0,1)$ (b) $Par(1)$

Abbildung 3.6: Q-Q-Plots

der Log-Log-Plot

$$\left\{ \left(\log\left(1 - \frac{i}{n}\right), \log\left(X_{i,n}\right) \right),\ i = 1, \ldots, n \right\}$$

eine negative Steigung aufweisen, welche $-\alpha$ entspricht.

In Abbildung 3.7 sind die Log-Log-Plots der simulierten Stichproben einer $Par(1)$-Verteilung (schwarz) und einer Cauchy-Verteilung (grau) mit je 1000 Elementen dargestellt. Man kann deutlich erkennen, dass sich beide Verteilungen an die blaue Linie mit Steigung -1 annähern. Da für die Pareto-Verteilung die Gleichung 3.9 nicht nur approximativ gilt, weist diese sofort das erwünschte Verhalten auf, während es für die Cauchy-Verteilung erst mit größerem x eintritt. Es stellt sich also die Frage, ab welchem i bzw. welcher Schranke $u = X_{i,n}$ diese Gleichung erfüllt ist.

Die Annahme,dass $\bar{F} \in RV_{-\alpha}$ mit $\alpha > 0$, liegt auch dem populären Hill-Schätzer zugrunde. In seiner ursprünglichen Form gilt er somit nur für Extremwertverteilungen des Frechet-Typs. Bevor der Hill-Schätzer berechnet werden kann, muss allerdings erst der Satz von Karamata eingeführt werden:

Satz 8: Karamata

Sei L langsam variierend. Dann gilt für

a) $\kappa > -1$, dass

$$\lim_{x \to \infty} \frac{x^{\kappa+1} L(x)}{\int_0^x t^{\kappa} L(t) dt} = \kappa + 1.$$

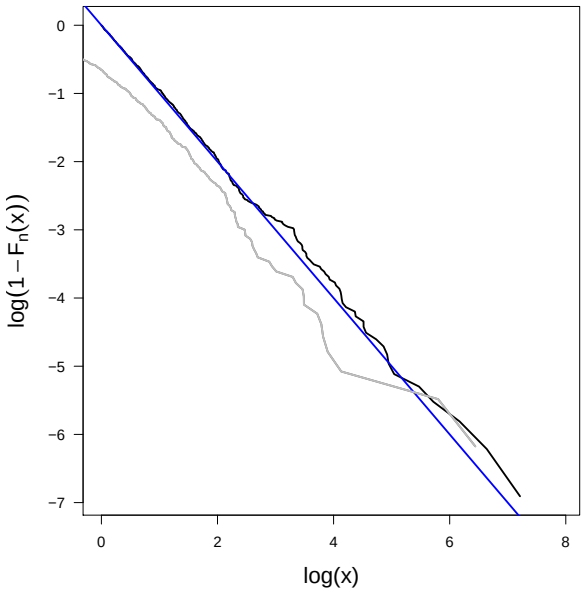


Abbildung 3.7: Log-Log-Plot der Pareto-Verteilung (schwarz) und der Cauchy-Verteilung (grau)

b) $\kappa < -1$, dass $\int_x^\infty t^\kappa L(t)dt < \infty$ und

$$\lim_{x\to\infty} \frac{x^{\kappa+1}L(x)}{\int_x^\infty t^\kappa L(t)dt} = -\kappa - 1.$$

Beweis. siehe [43], S. 26-28.

Nun kann die mittlere Überschreitungsfunktion der logarithmierten Zufallsvariablen X folgendermaßen vereinfacht werden:

$$e_{\log(X)}(\log(u)) = \mathbb{E}[\log(X) - \log(u) | \log(X) > \log(u)] = \frac{1}{\bar{F}(u)} \int_u^\infty \log(x) - \log(u) dF(x)$$

$$\overset{part.Int.}{=} \frac{1}{\bar{F}(u)} \int_u^\infty \frac{\bar{F}(x)}{x} dx \overset{\bar{F}\in RV_{-\alpha}}{\approx} \frac{1}{\bar{F}(u)} \int_u^\infty x^{-\alpha-1} L(x) dx$$

$$\overset{Satz\ 8b}{\approx} \frac{1}{\bar{F}(u)} \left(-\frac{1}{(-\alpha-1)+1} x^{-\alpha-1+1} L(x) \right) = \frac{1}{\alpha}.$$

Somit ergibt sich der Hill-Schätzer $\hat{\gamma}_H(k)$ der geordneten Stichprobe $X_{1,n}, \ldots, X_{n,n}$, als

$$\hat{\gamma}_H(k) = \frac{1}{\hat{\alpha}_H(k)} = \frac{1}{k} \sum_{j=1}^{k} \left[\log(X_{n-j+1} - \log(X_{n-k}) \right].$$

Dabei soll X_{n-k} die Schranke u ersetzen.
Eine Erweiterung des Hill-Schätzers auf $\alpha \in \mathbb{R}$ bieten Dekkers, Einmahl und de Hann (siehe [11]):

$$\hat{\gamma}_h^*(k) = \frac{1}{\hat{\alpha}_H^*(k)} = \hat{\gamma}_H(k) + 1 - \frac{1}{2}\left[1 - \frac{\left(\hat{\gamma}_H(k)\right)^2}{M_2(k)}\right]^{-1}$$
$$\text{mit} \quad M_2(k) = \frac{1}{k}\sum_{j=1}^{k}\left[\log(X_{n-j+1} - \log(X_{n-k})\right]^2.$$

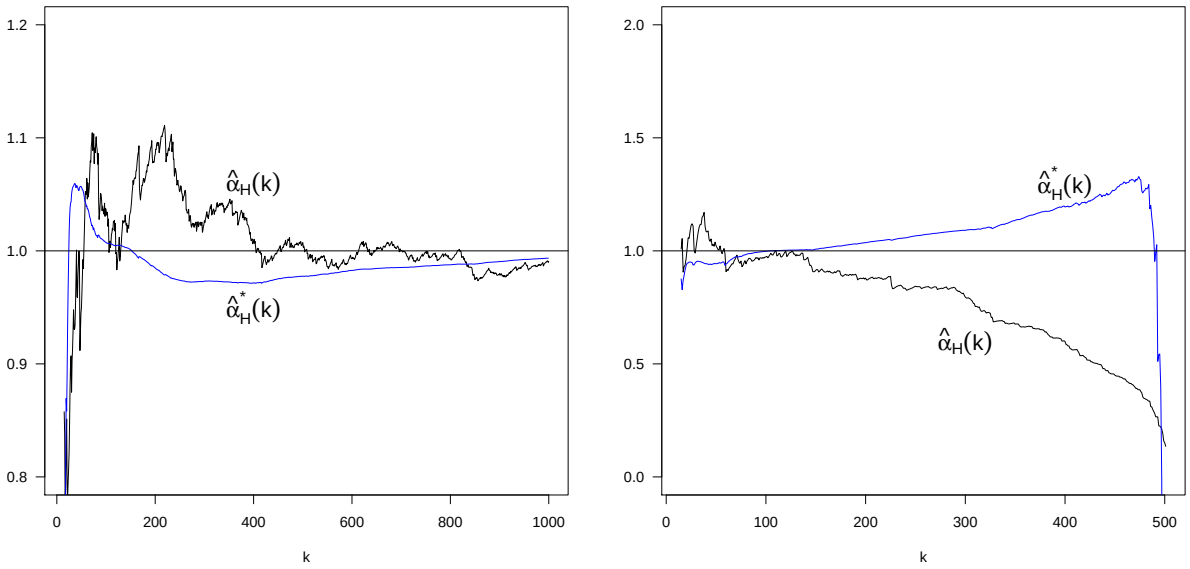
Offensichtlich hängen beiden Schätzer des Tail-Indexes α von der Wahl der Schranke u ab. Da unter bestimmten Voraussetzungen für $\hat{\gamma}_H$ als auch $\hat{\gamma}_H^*$ schwache bzw. starke Konsistenz gezeigt werden kann, sollten sich diese Schätzer ab einem Index k, sodass $X_{n-k,n} > u$, um den wahren Wert von $\gamma = 1/\alpha$ einpendeln. Dieses Verhalten wird meist grafisch im sogenannten Hill-Plot

$$\left\{\left(k, \frac{1}{\hat{\alpha}_H(k)}\right),\ k = 1, \ldots, n-1\right\}$$
$$\text{bzw.} \quad \left\{\left(k, \frac{1}{\hat{\alpha}_H^*(k)}\right),\ k = 1, \ldots, n-1\right\}$$

überprüft.
In Abbildung 3.8 sind die Hill-Plots für Simulationen von je 1000 unabhängigen Zufallsvariablen für die Pareto-, Cauchy-, Gleich- und Exponentialverteilung gegeben. Die wahren Werte für α sind als waagrechte Linien in der Höhe von 1, 1, -1 und 0 eingezeichnet. Während in Abbildung 3.8(a) sowohl $\hat{\alpha}_H(k)$ als auch $\hat{\alpha}_H^*(k)$ ein stabiles Verhalten um α zeigen, ist in den Abbildungen 3.8(c) und 3.8(d) zu erkennen, dass über den Hill-Schätzer $\hat{\alpha}_H(k)$ bzw. die richtige Schranke u keine Aussage getroffen werden kann. Der Schätzer $\hat{\alpha}_H^*(k)$ scheint gegen den richtigen Wert zu konvergieren. Abbildung 3.8(b) zeigt eine Cauchy-Verteilung, welche einen Tail-Index von 1 besitzt. Wie man im Log-Log-Plot in Abbildung 3.7 sehen kann, ist die Konvergenz allerdings nicht sehr schnell. In [15] wird eine ähnliche Verteilung abgebildet, die wie die Cauchy-Verteilung über einen positiven Tail-Index verfügt, deren Hill-Plot dennoch nicht aussagekräftig ist. Drees, de Haan und Resnick muten dort jenen, die den wahren Wert von α erkennen, „hellseherische“ Kräfte zu[1].

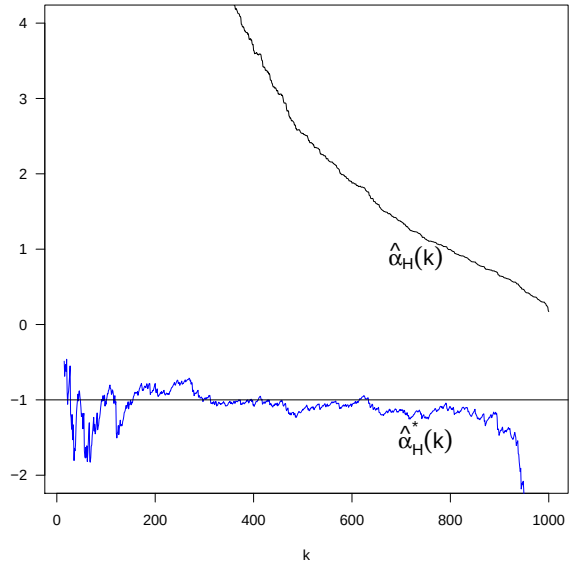
Eine weitere Möglichkeit die Schranke u zu bestimmen ist der Mean-Excess-Plot. Dabei gilt für die erwartete Überschreitungfunktion („mean excess function “) für ein ausreichend

[1] „One would have to be paranormal to discern with confidence the true value from the Hill plot.“([15], Seite 256, über den Hill-Plot einer symmetrischen stabilen Verteilung mit $\alpha = 0,2$)

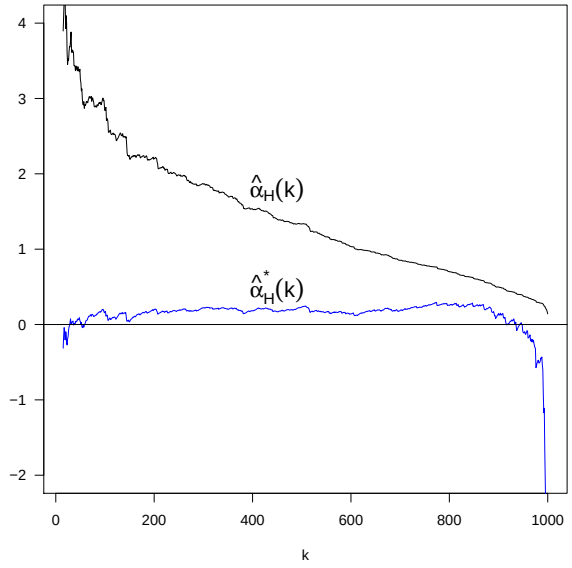


(a) $Par(1)$

(b) Cauchy



(c) $U(0,1)$



(d) $Exp(\lambda)$

Abbildung 3.8: Hillschätzer

großes u laut Satz 7, dass

$$\begin{aligned} e(u) =& \mathbb{E}[X-u|X>u] = \frac{\mathbb{E}\left[(X-u)\mathbb{1}_{\{X>u\}}\right]}{\mathbb{P}[X>u]} = \frac{1}{\bar{F}(u)}\int_u^{x_F}(x-u)dx \\ \overset{\text{part. Int.}}{=}& \frac{1}{\bar{F}(u)}\left[(x_F-u)-\int_u^{x_F}F(x)dx\right] = \int_u^{x_F}\frac{\bar{F}(x)}{\bar{F}(u)}dx \\ \overset{\text{Satz 7}}{=}& \left(1+\gamma\frac{u}{\beta}\right)^{1/\gamma}\int_u^{x_F}\left(1+\gamma\frac{x}{\beta}\right)^{-1/\gamma}dx = \frac{\beta}{1-\gamma}\left(1+\gamma\frac{x}{\beta}\right)\frac{\bar{G}_{\gamma,0,\beta}(x)}{\bar{G}_{\gamma,0,\beta}(u)}\Big|_{x=u}^{x_F} \\ =& \frac{\beta+\gamma u}{1-\gamma}. \end{aligned}$$

Diese Funktion sollte also linear werden, wenn $x > u$. Durch Verwendung des empirischen Schätzers der Mean-Excess-Funktion, $\hat{e}_n(u)$,

$$\hat{e}_n(u) = \frac{1}{N_u}\sum_{i:X_i>u}(x-u) \qquad \text{mit} \qquad N_u = |\{i : X_i > u\}|,$$

kann der Mean-Excess-Plot gezeichnet werden:

$$\{(X_{k,n}, \hat{e}_n(X_{k,n})),\ k = 1,\dots,n\}.$$

In Abbildung 3.9(a) kann dieses Verhalten anhand der $Par(2)$-Verteilung beobachtet werden. Die Steigung dieser Punkte liegt mit 1.15 schon nahe am wahren Wert von q. Allerdings sollte man den Plot eher als Anhaltspunkt für den Typ der Verteilung betrachten bzw. als Möglichkeit, die Schranke u zu bestimmen. Deshalb sind in den Abbildungen 3.9(b) die Mean-Excess-Plots von je 100 Simulationen einer $Par(2)$-verteilten Zufallsvariablen abgebildet. Die Steigungen der Ausgleichsgerade schwankt hier aber zwischen 0,26 und 1,88. Auch der Plot der Lognormal-Verteilung in Abbildung 3.9(c) könnte zur falschen Annahme einer Pareto-ähnlichen Verteilung führen. Die Mean-Excess-Funktion der Lognormalverteilung ist aber (siehe [17])

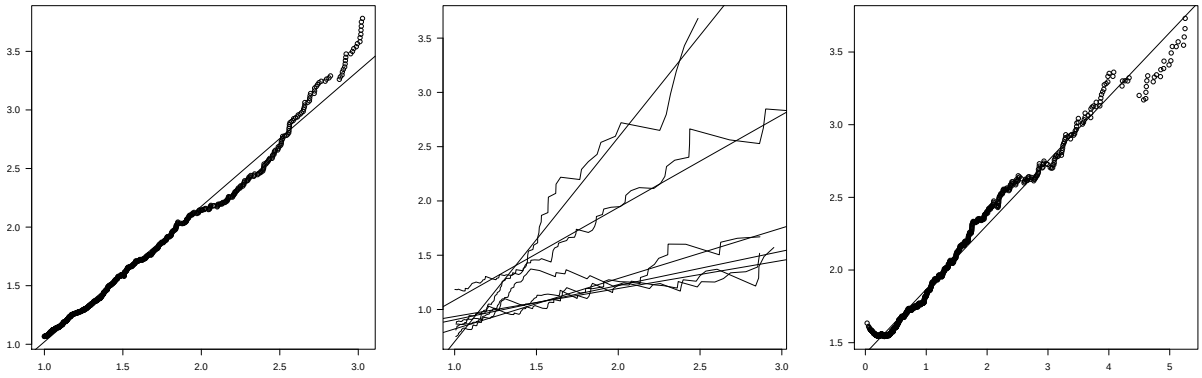
$$e_{LN}(u) = \frac{\sigma^2 u}{\log(u)-\mu}(1+o(1)) \qquad \text{für } u \to \infty.$$

Die Linearität dieser Funktion würde also fälschlich angenommen werden.

Skalenparameter $\sigma(u)$

Um den Parameter σ der Generalisierten Pareto-Verteilung zu ermitteln, wird meist die Maximum-Likelihood-Methode für $\gamma \neq 0$ verwendet. Dafür betrachtet man jene Überschüsse $Y_1 = X_{n,n} - u, \dots, Y_{N_u} = X_{n,n-N_u+1} - 1$, sodass $X_{n-k} > u,\ k = 0,\dots,N_u - 1$. Laut Satz 7 gilt

$$\bar{F}(y) \approx \bar{G}_{\gamma,0,\sigma(u)}.$$



(a) $Par(2)$ - 1000 Simulationen (b) $Par(2)$ - je 100 Simulationen (c) $LN(0,1)$

Abbildung 3.9: Mean-Excess-Plots

Durch diese Approximation gilt für die Dichte $f(y)$ bzw. die logarithmierte Maximum-Likelihood-Funktion $l(\sigma, \gamma | Y_1, \ldots, Y_{N_u})$:

$$f(y) \approx \frac{1}{\sigma}\left(1+\gamma\frac{y}{\sigma}\right)^{-\frac{1}{\gamma}-1}$$

$$l(\sigma,\gamma|Y_1,\ldots,Y_{N_u}) \approx -N_u\log(\sigma) + \left(-\frac{1}{\gamma}-1\right)\sum_{i=1}^{N_u}\log\left(1+\gamma\frac{Y_i}{\sigma}\right).$$

Die Maximierung der Log-Likelihood-Funktion muss allerdings numerisch erfolgen und liefert gute Ergebnisse für $\gamma > -1/2$.
Weiters wird zur Schätzung der Paramter γ und σ auch die Momentenmethode bzw. die Methode der wahrscheinlichkeitsgewichteten Momente verwendet (vgl. [17]). Ein möglicher Schätzer in Anlehnung an $\hat{\gamma}_H^*$ ist in [21] gegeben und lautet

$$\hat{\sigma}(k) = X_{n-k,n}\hat{\gamma}_H(k)\frac{1}{2}\left[\frac{(\hat{\gamma}_H(k))^{-2}}{M_2(k)}-1\right]^{-1}.$$

4 Modellierung der Abhängigkeiten: Copulas

In dieser Arbeit sollen vor allem Modelle für die räumliche Abhängigkeit der Jährlichkeiten gefunden werden. Während im ursprünglichen Distanzmatrixmodell nach Url Annahmen über die Verteilung der multiplikativen Abweichung getroffen werden, sind Copula-Funktionen ein allgemeiner Ansatz zur Modellierung von Abhängigkeiten. Zum besseren Verständnis wird in diesem Abschnitt eine kurze Einführung in die Theorie der Copulas gegeben. Eine detaillierte Beschreibung ist z.B. in [37] zu finden, wo auch die Definitionen und Sätze der nächsten Punkte nachzulesen sind.

4.1 Grundlegende Definitionen und Eigenschaften

Eine Copulafunktion wird vor allem dazu verwendet, nichtlineare Abhängigkeiten von Zufallsvariablen zu beschreiben. Beschränkt man die Analyse von multivariaten Zufallsvariablen auf lineare Abhängigkeitsmaße, wie z.B. den linearen Korrelationskoeffizienten, so können diverse Abhängigkeitsstrukturen unentdeckt bleiben.

Beispiel 4. In diesem Beispiel werden zwei Zufallsvariablen betrachtet, welche sich nur durch eine multiplikative Abweichung voneinander unterscheiden (vgl. Modul 1). Sei $X \sim LN(0,1)$ und $Y = XZ$ mit $Z \sim N(0,\sigma^2)$ unabhängig von X. Für den Erwartungswert und die Varianz von X und Y gilt:

$$\begin{aligned} \mathbb{E}[X] &= \sqrt{\mathrm{e}} & Var[X] &= \mathrm{e}(\mathrm{e}-1) \\ \mathbb{E}[Y] &= \mathbb{X}\mathbb{E}[Z] = 0 & Var[Y] &= \mathbb{E}[X^2]\mathbb{E}[Z^2] = \mathrm{e}^2\,\sigma^2. \end{aligned}$$

Der Korrelationskoeffizient der Zufallsvariablen X und Y ist

$$\rho_{X,Y} = \frac{\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]}{\sqrt{Var[X]}\sqrt{Var[Y]}} = \frac{\mathbb{E}[X^2Z]}{\sqrt{\mathrm{e}}\sqrt{\mathrm{e}-1}\,\mathrm{e}} = 0.$$

Somit besteht zwar kein linearer Zusammenhang, offensichtlich sind die beiden Zufallsvariablen aber nicht unabhängig.

Statt der Analyse der linearen Abhängigkeitsmaße kann auch ein Zugang über Copulas verwendet werden. Die formale Definition einer Copula beschreibt diese als die multivariate Verteilung von d gleichverteilten Zufallsvariablen.

Definition 4.1.1: Copula
Eine d-dimensionale Copula C ist eine Funktion $C : [0,1]^d \to [0,1]$, welche die folgenden Eigenschaften erfüllt:

1. $C(u_1, \ldots, u_d) = 0$, wenn mindestens ein $u_j,\ j = 1, \ldots, n$ gleich 0 ist,
2. $C(1, \ldots, 1, u_j, 1, \ldots, 1) = u_j \qquad \forall j = 1, \ldots, d,$
3. Für zwei Vektoren $u^i = (u_1^i, \ldots, u_n^i),\ i = 1, 2$, mit $u_j^1 \leq u_j^2 \quad \forall j$, gilt:

$$\sum_{i_1=1,2} \cdots \sum_{i_d=1,2} (-1)^{i_1+\cdots+i_d} C(u_1^{i_1}, \ldots, u_d^{i_d}) \geq 0.$$

Der Satz von Sklar zeigt, dass eine Copula in direkter Relation mit der gemeinsamen Verteilung von beliebigen Zufallsvariablen bzw. deren Randverteilungen steht. Bevor dieser Satz betrachtet wird, müssen zuerst zwei einfache Eigenschaften von Zufallsvariablen und deren Verteilungen beschrieben werden. Die erste grundlegende Eigenschaft ist, dass $F(X)$ für jede Zufallsvarible X mit Verteilungsfunktion F gleichverteilt ist. Eine ähnliche Eigenschaft gilt auch für die Inverse einer Verteilungsfunktion bzw. für deren Quasi-Inverse, wenn F nicht monoton steigend und stetig ist.

Definition 4.1.2: Quasi-Inverse
Sei $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion. Dann ist die Quasi-Inverse $F^{[-1]}$ der Verteilungsfunktion definiert als

$$F^{[-1]}(p) = \inf\{x | F(x) \leq p\}.$$

Bemerkung.
1. Die Quasi-Inverse einer Verteilungsfunktion wird auch oft als Quantilfunktion bezeichnet. Der Zusammenhang zwischen Quantilen und der Quasi-Inversen ergibt sich aus deren Definition.
2. Ist F streng monoton wachsend, dann gilt $F^{-1} = F^{[-1]}$

Mithilfe dieser Definition können nun die zuvor erwähnten Eigenschaften beschrieben werden.

Lemma 9:
Sei F eine Verteilungsfunktion. Dann gilt

1. Wenn F stetig ist und $X \sim F$, dann gilt $F(X) \sim U(0,1)$.
2. Wenn $U \sim U(0,1)$ und $F^{[-1]}$ die Quasi-Inverse von F, dann gilt $X = F^{[-1]}(U) \sim F$.

Beweis. Da diese beiden Transformationen für Copulas und auch für die Verteilung der Jährlichkeiten (siehe Abschnitt 5.2) sehr wichtig sind, soll auch der Beweis dieses Lemmas geführt werden:

1. Da F monoton steigend und stetig ist, ist $F^{[-1]}$ streng monoton steigend. Weiters gilt, dass

$$F(F^{[-1]}(u)) = F(\inf\{x \in \mathbb{R} | F(x) \geq u\}) = F(\inf\{x \in \mathbb{R} | F(x) = u\}) = u.$$

Somit gilt

$$\begin{aligned}\mathbb{P}[F(X) \leq u] &= \mathbb{P}[F^{[-1]}(F(X)) \leq F^{[-1]}(u)] \\ &\overset{f.s.}{=} \mathbb{P}[X \leq F^{[-1]}(u)] = F(F^{[-1]}(u)) = u\end{aligned}$$

und $F(X)$ folgt einer Gleichverteilung auf $[0, 1]$.

2. Da die Ungleichung $F^{[-1]}(u) \leq x$ äquivalent mit $F(x) \geq u$ ist, gilt

$$\mathbb{P}[X \leq x] = \mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[U \leq F(x)] \overset{U \sim U(0,1)}{=} F(x)$$

Satz 10: Sklar
Sei F eine multivariate Verteilung und $F_1 \dots F_d$ die dazugehörigen Randverteilungen. Dann existiert eine d-dimensionale Copula C, sodass

$$F(x_1, \dots, x_d) = C(F_1(x_1), \cdots, F_d(x_d)) \qquad \forall \mathbf{x} \in \mathbb{R}. \tag{4.1}$$

Sind die Randverteilungen $F_1 \dots F_d$ stetig, dann ist C eindeutig.
Auch die Umkehrung gilt, d.h. wenn C eine d-dimensionale Copula ist und $F_1 \dots F_d$ univariate Verteilungsfunktionen sind, dann ist F aus Gleichung 4.1 eine multivariate Verteilung mit Randverteilungen $F_1 \dots F_d$.

Beweis. siehe [37].

Eine der wichtigsten Folgerungen aus dem Satz von Sklar ist, dass eine Copula direkt aus der gemeinsamen Verteilung sowie den Randverteilungen der zugrundeliegenden Zufallsvariablen berechnet werden kann. Für stetige Randverteilungen gilt sofort, dass $C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))$. Sind die Randverteilungen nicht stetig, kann dieselbe Folgerung über die Quasi-Inversen der Randverteilungen getroffen werden.

Beispiel 5 (Gauss'sche Copula). Seien $F_i = \Phi,\ i = 1, \dots, d$ die Verteilungsfunktionen der Standardnormalverteilung und sei $F = \Phi_d^\Sigma$ die Verteilungsfunktion der d-dimensionalen Normalverteilung mit Erwartungswert 0 und Korrelationsmatrix Σ. Die daraus resultierende Copula C_Σ,

$$C_\Sigma(u_1, \dots, u_d) = \Phi_d^\Sigma(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)),$$

wird Gauss'sche Copula genannt.

Lemma 11: Invarianz einer Copula
Seien $X_1, \dots, X_d$ Zufallsvariablen mit stetigen Randverteilungen und gemeinsamer Copula C. Seien außerdem $\alpha_i : \mathbb{R} \to \mathbb{R},\ i = 1, \dots, d$ streng monoton steigende Funktionen. Dann ist die Copula C von $X_1, \dots, X_d$ auch eine Copula von $\alpha_1(X_1), \dots, \alpha_d(X_d)$.

Beweis. Für die Verteilungsfunktion F_i^α der transformierten Zufallsvariablen $\alpha_i(X_i)$ gilt

$$F_i^\alpha(x) = \mathbb{P}[\alpha_i(X_i) \leq x] = \mathbb{P}[X_i \leq \alpha_i^{-1}(x)] = F_i(\alpha_i^{-1}(x)).$$

Da die Funktionen α_i streng monoton steigend sind, ist α_i^{-1} und somit auch F_i^α stetig. Mit dem Satz von Sklar gibt es also eine eindeutig bestimmte Copula C^α für $\alpha_1(X_1), \ldots, \alpha_d(X_d)$.

$$\begin{aligned} C^\alpha(F_1^\alpha(x_1), \ldots, F_d^\alpha(x_d)) &= F^\alpha(\alpha_1(x_1), \ldots, \alpha_d(x_d)) \\ &= F(\alpha_1^{-1}(x_1), \ldots, \alpha_d^{-1}(x_d)) \\ &= C(F_1(\alpha_1^{-1}(x_1)), \ldots, F_d(\alpha_d^{-1}(x_d))) \\ &= C(F_1^\alpha(x_1), \ldots, F_d^\alpha(x_d)). \end{aligned}$$

4.2 Archimedische Copula

Eine breite Klasse der Copulas bilden die sogenannten Archimedischen Copulas. Ihre Vorteile liegen vor allem darin, dass sie durch eine Generator-Funktion erzeugt werden. Durch diese hilfreiche Darstellung können viele Copula-Eigenschaften vereinfacht werden.
Eine wichtige technische Voraussetzung an den Generator einer Archimedischen Copula ist die sogenannte d-Monotonie. In dieser Arbeit soll die Definition aus [35, S. 7] verwendet werden:

Definition 4.2.1: d-monoton, vollständig monoton
Eine Funktion $f : \mathbb{R} \to \mathbb{R}$ heißt d-monoton, $d \geq 2$, im Intervall (a,b) mit $a, b \in \bar{\mathbb{R}}$, wenn f $(d-2)$-mal differenzierbar ist, ihre k-te Ableitung, $k = 0, 1, \ldots, d-2$, folgende Eigenschaft erfüllt

$$(-1)^k f^{(k)}(x) \geq 0, \qquad \forall x \in (a, b), \tag{4.2}$$

und die $(d-2)$-te Ableitung fallend und konvex in (a, b) ist. Eine Funktion f heißt vollständig monoton, wenn alle Ableitungen existieren und (4.2) für $k \in \mathbb{N}_0$ gilt. $(d = \infty)$

Teilweise wird in der Literatur (vgl. [47] und [37]) sogar vollständige Monotonie des Generators vorausgesetzt. Dies hat den Grund, dass ein d-monotoner Generator eine d-dimensionale Copula-Funktion erzeugen kann. Für $k > d$ werden die definierenden Eigenschaften einer Copula verletzt. Wie man anhand der Definition der Archimedischen Copula erkennen kann, können diese relativ leicht auf höhere Dimensionen erweitert werden:

Definition 4.2.2: Archimedische Copula
Generator $\varphi : [0, 1]^d \to [0, \infty]$, stetig und streng monoton fallend mit $\varphi(1) = 0$. Wenn φ^{-1} vollständig monoton ist, definiert $C : [0, 1]^d \to [0, 1]$ mit

$$C(\vec{u}) = \varphi^{-1}\left(\varphi(u_1) + \cdots + \varphi(u_d)\right) \tag{4.3}$$

eine Copula für $d \geq 2$. Diese Copula heißt Archimedische Copula.

Eine ausführliche Aufzählung von Archimedischen Copulas C_θ mit einem einparametrigen Generator $\varphi_\theta(t)$ ist in [37, S. 116-122] zu finden. Die bekanntesten Archimedischen Copulas sind wohl die Clayton-, Gumbel- und Frank-Copula, welche in den Beispielen 6, 7 und 8 beschrieben werden. Der Vorteil dieser Copulas ist, dass die Inverse ihres Generators nicht nur konvex, sondern vollständig monoton ist. Dadurch sind sie nicht auf den bivariaten Fall beschränkt.

Beispiel 6 (Clayton-Copula). Seien der Generator φ_θ und dessen Inverse φ_θ^{-1} für $\theta > 0$ gegeben als

$$\varphi_\theta(t) = \frac{t^{-\theta} - 1}{\theta} \qquad \text{bzw.} \qquad \varphi_\theta^{-1}(s) = (\theta s + 1)^{-\frac{1}{\theta}}.$$

Die daraus resultierende Copula

$$C_\theta(\mathbf{u}) = \left(\theta \left(\sum_{i=1}^{d} \frac{u_i^{-\theta} - 1}{\theta} \right) + 1 \right)^{-\frac{1}{\theta}} = \left(\sum_{i=1}^{d} \left(u_i^{-\theta} \right) - d + 1 \right)^{-\frac{1}{\theta}}$$

heißt Clayton-Copula.

Beispiel 7 (Gumbel-Copula). Seien der Generator φ_θ und dessen Inverse φ_θ^{-1} für $\theta \in [1, \infty)$ gegeben als

$$\varphi_\theta(t) = (-\log(t))^\theta \qquad \text{und} \qquad \varphi_\theta^{-1}(s) = \exp\left\{ -s^{\frac{1}{\theta}} \right\}.$$

Die Copula dieses Generators heißt Gumbel-Copula und ist gegeben als

$$C_\theta(t) = \exp\left\{ -\left(\sum_{i=1}^{d} (-\log(t))^\theta \right)^{\frac{1}{\theta}} \right\}.$$

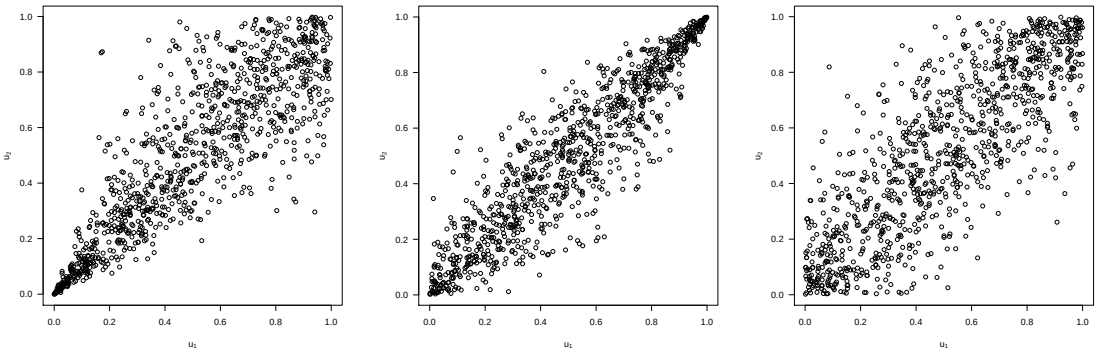
Beispiel 8 (Frank-Copula). Der Generator φ_θ und dessen Inverse φ_θ^{-1} für $\theta > 0$ der Frank-Copula sind

$$\varphi_\theta(t) = -\log\left(\frac{\mathrm{e}^{-\theta t} - 1}{\mathrm{e}^{-\theta} - 1} \right) \qquad \text{und} \qquad \varphi_\theta^{-1}(s) = -\frac{1}{\theta} \log\left(1 + \mathrm{e}^{-s} \left(\mathrm{e}^{-\theta} - 1 \right) \right).$$

Damit erhält man die Copula

$$\begin{aligned} C_\theta(t) &= -\frac{1}{\theta} \log\left(1 + \mathrm{e}^{-\sum_{i=1}^{d} -\log\left(\frac{\mathrm{e}^{-\theta u_i} - 1}{\mathrm{e}^{-\theta} - 1} \right)} \left(\mathrm{e}^{-\theta} - 1 \right) \right) \\ &= -\frac{1}{\theta} \log\left(1 + \left(\mathrm{e}^{-\theta} - 1 \right)^{1-d} \prod_{i=1}^{d} \left(\mathrm{e}^{-\theta u_i} - 1 \right) \right). \end{aligned}$$

Wie man in den Abbildungen 4.1 sehen kann, sammeln sich die Werte der u_1 und u_2 für die verschiedenen Copulas um $(0, 0)$ bzw. $(1, 1)$. Dieser Effekt wird als Tail-Dependence bezeichnet.



(a) Clayton-Copula ($\theta = 4$) (b) Gumbel-Copula ($\theta = 4$) (c) Frank-Copula ($\theta = 8$)

Abbildung 4.1: Scatterplots von gleichverteilten Zufallsvariablen mit verschiedenen Copula-Funktionen

4.3 Tail-Abhängigkeit

Die Koeffizienten der Tailabhängigkeit, λ_U und λ_L, geben Auskunft über die Abhängigkeit der Extrema zweier Zufallsvariablen mit gemeinsamer Copula C. Sie beschreiben die Wahrscheinlichkeit, dass eine Zufallsvariable sehr groß oder klein wird, wenn die andere bereits sehr groß oder klein ist.

Definition 4.3.1: Obere/untere Tail-Abhängigkeit
Seien X und Y Zufallsvariablen mit Randverteilungen F_X und F_Y. Dann ist der Koeffizient der oberen Tail-Abhängigkeit (X, Y), λ_U, gegeben als

$$\lambda_U = \lim_{t \to 1^-} \mathbb{P}[Y > F_Y^{[-1]}(t) | X > F_X^{[-1]}(t)],$$

und der Koeffizient der unteren Tailabhängigkeit, λ_L, ist

$$\lambda_L = \lim_{t \to 0^+} \mathbb{P}[Y \leq F_Y^{[-1]}(t) | X \leq F_X^{[-1]}(t)].$$

Satz 12:
Seien X und Y zwei Zufallsvariablen mit Randverteilungen F_X und F_Y und gemeinsamer Copula C. Dann gilt für die Koeffizienten der Tail-Abhängigkeit:

$$\lambda_U = 2 - \lim_{t \to 1^-} \frac{1 - C(t,t)}{1 - t}$$
$$\lambda_L = \lim_{t \to 0^+} \frac{C(t,t)}{t}.$$

Beweis. Da $U_1 = F_X(X) \sim U(0,1)$ und $U_2 = F_Y(Y) \sim U(0,1)$ mit gemeinsamer Vertei-

lungsfunktion C, gilt

$$\begin{aligned}\lambda_U &= \lim_{t\to 1^-} \mathbb{P}[Y > F_Y^{[-1]}(t) | X > F_X^{[-1]}(t)] = \lim_{t\to 1^-} \mathbb{P}[F_Y(Y) > t | F_X(X) > t] \\ &= \lim_{t\to 1^-} \frac{\mathbb{P}[U_1 > t, U_2 > t]}{\mathbb{P}[U_1 > t]} = \lim_{t\to 1^-} \mathbb{P}\frac{1 - \mathbb{P}[U_1 \le t] - \mathbb{P}[u_2 \le t] + \mathbb{P}[U_1 \le t, U_2 \le t]}{1 - \mathbb{P}[U_1 \le t]} \\ &= \lim_{t\to 1^-} \frac{1 - t - t - C(t,t)}{1-t} = 2 - \lim_{t\to 1^-} \frac{1 - C(t,t)}{1-t}\end{aligned}$$

und

$$\begin{aligned}\lambda_L &= \lim_{t\to 0^+} \mathbb{P}[Y \le F_Y^{[-1]}(t) | X \le F_X^{[-1]}(t)] = \lim_{t\to 0^+} \mathbb{P}[F_Y(Y) \le t | F_X(X) \le t] \\ &= \lim_{t\to 0^+} \frac{\mathbb{P}[U_1 \le t, U_2 \le t]}{\mathbb{P}[U_1 \le t]} = \lim_{t\to 0^+} \frac{C(t,t)}{t}.\end{aligned}$$

Beispiel 9 (Clayton-Copula). Die bivariate Clayton-Copula $C_\theta(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 2 + 1\right)$ verfügt über die Tail-Koeffizienten

$$\begin{aligned}\lambda_U &= 2 - \lim_{t\to 1^-} \frac{1 - \left(2t^{-\theta} - 1\right)^{-1/\theta}}{1-\theta} = 2 - \lim_{t\to 1^-} \frac{2t^{-\theta-1}\left(2t^{-\theta} - 1\right)^{-1/\theta-1}}{-1} = 0 \\ \lambda_L &= \lim_{t\to 0^+} \frac{\left(2t^{-\theta} - 1\right)^{-1/\theta}}{t} = \lim_{t\to 0^+} \left(2 - t^\theta\right)^{-1/\theta} = 2^{-1/\theta}\end{aligned}$$

Beispiel 10 (Gumbel-Copula). Die Tailkoeffizienten der bivariate Gumbel-Copula,

$$C_\theta = \exp\left\{-\left[\left(-\log(u_1)\right)^\theta + \left(-\log(u_2)\right)^\theta\right]^{-1/\theta}\right\} = t^{2^{1/\theta}},$$

sind

$$\begin{aligned}\lambda_u &= 2 - \lim_{t\to 1^-} \frac{1 - t^{2^{1/\theta}}}{1-t} = 2 - \lim_{t\to 1^-} \frac{-2^1/\theta t^{2^{1/\theta-1}} - 1}{-1} = 2 - 2^{1/\theta} \\ \lambda_L &= \lim_{t\to 0^+} \frac{t^{2^{1/\theta}}}{t} = 0\end{aligned}$$

4.4 Kendalls τ

In Abschnitt 4.1 wurde ein mögliches Maß der Abhängigkeit, der lineare Korrelationskoeffizient, beschrieben. Wie anhand des Beispieles 4 gezeigt wurde, könnte dieses Maß zu irreführenden Resultaten führen. Abgesehen davon ist es möglich, dass der lineare Korrelationskoeffizient nicht für alle Verteilungen existiert, da im Fall unbeschränkter zweiter Momente durch unendlich dividiert wird. Ein weiteres verbreitetes Maß der Abhängigkeit wird durch die Betrachtung der sogenannten „Konkordanzen“ bzw. „Diskonkordanzen“ zweier Zufallsvektoren gefunden.

Definition 4.4.1: (dis-)konkordant
Seien X und Y zwei Zufallsvektoren mit stetigen Randverteilungen. Die Beobachtungen (x_i, y_i) und (x_j, y_j) sind konkordant, wenn $x_i > x_j$, $y_i > y_j$ oder $x_i < x_j$, $y_i < y_j$. Die Beobachtungen, für die $x_i > x_j$, $y_i < y_j$ oder $x_i < x_j$, $y_i > y_j$ gilt, heißen diskonkordant.

Bemerkung. Die Bedingungen aus Definition 4.4.1 für konkordanten Beobachtungen kann zu

$$(x_i - x_j)(y_i - y_j) > 0, \tag{4.4}$$

vereinfacht werden. Im Fall von diskonkordanten Beobachtungen ändert sich das Vorzeichen in 4.4.

Für die Differenz der Wahrscheinlichkeiten der Konkordanzen und Diskonkordanzen ergibt sich eine Integraldarstellung durch Copulas, welche eine einfache Möglichkeit zur Bestimmung der Parameter Archimedischer Copulas liefert.

Satz 13:
Seien (X, Y) und $(\tilde{X}, \tilde{Y})$ zwei Zufallsvektoren mit stetigen Randverteilungen F_X bzw. F_Y und gemeinsamer Verteilung H bzw. $\tilde{H}$. Seien C und $\tilde{C}$ die Copulas von (X, Y) bzw. $(\tilde{X}, \tilde{Y})$ und Q definiert als

$$Q = \mathbb{P}[(X - \tilde{X})(Y - \tilde{Y}) > 0] - \mathbb{P}[(X - \tilde{X})(Y - \tilde{Y}) < 0]. \tag{4.5}$$

Dann gilt

$$Q(C, \tilde{C}) = 4 \int_0^1 \int_0^1 \tilde{C}(u, v) dC(u, v) - 1.$$

Beweis. Im ersten Schritt wird die Wahrscheinlichkeit Q so umgeformt, dass Q nur noch von der Wahrscheinlichkeit der Konkordanzen abhängt:

$$\begin{aligned} Q &= \mathbb{P}[(X - \tilde{X})(Y - \tilde{Y}) > 0] - \mathbb{P}[(X - \tilde{X})(Y - \tilde{Y}) < 0] \\ &= \mathbb{P}[(X - \tilde{X})(Y - \tilde{Y}) > 0] - (1 - \mathbb{P}[(X - \tilde{X})(Y - \tilde{Y}) > 0]) \\ &= 2\mathbb{P}[(X - \tilde{X})(Y - \tilde{Y}) > 0] - 1 \\ &= 2(\mathbb{P}[X > \tilde{X}, Y > \tilde{Y}] + \mathbb{P}[X < \tilde{X}, Y < \tilde{Y}]) - 1 \end{aligned}$$

Für die Wahrscheinlichkeit $\mathbb{P}[X > \tilde{X}, Y > \tilde{Y}]$ gilt:

$$\begin{aligned} \mathbb{P}[X > \tilde{X}, Y > \tilde{Y}] &= \iint_{\mathbb{R}^2} \tilde{H}(x, y) dH(x, y) \\ &= \iint_{\mathbb{R}^2} \tilde{C}(F_X(x), F_Y(y)) dC(F_X(x), F_Y(y)) \end{aligned}$$

Mit der Transformation $u = F_X(x),\ v = F_Y(y)$ kann das obige Ergebnis in die gewünschte Form gebracht werden:

$$\mathbb{P}[X > \tilde{X}, Y > \tilde{Y}] = \int_0^1 \int_0^1 \tilde{C}(u, v) dC(u, v).$$

Damit gilt

$$Q = 2(\mathbb{P}[X > \tilde{X}, Y > \tilde{Y}] + \mathbb{P}[X < \tilde{X}, Y < \tilde{Y}]) - 1 = 4 \int_0^1 \int_0^1 \tilde{C}(u,v) dC(u,v) - 1.$$

Um auch die Wahrscheinlichkeit $\mathbb{P}[X < \tilde{X}, Y < \tilde{Y}]$ in einer ähnlichen Integralform darzustellen, wird sie zuerst vereinfacht:

$$\begin{aligned}\mathbb{P}[\tilde{X} > x, \tilde{Y} > y] &= 1 - \mathbb{P}[\tilde{X} < x] - \mathbb{P}[\tilde{Y} < y] + \mathbb{P}[\tilde{X} < x, \tilde{Y} < y], \\ &= 1 - F_X(x) - F_Y(y) + \tilde{H}(x,y),\end{aligned}$$

und kann dann als Integral nach einer Copula gbildet werden.

$$\begin{aligned}\mathbb{P}[\tilde{X} > X, \tilde{Y} > Y] &= \iint_{\mathbb{R}^2} (1 - F_X(x) - F_Y(y) + \tilde{H}(x,y)) dH(x,y) \\ &= \int_0^1 \int_0^1 (1 - u - v + \tilde{C}(u,v)) dC(u,v)\end{aligned}$$

Da F_X bzw. F_Y sowohl die Randverteilung von (X,Y) als auch von $(\tilde{X}, \tilde{Y})$ ist, ist C auch hier die bivariate Verteilungsfunktion von u und v. Somit gilt

$$\begin{aligned}\mathbb{P}[\tilde{X} > X, \tilde{Y} > Y] &= \underbrace{\int_0^1 \int_0^1 1 dC(u,v)}_{1} - \underbrace{\int_0^1 \int_0^1 u dC(u,v)}_{\mathbb{E}[u]=\frac{1}{2}} \\ &\quad - \underbrace{\int_0^1 \int_0^1 v dC(u,v)}_{\mathbb{E}[v]=\frac{1}{2}} + \int_0^1 \int_0^1 \tilde{C}(u,v)) dC(u,v) \\ &= \int_0^1 \int_0^1 \tilde{C}(u,v)) dC(u,v).\end{aligned}$$

Sind (X,Y) und $(\tilde{X}, \tilde{Y})$ identisch verteilt, wird die Wahrscheinlichkeit Q aus Satz 13 als Kendalls τ bezeichnet.

Definition 4.4.2: Kendalls τ
Seien (X_1, Y_1) und (X_2, Y_2) unabhängig und identisch verteilte Zufallsvektoren mit stetigen Randverteilungen. Dann bezeichnet man die Differenz zwischen Konkordanzen und Diskonkordanzen,

$$\tau = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0],$$

als das Kendalls τ von X und Y.

Eine direkte Folgerung aus Satz 13 der Zusammenhang zwischen Kendalls Tau und der gemeinsamen Copula:

Satz 14:
Seien X und Y Zufallsvariablen mit stetigen Randverteilungen (d.h. ihre Copula ist direkt aus dem Satz von Sklar gegeben). Dann ist Kendalls τ gegeben als

$$\tau = Q(C, \tilde{C}) = 4 \int_0^1 C(u,v) dC(u,v) - 1. \tag{4.6}$$

Für Archimedische Copulas kann die Gleichung 4.6 weiter vereinfacht werden.

Lemma 15:
Seien X und Y Zufallsvariablen mit Archimedischer Copula C und Generator φ. Dann ist Kendalls τ gegeben als

$$\tau = 1 + 4 \int_0^1 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt \tag{4.7}$$

Beweis. Siehe [37, S. 163]

Beispiel 11 (Clayton-Copula). Setzt man den Generator der Clayton-Copula mit Parameter $\theta > 0$, $\varphi(t) = (t^{-\theta} - 1)/\theta$, und dessen Ableitung $\varphi'(t) = -t^{-(\theta+1)}$ in Gleichung 4.7 ein, erhält man folgende Relation zwischen θ und τ:

$$\begin{aligned}\tau &= 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt = 1 + 4 \int_0^1 \frac{t(t^\theta - 1)}{\theta} dt = 1 + \frac{4}{\theta} \left(\frac{t^{\theta+1}}{\theta+1} - \frac{t^2}{2} \right) \Bigg|_0^1 \\ &= 1 + \frac{4}{\theta} \left(\frac{1}{\theta+2} - \frac{1}{2} \right) = 1 - \frac{2}{\theta+2} = \frac{\theta}{\theta+2}\end{aligned}$$

Beispiel 12 (Gumbel-Copula). Auch für die Gumbel-Copula mit Generator $\varphi(t) = (-\log(t))^\theta$ und seiner Ableitung $\varphi'(t) = -\theta/t(-\log(t))^{\theta-1}$ kann man Kendalls τ in Abhängigkeit von θ ausdrücken:

$$\begin{aligned}\tau &= 1 + 4 \int_0^1 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt = 1 + \frac{4}{\theta} \int_0^1 t \log(t) dt = 1 + \frac{4}{\theta} \left(-\frac{t^2}{4} + \frac{t^2}{2} \log(t) \right) \Bigg|_0^1 \\ &= 1 + \frac{4}{\theta} \frac{1}{4} = 1 - \frac{1}{\theta}\end{aligned}$$

Bemerkung. Der Schätzer von Kendalls τ für Stichprobe $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ ist folgendermaßen gegeben:

$$\begin{aligned}\hat{\tau} =& \frac{\#\{\text{übereinstimmende Paare}\} - \#\{\text{nicht übereinstimmende Paare}\}}{\#\{\text{alle Paare}\}} \\ =& \frac{\sum_{i=1}^{n} \sum_{j=1}^{i-1} \operatorname{sgn}((x_i - x_j)(y_i - y_j))}{\frac{n(n-1)}{2}}.\end{aligned} \tag{4.8}$$

Durch den Satz 14 kann somit anhand des geschätzten Kendalls τ, $\hat{\tau}$, ein Schätzer für den Parameter Archimedischer Copulas, $\tilde{\theta}$, gefunden werden.

Beispiel 13 (Schätzer des Parameters der Clayton-Copula und der Gumbel-Copula). Die Gleichungen aus Beispiel 11 und 12 können nach θ aufgelöst werden. Setzt man den Schätzer des Kendalls τ aus 4.8 ein, erhält man für die Parameterschätzer $\hat{\theta}_{Cl}$ der Clayton Copula bzw. $\hat{\theta}_{Gu}$ der Gumbel-Copula

$$\hat{\theta}_{Cl} = \frac{2\hat{\tau}}{1-\hat{\tau}} \qquad \text{und} \qquad \hat{\theta}_{Gu} = \frac{1}{1-\hat{\tau}}$$

Ein anderes Konkordanzmaß in der Literatur ist Spearman's Rho, siehe [37, S. 167-170].

4.5 Der hierarchische Simulationsalgorithmus für Copulas

Die naheliegendste Simulationsmethode zur Erzeugung von gleichverteilten Zufallsvariablen $u_1, \ldots, u_d$ ist die hierarchische Simulation. Dabei wird durch Verwendung der bedingten Copula-Funktion rekursiv aus den bisher simulierten $u_1, \ldots, u_k$ die neue Zufallsvariable u_{k+1} gewonnen. Die bedingte Copula ist dabei folgendermaßen gegeben:

$$\begin{aligned} C(u_{k+1}|u_1,\ldots,u_k) &= \mathbb{P}[U_{k+1} \le u_{k+1}, U_{k+2} = 1, \ldots, U_d = 1 \mid U_1 = u_1, \ldots, U_k = u_k] \\ &= \mathbb{P}[U_{k+1} \le u_{k+1} \mid U_1 = u_1, \ldots, U_k = u_k] \\ &\overset{f.s.}{=} \lim_{h_1 \to 0, \ldots, h_k \to 0} \frac{\mathbb{P}[U_{k+1} \le u_{k+1}, u_1 \le U_1 \le u_1 + h_1, \ldots, u_k \le U_k \le u_k + h_k]}{\mathbb{P}[u_1 \le U_1 \le u_1 + h_1, \ldots, u_k \le U_k \le u_k + h_k]} \\ &= \frac{\frac{\partial^k C}{\partial u_1 \ldots \partial u_k}(u_1, \cdots, u_k, u_{k+1}, u_{k+2} = \cdots = u_d = 1)}{\frac{\partial^k C}{\partial u_1 \ldots \partial u_k}(u_1, \cdots, u_k, u_{k+1} = u_{k+2} = \cdots = u_d = 1)} \end{aligned} \tag{4.9}$$

Durch Verwendung von Lemma 9 für bedingte Verteilungsfunktionen kann nun ein Algorithmus zur Erzeugung von Zufallsvariablen mit gemeinsamer Copula-Funktion C beschrieben werden (siehe [20] oder [47]):

Hierarchischer Simulationsalgorithmus

1. Erzeuge $u_1 \sim U(0,1)$
2. Für $k = 2, \ldots, d-1$ erzeuge $v \sim U(0,1)$ und löse

$$v = \frac{\frac{\partial^k C}{\partial u_1 \ldots \partial u_k}(u_1, \cdots, u_k, u_{k+1}, u_{k+2} = \cdots = u_d = 1)}{\frac{\partial^k C}{\partial u_1 \ldots \partial u_k}(u_1, \cdots, u_k, u_{k+1} = u_{k+2} = \cdots = u_d = 1)}. \tag{4.10}$$

   nach u_{k+1}.
3. Das Ergebnis $(u_1, \ldots, u_d)$ hat die gemeinsame Verteilungsfunktion C.

Auch die bedingte Copula kann für die Archimedische Copula als eine Funktion des Generators dargestellt werden, wie das folgende Lemma zeigt.

Lemma 16: Bedingte Archimedische Copula
Sei C eine Archimedische Copula mit Generator φ. Dann gilt für die bedingte Copula

$$C(u_{k+1}|u_1, \ldots, u_k) = \frac{\frac{\partial \varphi^{-1}}{\partial u_1 \ldots \partial u_k}\left(\varphi(u_1) + \cdots + \varphi(u_{k+1})\right)}{\frac{\partial \varphi^{-1}}{\partial u_1 \ldots \partial u_k}\left(\varphi(u_1) + \cdots + \varphi(u_k)\right)}.$$

Beweis. Wenn C eine Archimedische Copula ist, gilt für den Zähler aus (4.9):

$$\begin{aligned}\frac{\partial^k C}{\partial u_1 \cdots \partial u_k}(u_1, \ldots, u_{k+1}, u_{k+2} = \cdots = u_d = 1) = \\ \frac{\partial^k \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_{k+1}) + \varphi(1) + \cdots + \varphi(1))}{\partial u_1 \cdots \partial u_k} \\ \overset{\varphi(1)=0}{=} \frac{\partial^k \varphi^{-1}}{\partial u_1 \ldots \partial u_k}(\varphi(u_1) + \cdots \varphi(u_{k+1})) \prod_{i=1}^{k} \varphi'(u_i).\end{aligned}$$

Analog kann der Nenner aus (4.9) umgeformt werden:

$$\frac{\partial^k C}{\partial u_1 \cdots \partial u_k}(u_1, \ldots, u_k, u_{k+1} = \cdots = u_d = 1) = \frac{\partial^k \varphi^{-1}}{\partial u_1 \ldots \partial u_k}(\varphi(u_1) + \cdots \varphi(u_k)) \prod_{i=1}^{k} \varphi'(u_i).$$

Damit ist die bedingte Archimedische Copula

$$\begin{aligned} C(u_{k+1}|u_1,\dots,u_k) &= \frac{\frac{\partial^k C}{\partial u_1 \cdots \partial u_k}(u_1,\dots,u_{k+1},u_{k+2}=\cdots=u_d=1)}{\frac{\partial^k C}{\partial u_1 \cdots \partial u_k}(u_1,\dots,u_k,u_{k+1}=\cdots=u_d=1)} \\ &= \frac{\frac{\partial^k \varphi^{-1}}{\partial u_1 \dots \partial u_k}(\varphi(u_1)+\cdots\varphi(u_{k+1}))\prod_{i=1}^{k}\varphi'(u_i)}{\frac{\partial^k \varphi^{-1}}{\partial u_1 \dots \partial u_k}(\varphi(u_1)+\cdots\varphi(u_k))\prod_{i=1}^{k}\varphi'(u_i)} \\ &= \frac{\frac{\partial \varphi^{-1}}{\partial u_1 \dots \partial u_k}(\varphi(u_1)+\cdots+\varphi(u_{k+1}))}{\frac{\partial \varphi^{-1}}{\partial u_1 \dots \partial u_k}(\varphi(u_1)+\cdots+\varphi(u_k))}. \end{aligned}$$

5 Sturmschadenmodelle

In den folgenen Abschnitten sollen die Überlegungen der letzten Kapitel verwendet werden, um die Modelle der Hochwasserstudie ([39]) auf Sturmschäden anzupassen.

5.1 Trend und Saisonalität der Daten

In Abschnitt 2.1 wurde beschrieben, wie die Schäden an den einzelnen Sturmtagen zu Gesamtschäden von Sturmereignissen zusammengefasst wurden, um unabhängige Ereignisse zu schaffen und Fehler durch zeitlich verzögerte Meldungen der Schäden zu vermeiden. Außerdem wurde auch eine saisonale Komponente in den Daten festgestellt, welche aber aufgrund der geringen Anzahl von 36 Datenpunkten nicht berücksichtigt werden kann. Außerdem wurde ein log-lineares Poisson-Modell auf die Anzahl der Sturmereignisse angepasst. Allerdings verbesserte die Hinzunahme eines zeitlichen Trends dieses Modell kaum und aufgrund der Struktur der Abweichungen der Residuen von einer Geraden kann nicht mit Sicherheit von einer Zunahme extremer Sturmereignisse ausgegangen werden.

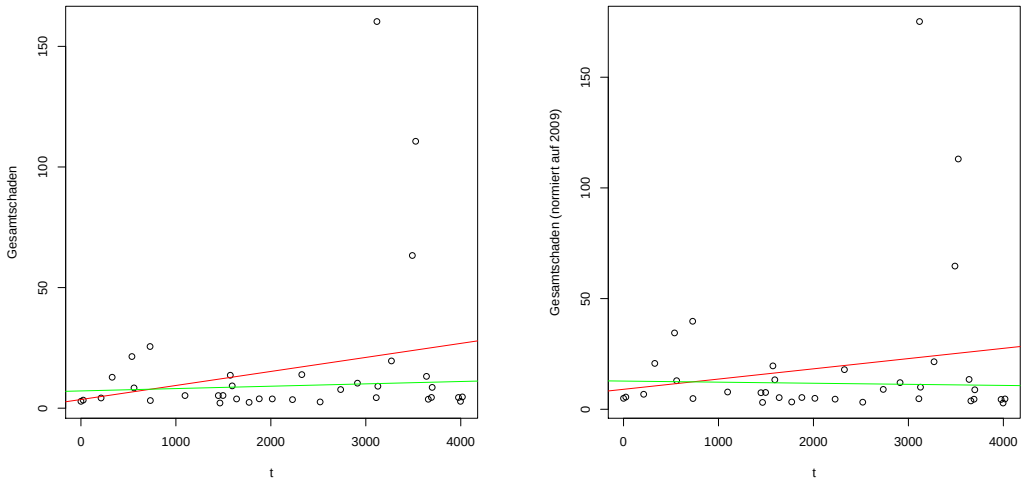
Nun soll betrachtet werden, ob die Sturmereignisse im Laufe der Zeit heftiger geworden sind und somit zu einem Trend in den Schadenshöhen geführt haben. In Abschnitt 1.1 wurden bereits einige Studien erwähnt, die sich mit der zukünfigen Entwickung von extremen Sturmereignissen beschäftigen. Deren Ergebnisse werden von Schönwiese folgendermaßen zusammengefasst:

> „Jedoch sind gerade bei Extremereignissen die Klimamodellrechnungen besonders unsicher und eine systematische Zunahme von Gewittern, Hagel und Tornados bisher nicht signifikant nachweisbar. Die Zugbahnen der Herbst- und Winterstürme gemäßigter Breiten könnten sich sogar polwärts verlagern, was z.B. in Mitteleuropa zu einer Abnahme, in Nordwest- und Nordeuropa aber zu einer Zunahme führen könnte.“([49], S. 363)

Die Expertenmeinungen über die Auswirkungen von extremen Sturmereignissen sind eher pessimistisch. So haben z.B. Pinto et al. in [38] durch die Betrachtung von verschiedenen Klimaänderungs-Szenarien des Max-Planck-Institutes (MPI) festgestellt, dass in West- und Zentraleuropa mit einem Anstieg der durch Stürme verursachten Schäden zu rechnen ist, selbst wenn eine Anpassung an die erhöhten Gefahren erfolgt. Auch die Rückversicherungsgesellschaft SwissRe geht von einer Zunahme von $16 - 68\%$ der Schadenshöhen über die Periode von 1975 bis 2085, welche vor allem auf ein häufigeres Auftreten sehr schwerer

Stürme zurückzuführen ist, aus ([51]).

In Abbildung 5.1(a) kann man erkennen, dass in den letzten Jahren vermehrt Sturmereignisse mit extrem hohem Schaden aufgetreten sind. Wird nun anhand dieser Daten eine einfache lineare Trendfunktion durch die Methode der kleinsten Quadrate verwendet, erhält man einen Anstieg von ca. 23 Mio. Euro über die betrachtete Periode. Die Steigung dieser Geraden wird aber vor allem durch die große Höhe der Stürme Kyrill, Paula und Emma beeinflusst. Nimmt man zumindest die beiden größeren Stürme aus der Trendschätzung aus, verschiebt dieser sich von der roten Geraden zur grünen. Über den gesamten Zeitraum ist nun nur mehr eine Zunahme von 4 Mio. Euro auf diesen Trend zurückzuführen.
Bevor sich ein Trend in den Schadensdaten aber auf eine Zunahme der Intensität von Stur-



(a) Gesamtschaden

(b) normierter Gesamtschaden

Abbildung 5.1: Gesamtschaden mit linearen Trendfunktionen

mereignissen zurückspielen lässt, müssen die Daten erst von wirtschaftlichen Zuwächsen bereinigt werden. Eine Inflations-Bereinigung wurde schon im Vorfeld vom VVÖ durchgeführt. Andererseits ist aber auch die versicherte Gesamtsumme im betrachteten Zeitraum in allen PLZ-2-Stellern angestiegen, was zu einem positiven Trend der Gesamtschäden führen kann. Um diesen Effekt auszuschalten und die einzelnen Datenpunkte vergleichbar zu machen, werden die Schadensgrade der einzelnen PLZ-2-Steller betrachtet. Diese beschreiben den Promille-Anteil des aufgetretenen Schadens an der versicherten Summe im jeweiligen Jahr in jedem PLZ-2-Steller. Somit können durch Multiplikation der versicherten Summe aus dem Jahr 2009 die Schäden der einzelnen Jahre von der Zunahme im Versicherungsvolumen bereinigt werden.
Für die auf das Jahr 2009 normierten Gesamtschäden wurde nun wieder eine lineare Trendfunktion berechnet. Diese ist als rote Linie in Abbildung 5.1(b) eingezeichnet. Durch die Normierung der Schadenshöhen tritt in diesem Fall nur mehr ein Anstieg von 18 Mio. Euro

zwischen den Jahren 1998 und 2009 auf. Schließt man auch hier die Gesamtschäden von Kyrill und Emma aus, ist der Trend der Schadenshöhen sogar leicht negativ (-2 Mio. seit 1998). Da alle verwendeten mathematischen Methoden aus Kapitel 3 invariant bzgl. der Multiplikation mit einer Konstanten sind, treten durch diese Umrechnung keine Probleme auf.

5.2 Schadensgradverteilung

Da saisonale Abhängigkeiten der Schadensdaten nicht in die Analyse miteinbezogen werden können und nach Normierung kein signifikanter Trend feststellbar ist, kann das einfache Modell der Jährlichkeiten aus Abschnitt 3.1.3 mit

$$J_u = \frac{\mathbb{E}[t_1]}{1 - F(u)}$$

verwendet werden. In dieser Arbeit soll aber nicht die Wiederkehrperiode einer gewissen Schranke, sondern jene der gemessenen Schadenshöhen betrachtet werden. Dafür setzt man den Schadensgrad $S_{j,k}$ des Sturmes k im PLZ-2-Steller j als Schranke u in die obige Gleichung ein. Vernachlässigt man auch den Erwartungswert des zeitlichen Abstandes zwischen zwei Schadensgraden beliebiger Höhe, erhält man

$$J_{j,k} = \frac{1}{1 - F_j(S_{j,k})}, \tag{5.1}$$

wobei F_j die Verteilung der Schadensgrade im PLZ-2-Steller j beschreibt.

Bemerkung. Diese Form der Jährlichkeit kann natürlich nicht als Anzahl von Jahren interpretiert werden, da der Faktor $\mathbb{E}[t_1]$ nicht berücksichtigt wurde. Die Ergebnisse für den Jahresgesamtschaden dieser Berechnungsmethoden wurden im Endeffekt aber wieder um die zeitliche Komponente des markierten Punktprozesses ergänzt, indem die Anzahl der Sturmereignisse pro Jahr als poissonverteilte Zufallsvariable angenommen wurde.

Um die Jährlichkeiten wie in (5.1) zu berechnen, ist es notwendig, die Verteilung der Schadensgrade zu kennen. Im ersten Schritt muss deshalb für jeden PLZ-2-Steller j die Verteilung der positiven Schadensgrade $S_{j,k}|S_{j,k} > 0$ der Ereignisses $k = 1, \ldots, 36$ gefunden werden. Dazu wurden statistische Methoden, wie z.B. Q-Q-Plots, P-P-Plots, Plots der empirischen Verteilungsfunktionen, der verschiedenen Verteilungsannahmen miteinander verglichen. Von einer Darstellung dieser Plots für die 76 betrachteten PLZ-2-Steller wird aber aus Gründen der Lesbarkeit verzichtet.
In den meisten PLZ-2-Stellern lieferte die Lognormal-Verteilung eine gute Anpassung. Diese Wahl ist passend, weil die Lognormalverteilung rechtsschief ist und deshalb oft zur Modellierung von Versicherungschäden verwendet wird. Die Verteilungsfunktion einer lognormalverteilten Zufallsvarible X mit Parametern $\mu \in \mathbb{R}$ und $\sigma > 0$, d.h. $X \sim \mathcal{LN}(\mu, \sigma)$,

ist durch

$$F_{\mu,\sigma}(x) = \int_0^x \frac{1}{\sqrt{2\pi}\sigma t} \exp\left(-\frac{(\ln(t)-\mu)^2}{2\sigma^2}\right) \mathrm{d}t$$

gegeben.
Nach Schätzung der Parameter μ und σ wurden auch der χ^2-Anpassungs-Test und der Shapiro-Wilk auf Normalverteilung auf die Daten bzw. die logarithmierten Daten angewendet. In Tabelle 5.1 sind die p-Werte dieser beiden Tests wiedergegeben. Der p-Wert

PLZ_j	N_j	p_{χ^2}	p_s	PLZ_j	N_j	p_{χ^2}	p_s	PLZ_j	N_j	p_{χ^2}	p_s
10	36	0.870	0.823	44	34	0.210	0.493	72	27	0.102	0.567
11	36	0.788	0.417	45	32	**0.002**	0.100	73	28	0.330	0.668
12	36	0.492	0.425	46	35	0.368	0.094	74	30	0.574	0.749
20	31	0.391	0.334	47	34	0.566	0.186	75	23	**0.000**	**0.002**
21	34	0.526	0.304	48	34	0.242	0.623	80	36	0.309	0.837
22	32	0.947	0.861	49	34	0.679	0.351	81	32	0.955	0.883
23	36	0.516	0.154	50	35	0.946	0.977	82	35	0.240	0.326
24	34	0.199	0.431	51	31	0.688	0.496	83	31	0.306	0.419
25	35	0.533	0.247	52	32	0.599	0.459	84	34	0.060	0.890
26	36	0.161	0.573	53	32	0.923	0.665	85	32	0.269	0.239
27	32	0.242	0.173	54	30	0.616	0.813	86	34	0.291	0.246
28	32	0.442	0.626	55	30	0.641	0.790	87	35	0.922	0.932
30	34	0.455	0.890	56	29	0.273	0.522	88	25	0.611	0.456
31	35	0.537	0.905	57	31	0.338	0.935	89	33	0.134	0.673
32	32	0.682	0.933	60	28	0.684	0.079	90	33	0.574	0.278
33	36	**0.041**	0.313	61	26	0.220	0.502	91	32	0.722	0.588
34	33	0.485	0.954	62	29	0.504	0.107	92	23	0.218	0.358
35	31	0.570	0.668	63	31	0.938	0.959	93	32	0.762	0.324
36	29	0.319	0.591	64	24	0.231	0.231	94	34	0.594	0.906
37	28	0.237	0.285	65	19	0.410	0.743	95	30	0.890	0.791
38	29	0.748	0.943	66	17	**0.000**	0.793	96	26	0.924	0.888
39	31	0.311	0.357	67	24	0.889	0.620	97	27	0.419	0.444
40	35	0.719	0.540	68	33	0.422	0.192	98	30	0.799	0.834
41	31	0.619	0.946	69	29	0.524	0.365	99	22	0.133	0.430
42	34	0.482	0.745	70	30	0.730	**0.048**				
43	34	0.883	0.920	71	26	0.473	0.108				

Tabelle 5.1: p-Werte der Lognormalverteilung

p_{χ^2} des χ^2-Tests unterschreitet das Signifikanzniveau nur in 4 PLZ-2-Stellern. Dieser Test ist allerdings nicht sehr genau, wenn die Wahrscheinlichkeiten der einzelnen Klassen zu klein sind. Deshalb sollte der Ablehnung der Hypothese einer Lognormalverteilung in den

PLZ-2-Stellern 66 und 75 nicht zu viel Bedeutung beigemessen werden. Die Normalverteilungsannahme der logarithmierten Daten wird durch den Shapiro-Wilk-Test hingegen nur im PLZ-2-Steller 70 und 75 verworfen. Da für den Großteil der PLZ-2-Steller keine Ablehnung der Nullhypothese einer Lognormalverteilung erfolgt, wird die Lognormalverteilung zur Modellierung der Schadensgrade verwendet.

Hierbei ist zu beachten, dass die Lognormalverteilung keine Schadensgrade der Höhe 0 zulässt. In den gegebenen Daten kommen jedoch in den meistem PLZ-2-Stellern auch Sturmereignisse ohne Schaden vor. Die Anteile der Schäden der Höhe 0 ist in Tabelle 5.2 für jeden PLZ-2-Steller aufgelistet. Dies ist vor allem dadurch zu erklären, dass die jeweiligen Sturmereignisse örtlich auf einige PLZ-2-Steller beschränkt waren und somit in anderen PLZ-2-Stellern nur wenige oder keine Sturmschäden auftraten.

PLZ_j	$\hat{p}_j$	PLZ_j	$\hat{p}_j$	PLZ_j	$\hat{p}_j$	PLZ_j	$\hat{p}_j$	PLZ_j	$\hat{p}_j$
10	0.000	34	0.083	50	0.028	68	0.083	88	0.306
11	0.000	35	0.139	51	0.139	69	0.194	89	0.083
12	0.000	36	0.194	52	0.111	70	0.167	90	0.083
20	0.139	37	0.222	53	0.111	71	0.278	91	0.111
21	0.056	38	0.194	54	0.167	72	0.250	92	0.361
22	0.111	39	0.139	55	0.167	73	0.222	93	0.111
23	0.000	40	0.028	56	0.194	74	0.167	94	0.056
24	0.056	41	0.139	57	0.139	75	0.361	95	0.167
25	0.028	42	0.056	60	0.222	80	0.000	96	0.278
26	0.000	43	0.056	61	0.278	81	0.111	97	0.250
27	0.111	44	0.056	62	0.194	82	0.028	98	0.167
28	0.111	45	0.111	63	0.139	83	0.139	99	0.389
30	0.056	46	0.028	64	0.333	84	0.056		
31	0.028	47	0.056	65	0.472	85	0.111		
32	0.111	48	0.056	66	0.528	86	0.056		
33	0.000	49	0.056	67	0.333	87	0.028		

Tabelle 5.2: Nullanteile der PLZ-2-Steller

Eine Möglichkeit, die Nullanteile der PLZ-2-Steller in das Modell einfließen zu lassen, ist die getrennte Modellierung der Ereignisse ohne Schaden und jener mit positivem Schaden. Man verwendet eine Bernoulli-verteilte Zufallsvariable S_j^0, welche die Schäden der Höhe Null im PLZ-2-Steller j beschreibt, und eine zweite Zufallsvariable S_j^1, welche die Verteilung der bedingten Schadensgrade $S_{j,k}|S_{j,k} > 0$ besitzt. Der zufällige Schadensgrad S_j kann dann als Produkt dieser beiden Zufallsvariablen dargestellt werden und besitzt somit die notwendigen Eigenschaften. Eine weitere Umsetzungsmethode bildet die Bestimmung der Verwendung einer unstetigen Verteilungsfunktion. Dies wurde schon z.B. in [8] in Verbindung mit einer generalisierten Pareto-Verteilung verwendet.

Bezeichne nun F_j die Verteilungsfunktion der $S_{j,k}|S_{j,k} > 0$ und p_j den Anteil an Schäden der Höhe 0 im PLZ-2-Steller j. Dann gilt für die Verteilungsfunktion $F_{j,0}$ der $S_{j,k}$:

$$\begin{aligned} F_{j,0}(x) &= \mathbb{P}[S_j = 0] + (1 - \mathbb{P}[S_j = 0])\mathbb{P}[S_j \leq x]\mathbb{1}_{x>0} \\ &\overset{F_j(0)=0}{=} p_j + (1 - p_j)F_j(x) \end{aligned}$$

Somit können die gegebenen Schadensgrade in Jährlichkeiten wie in (5.1) umgerechnet werden:

$$J_{j,k} = \frac{1}{\mathbb{P}[S > S_{k,j}]} = \frac{1}{1 - F_{i,0}(S_{i,k})} \tag{5.2}$$

Laut Lemma 9 gilt nun, dass die Jährlichkeiten als eine $Par(1)$-verteilte Zufallsvariable, der Kehrwert einer gleichverteilten Zufallsvariblen, beschrieben werden können. Dies wird auch durch die Plots der Verteilungsfunktion in Abbildung 5.2 und des Histogramms bzw. Log-Log-Plots in Abbildung 5.3 bestätigt. Dabei können die Abweichungen im unteren Teil des Log-Log-Plot möglicherweise durch eine obere Schranke der Schadensgrade erklärt werden (siehe [4]).

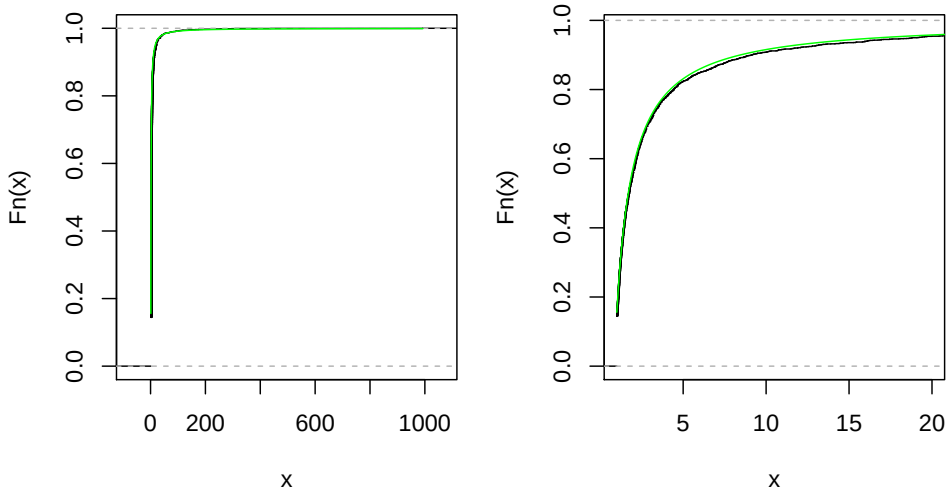


Abbildung 5.2: Empirische Verteilungsfunktion der Jährlichkeiten (grün: theoretische Verteilungsfunktion)

Die Umrechnung von Jährlichkeiten zu Schadensgraden erfolgt einfach über die Umkehrung von Definition in (5.1). Hier muss allerdings beachtet werden, dass deren Verteilungsfunktion in 0 eine Unstetigkeitsstelle besitzt. Die Jährlichkeiten können trotzdem als $Par(1)$-verteilte Zufallsvariable simuliert werden, bei der Rückrechnung zu Schadensgraden

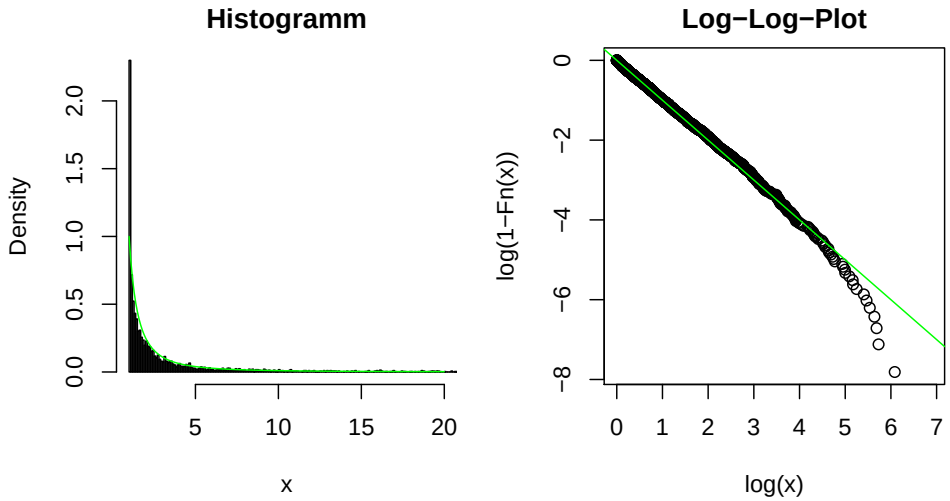


Abbildung 5.3: Histogramm und Log-Log-Plot der Jährlichkeiten (grün: theoretische Verteilungsfunktion)

muss aber Folgendes beachtet werden:

$$S_{j,k} = F_{j,0}^{-1}\left(1 - \frac{1}{J_{j,k}}\right) = \begin{cases} F_j^{-1}\left(1 - \frac{1}{(1-p_j)J_{j,k}}\right) & J_{j,k} > \frac{1}{1-p_j} \\ 0 & J_{j,k} \leq \frac{1}{1-p_j} \end{cases}.$$

Bemerkung. Eine weit verbreitete Modellierungsmethode von Schäden, welche eine bestimmte Schranke überschreiten, ist die Verwendung einer Generalisierten Pareto-Verteilung aus Abschnitt 3.2. Deshalb wurde auch versucht, diese Verteilungsfunktion an die gegebenen Sturmschadendaten anzupassen. Durch die wenigen Datenpunkte konnte aber innerhalb dieser keine Schranke gefunden werden, die passend schien. Die Auswahl wurde für die meisten PLZ-2-Steller weiters dadurch erschwert, dass viele Sturmereignisse ohne Schäden vorliegen. Auch ein durchgeführter χ^2-Anpassungstest wurde in 46 PLZ-2-Stellern abgelehnt. Der geschätzte Tail-Koeffizient $\hat{\alpha}$ liegt für 72 PLZ-2-Steller im Intervall $(0, 2)$. Somit würde für jene PLZ-2-Steller, in welchen $\alpha < 1$ gilt, kein endlicher Erwartungswert vorliegen. Grundsätzlich muss also davon ausgegangen werden, dass die Sturmschäden der einzelnen PLZ-2-Steller noch keiner generalisierten Pareto-Verteilung folgen.

5.3 Nachbarschaftsrelationen

Die Distanzmatrixmethode nach Url wird vor allem dazu verwendet, Extremereignisse mit großer Jährlichkeit (hier mindestens 20 Jahre) zu modellieren. Wie man in den Abbildungen 5.4, 5.5 und 5.6 erkennen kann, sammeln sich die Schäden mit hohen Jährlichkeiten in den drei Stürmen mit dem größten Gesamtschaden meist in einem Gebiet aus mehreren zusammenhängenden PLZ-2-Stellern. Die restlichen Jährlichkeitsplots der einzelnen Sturmereignisse zeigen auch lokale Ansammlungen von hohen Jährlichkeiten. Meist sind diese

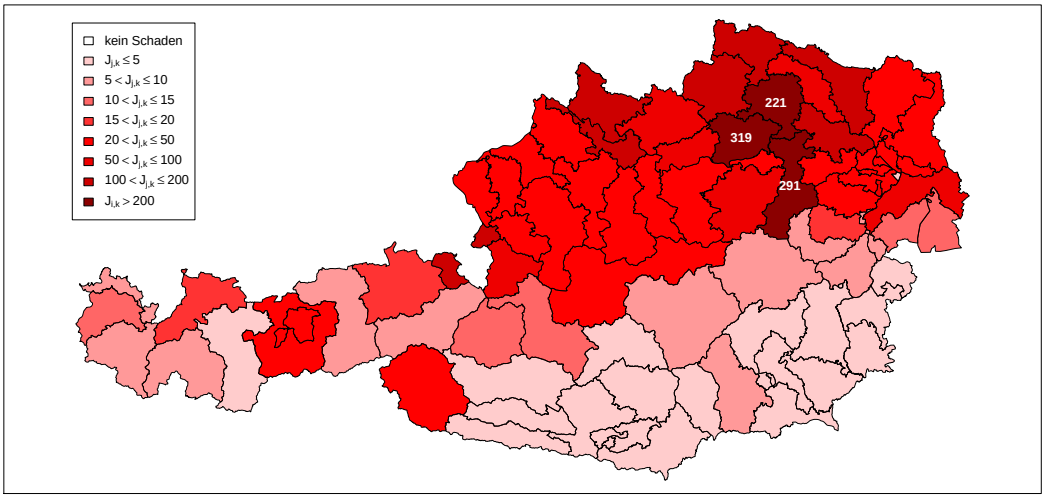


Abbildung 5.4: Kyrill

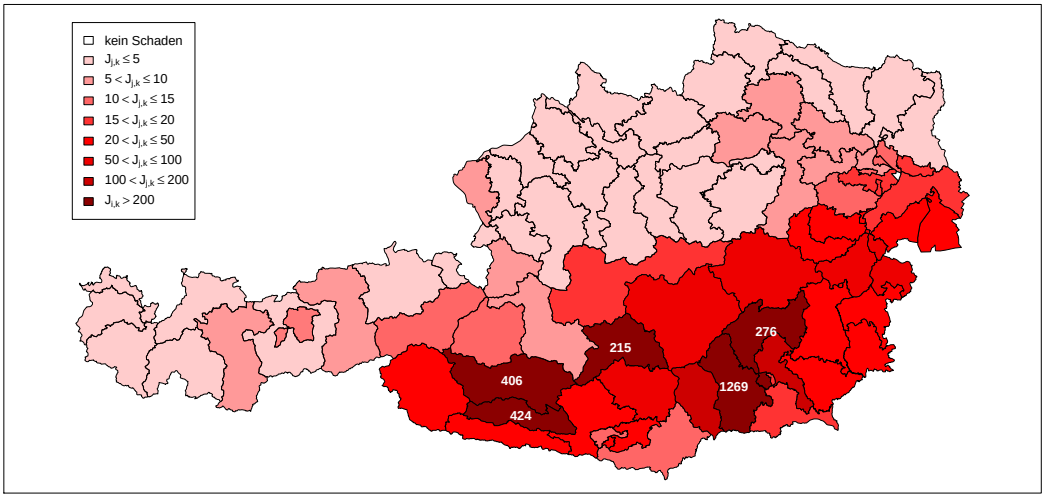


Abbildung 5.5: Paula

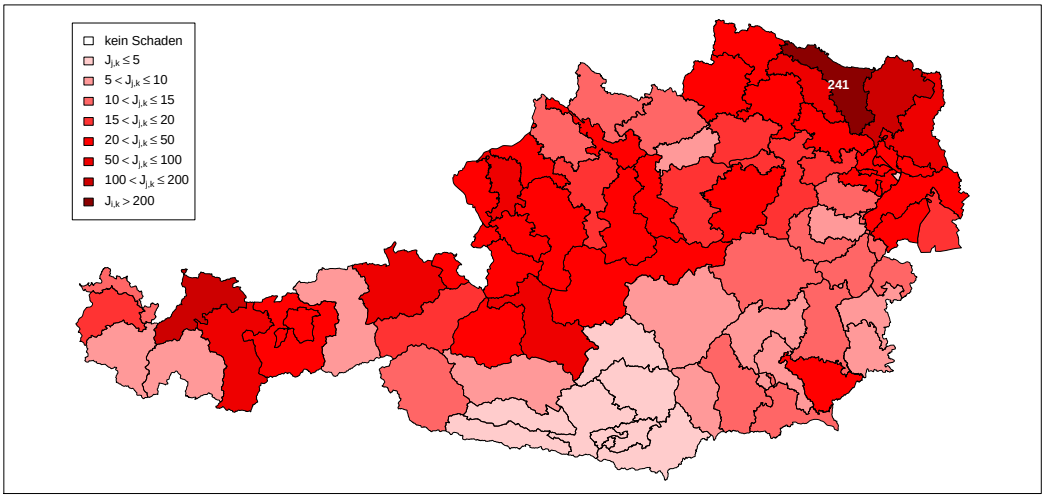


Abbildung 5.6: Emma

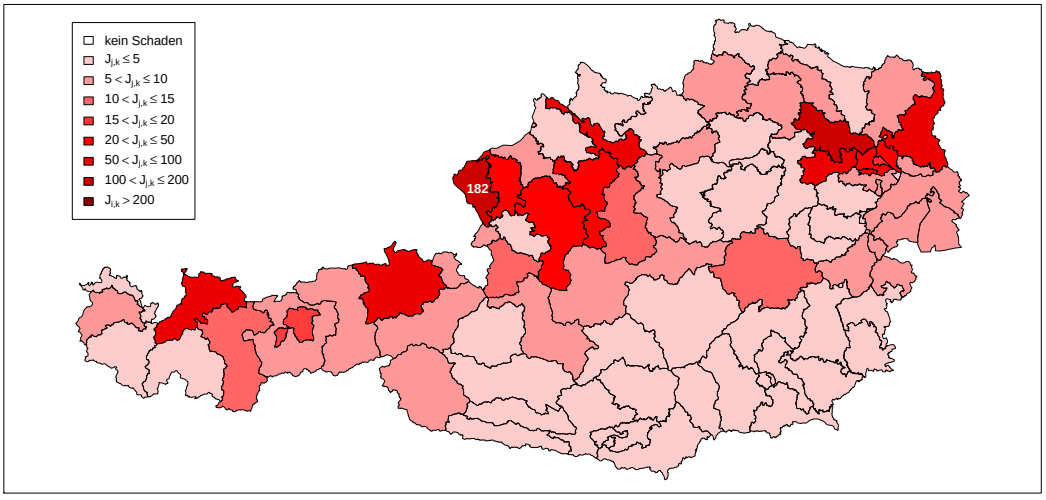


Abbildung 5.7: Wolfgang

Gebiete aber kleiner als die von Kyrill, Paula und Emma betroffenen.
Anders hingegen verhält sich der Sturm Nr. 38 (vgl. Abbildung 5.7), hier verteilen sich mehrere kleinere „Sturmzellen“ mit hohen Jährlichkeiten über ganz Österreich. Dieses Ereignis ist aber lt. [40] vermutlich durch Überflutungsschäden verunreinigt und wurde ohnehin von den Berechnungen ausgenommen.
Im Folgenden wird nun eine Nachbarschaftsbedingung für jeden PLZ-2-Steller definiert, sodass vor allem Sturmereignisse mit großer geografischer Ausbreitung dargestellt werden können.
In der Hochwasserstudie [39] wird eine Nachbarschaft zwischen zwei PLZ-2-Stellern dadurch induziert, dass die 1.5fache Entfernung zwischen den beiden Schwerpunkten weniger als 100km beträgt. Hierbei muss beachtet werden, dass für die Modellierung von Hochwassern die HQ_n-Zonen vorliegen. Dadurch ist bei Auftritt einer hohen Jährlichkeit nicht die gesamte Gemeinde betroffen, sondern nur die zu dieser Jährlichkeit gehörende Zone. Die Verwendung derselben Nachbarschaftsbedingung führte im Sturmschadenmodell aber auf zu große Sturmgebiete und somit auch zu einer Überschätzung der Schäden.

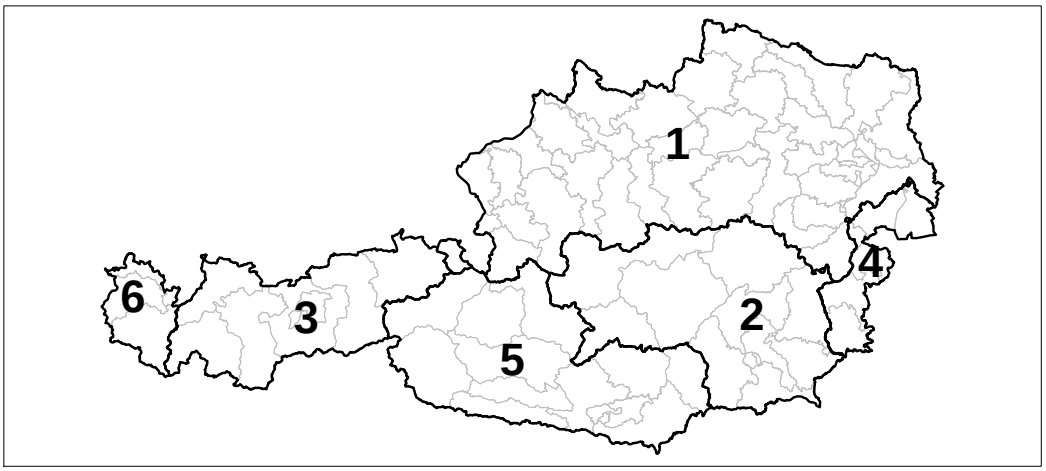


Abbildung 5.8: Einteilung Österreichs in verschiedene Regionen

Die Plots der aus den Daten berechneten Jährlichkeiten lassen vermuten, dass Sturmereignisse in Österreich örtlich durch die Alpen begrenzt sind. So zeigen sich die Windgebiete vor allem nördlich oder südöstlich der Alpen bzw. in deren Tälern (Tirol bis Salzburg). Dies entspricht den meteorologischen Eigenschaften zur Entstehung starker Winde in Österreich (siehe Kapitel 2). Aufgrund dieser Überlegungen wurde Österreich in 6 Regionen unterteilt (vgl. Abbildung 5.8). Vergleicht man diese Abbildung mit jenen der Jährlichkeiten der extremsten Sturmereignisse, so kann man erkennen, dass die Stürme Kyrill und Emma beinahe die gesamte Region 1 getroffen haben. Paula hingegen hat sich über die Regionen 2, 4 und 5 erstreckt. Die übrigen Stürme haben jeweils nur Teile der Regionen getroffen. Deshalb ist es wichtig, innerhalb dieser Regionen auch eine Nachbarschaftsbedingung der PLZ-2-Steller einzuführen und somit die Ausbreitung der möglichen Sturmereignisse zu

beschränken.
Dabei muss angemerkt werden, dass die hier definierten Regionen großteils nicht mit den sogenannten PLZ-1-Stellern übereinstimmen. Die PLZ-1-Steller sind dabei jene Gebiete Österreichs, deren Postleitzahlen mit derselben Ziffer beginnen. Bis auf wenige Ausnahmen stimmen die PLZ-1-Steller mit den Bundesländern überein. Die Tabelle 5.3 stellt dar, welche Bundesländer zu den oben definierten Regionen gehören.

Region 1	Oberösterreich, Niederösterreich, nördlicher Teil Salzburgs, Wien
Region 2	Steiermark
Region 3	Tirol
Region 4	Burgenland
Region 5	Kärnten, Osttirol, südlicher Teil Salzburgs
Region 6	Vorarlberg

Tabelle 5.3: Regionen

Weiters kann man in der Abbildung 5.8 erkennen, dass nicht nur die Größe der Regionen, sondern auch die Anzahl der PLZ-2-Steller pro Region unterschiedlich ist (siehe auch Tabelle 5.4).
Die unterschiedliche Größe ergibt sich aus den geografischen Eigenheiten der jeweiligen Regionen. So finden sich z.B. in Region 1 keine extremen Rauigkeiten der Erdoberfläche oder natürliche Wind-Grenzen wie Gebirge. Deshalb ist der gesamte Donauraum häufig starken und relativ großflächigen Stürmen ausgesetzt. Die große Anzahl an PLZ-2-Stellern in der Region 1 kann aber auch dadurch erklärt werden, dass die Bundesländer Oberösterreich, Niederösterreich und Wien in kleinere PLZ-2-Steller als der Rest Österreichs gegliedert sind.

	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
Anzahl	38	10	7	6	13	3

Tabelle 5.4: Anzahl der PLZ-2-Steller in den Regionen

Nun sollen für diese Regionen unterschiedliche Nachbarschaftsrelationen definiert werden. Wie bereits beschrieben, ist vor allem die Region 1 häufig von großen Windsystemen betroffen. Deshalb wurden hier die Nachbarschaftsbedingungen anders gewählt als in den übrigen Regionen. Da die meisten Stürme in Österreich Westwinde sind, wurden auch je nach Himmelsrichtung unterschiedliche Bedingungen definiert. Somit hängt die Nachbarschaft zweier PLZ-2-Steller davon ab, dass die Distanz ihrer Schwerpunkte in Nord-Süd- und in Ost-West-Richtung eine gewisse Entfernung nicht überschreitet. Diese Entfernungen sind in in Tabelle 5.5 angegeben.
Die Grenzen zwischen zwei PLZ-2-Steller müssen allerdings nicht an natürliche Sturmgrenzen wie z.B. hohe Gebirge gebunden sein. Dies trifft nicht immer auf Gebiete an der Grenze der einzelnen Regionen zu. Dadurch kann in einer Region ein Sturm entstehen, der auch auf andere Regionen übergreift. Dort sollte er aber schon durch die topografische Oberfläche

der PLZ-2-Stellern an den Grenzen der Region abgeschwächt sein, sodass er nur noch die Randgebiete betrifft.
Aus diesem Grund wurde hier die Nachbarschaftsrelation auch auf PLZ-2-Steller verschiedener Regionen ausgedehnt. Dazu darf die Entfernung der beiden PLZ-2-Steller aber nicht mehr so groß sein. Diese Einschränkung soll die Abschwächung des Sturmes abbilden.

	Nachbar in			
	gleicher Region		anderer Region	
	Nord-Süd	West-Ost	Nord-Süd	West-Ost
PLZ-2-Steller in				
- Region 1	100	200	40	80
- Region 2	50	100	20	40
- Region 3	50	100	20	40
- Region 4	50	100	20	40
- Region 5	50	100	20	40
- Region 6	50	100	20	40

Tabelle 5.5: Nachbarschafts-Distanzen der verschiedenen Regionen

Um zu bestimmen, ob der PLZ-2-Steller j ein Nachbar von i ist, wurden um den Schwerpunkt des Polygons i zwei Rechtecke mit den Kantenlängen aus Tabelle 5.5 gelegt. Durch Verwendung des Programmpackets **sp** der Programmiersprache **R** wurde überprüft, ob der Schwerpunkt des PLZ-2-Stellers j in das richtige Rechteck fällt. Eine Beschreibung der möglichen Bearbeitung von räumlichen Polygonen („spatial polygons“) in **R** findet sich in [1].
Dabei wurde die Erdkrümmung bei der Berechnung der Distanzen vernachlässigt. Diese führt aber auf einem recht kleinen Gebiet wie Österreich nicht zu ausschlaggebenden Fehlern.

5.4 Distanzmatrixmethode

Nun soll die Distanzmatrix-Methode nach Url verwendet werden, um eine räumliche Abhängigkeitsstruktur der Jährlichkeiten der Schadensgrade darzustellen. Dazu ist diesese Methode im Folgenden beschrieben und anschließend werden die Änderungen des Modell zur Anpassung auf extreme Sturmereignisse erläutert.

5.4.1 Hochwassermodell nach Url - Überblick

Die in [39] verwendeten Modelle können in zwei Module aufgeteilt werden. Im ersten Modul (siehe Punkt 5.4.2) soll für alle österreichischen Bezirke die jeweilige Jährlichkeit eines Hochwassers simuliert werden. Die Jährlichkeit eines Hochwassers ist definiert als der Kehrwert der Wahrscheinlichkeit, dass ein Hochwasser mit demselben oder höherem Pegelstand als jener des betrachteten Hochwassers innerhalb einer Beobachtungsperiode von einem

Jahr auftritt. Da die Wahrscheinlichkeit für ein Hochwasser mit Jährlichkeit x genau $\frac{1}{x}$ beträgt, kann die Jährlichkeit als $Par(1)$-verteilte Zufallsvariable modelliert werden. Da die Methoden vor allem auf extreme Hochwasser angewendet werden sollen, werden in dieser Simulationmethode allerdings nur Jährlichkeiten in der Höhe von mindestens 20 Jahren erzeugt. Anhand der im Modul 1 simulierten Jährlichkeiten werden im Modul 2 (siehe Punkt 5.4.3) die Einzelschäden generiert und anschließend zu Gesamtschäden zusammengefasst.

5.4.2 Modul 1

Den Mittelpunkt der Distanzmatrixmethode bildet die Nachbarschaftsrelation. In [53] gelten zwei Bezirke als benachbart, wenn die Fahrtstrecke zwischen deren Hauptstädten maximal 100km aufweist. In [39] wurde diese Bedingung verändert, sodass zwei Bezirke dann benachbart sind, wenn das 1.5fache der Distanz zwischen ihren Schwerpunkten weniger als 100km beträgt.
Im ersten Schritt wird für jeden Bezirk d, $d = 1, \ldots, D$, eine gleichverteilte Zufallsvariable x_d generiert. In jenen Bezirken, in welchen folgende Ungleichung erfüllt ist, tritt ein Hochwasser mit Jährlichkeiten von mindestens 20 auf:

$$x_d \leq \frac{c \cdot \frac{1}{D} \sum_{j=1}^{D} \#\{\text{Nachbarn des Bezirkes } j\}}{\#\{\text{Nachbarn des Bezirkes } d\}}.$$

Dabei wird die Konstante c so gewählt, dass jeder Bezirk im Durchschnitt mit Wahrscheinlichkeit $\frac{1}{20}$ betroffen ist. Da die Jährlichkeiten $Par(1)$-verteilt ist, kann die Jährlichkeit J_d eines betroffenen Bezirkes d als $Par(1, m20)$-verteilte Zufallsvariable, mit Verteilungsfunktion

$$F(x) = \begin{cases} 1 - \frac{20}{x}^{\alpha} & x \geq 20 \\ 0 & x < 20 \end{cases}$$

simuliert werden. Für alle Nachbarn dieses Bezirkes wird die Jährlichkeit als $J_d|1 + y_d|$ berechnet, wobei y_d eine normalverteilte Zufallsvariable mit bestimmter Varianz ist. Der Nachteil dieser Methode ist, dass die Jährlichkeiten somit nicht mehr Pareto-verteilt sind. Um die durch die Distanzmatrixmethode gewonnenen Ergebnisse zu vergleichen, wurde in der Hochwasser-Studie die Abhängigkeit der Jährlichkeiten in den einzelnen Bezirken mit einer Gumbel-Copula modelliert. Dabei wird angenommen, dass die Jährlichkeiten der Gemeinden Österreichs, bedingt auf eine unbekannte Zufallsvariable, unabhängig sind. Die Abhängigkeit der Jährlichkeiten von dieser zufälligen Größe wird dann durch eine Gumbel-Copula beschrieben. Das Problem der Copulamethode besteht darin, dass der Parameter θ der Gumbel-Copula nicht geschätzt werden kann.

5.4.3 Modul 2

Im Vorfeld der Studie wurden HQ_n-Zonen durch geeignete hydrologische, geografische und statistische Methoden entwickelt. Diesen Zonen sind in der HORA[1]-Datenbank gesammelt und über das Internet ist es möglich diese Einteilung Österreichs abzurufen ([36]).
Anhand dieser Zonen können nun mit den in Modul 1 simulierten Jährlichkeiten die betroffenen Gebiete Österreichs bestimmt werden. Je nach Jährlichkeit der simulierten Hochwasser ändert sich die Höhe des Schadens an den betroffenen Gebäuden. So wird z.B. angenommen, dass bei einer Jährlichkeit kleiner als 100 nur der Keller der betroffenen Gebäude überschwemmt ist. Die Höhe der Schäden wird als lognormalverteilte Zufallsvariable simuliert, deren Parameter allerdings von der Höhe der Jährlichkeit abhängen.

Die zentrale Überlegung der Distanzmatrixmethode ist, dass sich eine hohe Jährlichkeit in einem PLZ-2-Steller auf die Höhe die Jährlichkeiten in dessen Nachbarn auswirkt. Im ursprünglichen Hochwasser-Modell wird die örtliche Abhängigkeit für PLZ-2-Stellern mit einer Jährlichkeit größer 20 betrachet. Diese Schranke wurde für Sturmschäden übernommen, die folgenden Modelle können aber auf beliebige Grenzen angewandt werden. Gegebenfalls sollten jedoch neue Nachbarschaftsbedingungen definiert werden.
Im ersten Schritt wird zufällig ein PLZ-2-Steller s („Startgebiet“) ausgewählt, von dem aus sich der Sturm ausbreiten soll. Zur Bestimmung des Startgebiets wird für jeden PLZ-2-Steller j eine gleichverteilte Zufallsvariable U_j mit dem sogenannten „kritischen Wert“$cval_j$ verglichen. Dabei soll $cval_j$ garantieren, dass die Wahrscheinlichkeit für einen Schadengrad mit Jährlichkeit größer 20 im PLZ-2-Steller j genau $\frac{1}{20}$ beträgt. Gilt $U_s < cval_s$, ist der PLZ-2-Steller s ein Startgebiet. Da dessen Jährlichkeit J_s größer als 20 sein soll, folgt diese einer $Par(1, 20)$-Verteilung.

Bemerkung. Durch diese Auswahlmethode können auch Sturmereignisse mit mehreren Startgebieten auftreten. In diesem Fall wird für alle Startgebiete dieselbe Jährlichkeit angenommen.

In jenen PLZ-2-Stellern, die gemäß Abschnitt 5.3 keine Nachbarn des Startgebietes sind, wird die Jährlichkeit als eine Zufallsvariable mit einer nach oben beschränkten Pareto-Verteilung $TPar(1, 1, \max = 20)$ angenommen, d.h. die Verteilungsfunktion ist

$$G_{\max}(y) = \begin{cases} \frac{20}{20-1}\left(1 - \frac{1}{y}\right) & y \leq 20 \\ 1 & y > 20. \end{cases}$$

Für die Nachbarn des Startgebietes s wurden zwei verschiedene Modelle zur Berechnung der Jährlichkeit verwendet, welche in den folgenden Punkten (5.4.4 und 5.4.5) beschrieben werden.

[1]Natural Hazards Overview Risk Assessment Austria

5.4.4 Verteilung der Abweichungen

Im Distanzmatrixmodell nach Url wird angenommen, dass sich die Jährlichkeiten der Nachbarn j des Startgebietes s durch eine multiplikative Abweichung $Z_{i,s}$ von J_s unterscheiden. Dabei werden die $Z_{j,s}$ als Betrag einer normalverteilten Zufallsvariable $Y_{i,j}$ mit Erwartungswert $\mu = 1$ und Varianz $\sigma^2 = 0.1$ beschrieben. Die Jährlichkeit J_j kann also folgendermaßen dargestellt werden:

$$J_j = J_s \cdot Z_{j,s} = J_s \cdot |Y_{j,s}| \qquad \text{mit } Y_{i,s} \sim N(1; 0, 1)$$

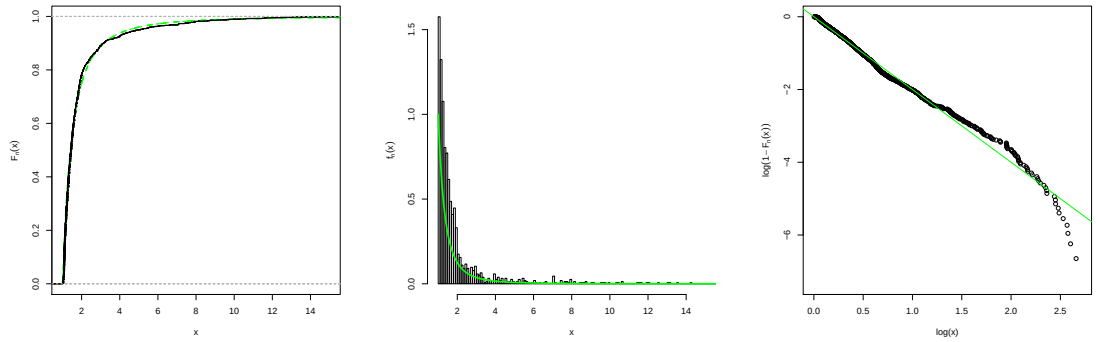
Die Verteilung des Betrages einer normalverteilten Zufallsvariablen wird in der englischsprachigen Literatur als „Folded Normal Distribution“bezeichnet. Sie wird vor allem zur statistischen Analyse von Messfehlern, deren Vorzeichen nicht bestimmt werden kann, verwendet (siehe z.B. [31]). Eine direkte Übersetzung wäre die Bezeichnung „gefaltete Normalverteilung“, diese kann allerdings leicht zu Verwechslungen führen. In [44], S. 215, wird sie als „gespiegelte “Normalverteilung übersetzt.
Die Verteilungsfunktion F_Z und Dichte f_z einer gespiegelten Verteilung können durch die Verteilungs- und Dichtefunktion (F_y bzw. f_y) der zugrundeliegenden Verteilung relativ einfach dargestellt werden:

$$\begin{aligned} F_Z(z) &= \mathbb{P}[Z \leq z] = \mathbb{P}[|Y| \leq z] = \mathbb{P}[-z \leq Y \leq z] = F_Y(z) - F_Y(-z) \\ f_Z(z) &= \frac{\partial F_Z(z)}{\partial z} = \frac{\partial F_Y(z)}{\partial z} - \frac{\partial F_Y(-z)}{\partial z} = f_Y(z) + f_Y(-z). \end{aligned}$$

Eine Parameterschätzung der gespiegelten Normalverteilung gestaltet sich im Allgemeinen eher schwierig, weil die beiden Parameter μ und σ Formparameter sind. In [16] wird vorgeschlagen, die 2. und 4. Momente zur Schätzung von μ und σ zu verwenden, wenn der Quotient des Stichprobenmittels und der Stichprobenstandardabweichung, m/s, kleiner als 1.35 ist.
Diese Schätzung wurde an den Abweichungen der berechneten Jährlichkeiten durchgeführt. Da durch die Daten nicht erklärt werden kann, welcher PLZ-2-Steller das Startgebiet bildet, wurden die $Z_{j,l,k}$ der Gebiete j und l und des Sturmereignisses k folgendermaßen bestimmt:

$$Z_{j,l,k} = \begin{cases} \frac{J_{j,k}}{J_{l,k}} & J_{j,k} > 20 \text{ und } J_{l,k} > 20 \\ \frac{\min(J_{j,k}, J_{l,k})}{\max(J_{j,k}, J_{l,k})} & J_{j,k} > 20 \text{ oder } J_{j,k} > 20, \ \min(J_{j,k}, J_{l,k}) > 0. \end{cases}$$

Für diese Daten liegt m/s bei ca. 0,5. Dieser Wert ist aber sehr klein und somit kann die Momentenmethode zu falschen Ergebnissen führen. Deshalb wurde auch mit einer Maximum-Likelihood-Schätzung versucht, passende Parameter zu finden. Beide Methoden führten jedoch nicht zur Anpassung der gespiegelten Normalverteilung an die Abweichungen der Jährlichkeiten.
Aus diesem Grund wurde mit den üblichen statistischen Methoden eine passende Verteilung der $Z_{j,l}$ ermittelt. Die beste Anpassung lieferte die verschobene Pareto(2)-Verteilung, d.h. $1 + Z_{j,l} \sim Par(2)$ (vgl. Abbildung 5.9).



(a) ECDF (b) Histogramm (c) Log-Log-Plot

Abbildung 5.9: Grafische Überprüfung der Anpassung der Pareto(2)-Verteing an die Abweichungen der Jährlichkeiten

Für beide Verteilungsannahmen der Schadensgrade könnte der kritische Wert jedes PLZ-2-Steller analytisch berechnet werden. Dazu muss aber zuerst die Verteilungsfunktion der Jährlichkeiten in den Nachbargebieten des Startgebietes s berechnet werden. Sei $J \sim Par(1, min = 20)$ und $z \sim Par(2)$, z und J unabhängig. Dann gilt:

$$\begin{aligned} F_{J\cdot(z-1)}(x) &= \mathbb{P}\left[J \cdot (z-1) \leq x\right] = \mathbb{P}\left[J \leq \frac{x}{z-1}\right] \\ &= \int_{1}^{\frac{x}{20}+1} \int_{20}^{\frac{x}{z-1}} f_J(j) f_Z(z) dj dz = \int_{1}^{\frac{x}{20}+1} \int_{20}^{\frac{x}{z-1}} \frac{20}{j^2} \cdot \frac{2}{z^3} dj dz \\ &= \frac{x}{x+20} \end{aligned}$$

Bemerkung. Diese Verteilung ist keine $Par(1)$-Verteilung, d.h. die Jährlichkeiten in diesem Modell stimmen nicht mehr mit jenen aus 5.1 überein. Bei der Rückrechnung zu Schadensgraden wird also ein Fehler in Kauf genommen.

Sei n die Anzahl der Nachbarn j_k, $k = 1, \ldots, n$, des PLZ-2-Stellers l. Für den kritischen wert $cval_l$ gilt somit:

$$\frac{1}{20} \stackrel{!}{=} \mathbb{P}\left[U_l < cval_l \vee (\exists \text{mind. ein } j_k,\ 1 \leq k \leq n:\ U_{j_k} < cval_{j_k} \wedge J_{j_l} \cdot Z_{j_l,l} > 20)\right] \tag{5.3}$$

$$\begin{aligned}
&= cval_l + \sum_{k=1}^{n} cval_{j_k} \mathbb{P}\left[J_{j_k} \cdot Z_{j_k,l} > 20\right] \\
&\quad + \sum_{\substack{k=1 \\ m=k+1}}^{n} cval_{j_k} cval_{j_m} \mathbb{P}\left[J_{j_k} \cdot Z_{j_k,l} > 20 \text{ und } J_{j_m} \cdot Z_{j_m,l} > 20\right] \\
&\quad + \sum_{\substack{k=1 \\ m=k+1 \\ i=k+2}}^{n} cval_{j_k} cval_{j_m} cval_{j_i} \mathbb{P}\left[J_{j_k} \cdot Z_{j_k,l} > 20 \text{ und } J_{j_m} \cdot Z_{j_m,l} > 20 \text{ und } J_{j_i} \cdot Z_{j_i,l} > 20\right] \\
&\quad + \cdots
\end{aligned}$$

PLZ	#	PLZ	#	PLZ	#	PLZ	#	PLZ	#	PLZ	#
10	28	31	37	44	35	57	3	72	6	89	3
11	29	32	36	45	33	60	6	73	5	90	8
12	28	33	35	46	33	61	6	74	3	91	8
20	26	34	31	47	28	62	5	75	3	92	8
21	24	35	31	48	27	63	4	80	6	93	10
22	25	36	33	49	28	64	5	81	8	94	6
23	30	37	27	50	17	65	5	82	8	95	8
24	28	38	18	51	20	66	5	83	6	96	7
25	32	39	28	52	22	67	2	84	3	97	8
26	32	40	33	53	22	68	2	85	7	98	10
27	31	41	31	54	18	69	1	86	5	99	5
28	24	42	33	55	6	70	6	87	6		
30	34	43	36	56	5	71	3	88	4		

Tabelle 5.6: Anzahl der Nachbarn der einzelnen PLZ-2-Steller

Durch die Definition der Nachbarschaften in Abschnitt 5.3 liegt die Anzahl der Nachbarn, n, zwischen 1 und 37. Die genauen Zahlen sind in Tabelle 5.6 gegeben. Durch diese große Anzahl wird das Gleichungssystem zur Berechnung der kritischen Werte zu rechenintensiv. Deshalb wurden die verschiedenen $cval_k$ bzw. $cval_l$ in (5.3) gleichgesetzt. Außerdem wurde noch die Vereinfachung getroffen, dass ein Gebiet nur von jeweils einem Startgebiet beeinflusst wird, auch wenn es mit mehreren Startgebieten benachbart ist. Damit folgt

$$\begin{aligned}
\frac{1}{20} \stackrel{!}{=} & cval + \sum_{k=1}^{n} \binom{n}{k} cval^k \cdot (1 - F_{J \cdot Z}(20))]) \\
= & cval + (1 - F_{J \cdot (z-1)}(20))((1 + cval)^n - 1)
\end{aligned}$$

Damit ist diese Annäherung des kritischen Wertes eines PLZ-2-Stellers nur von der Anzahl der Nachbarn abhängig. Deshalb wurden kritische Werte $cval_n$ für $n = 1, \ldots, 37$ berechnet und dann den PLZ-2-Stellern je nach Anzahl ihrer Nachbarn zugewiesen.

5.4.5 Copula-Modell der Nachbarn

Wie in Kapitel 4 beschrieben wurde, werden häufig Copula-Funktionen zu Modellierung der Abhängigkeiten von Zufallsvariablen verwendet. Die Wahl der richtigen Copula ist aber nicht offensichtlich. Manchmal kann durch Scatterplots des bivariaten gleichverteilten Vektors (U_1, U_2) wie in Abbildung 4.1 bzw. durch empirische Schätzer der Tailabhängigkeiten (vgl. [48] oder [19]) die Menge der möglichen Copula-Funktionen eingegrenzt werden. Da in den Sturmdaten allerdings nur 36 Datenpunkte pro PLZ-2-Steller vorliegen, sind diese Methoden nicht sehr aussagekräftig.
Wie die Distanzmatrixmethode soll auch das Copula-Modell vor allem die Abhängigkeiten extremer Sturmereignisse mit Jährlichkeiten von mindestens 20 Jahren beschreiben. Deshalb ist hier die Wahl einer Gumbel-Copula sinnvoll, die lt. Beispiel 10 eine hohe obere Tail-Abhängigkeit besitzt.
Die multivariate Gumbel-Copula hängt nur von einem Parameter θ ab, der durch Kendalls τ (siehe Beispiel 12 bzw. 13) geschätzt wird. Da die Schätzer der Kendalls τ allerdings stark variieren (vgl. Abb. 5.10), wird für jede Nachbarschaftsbeziehung ij eine eigene bivariate Gumbel-Copula mit Parameter θ_i angenommen.

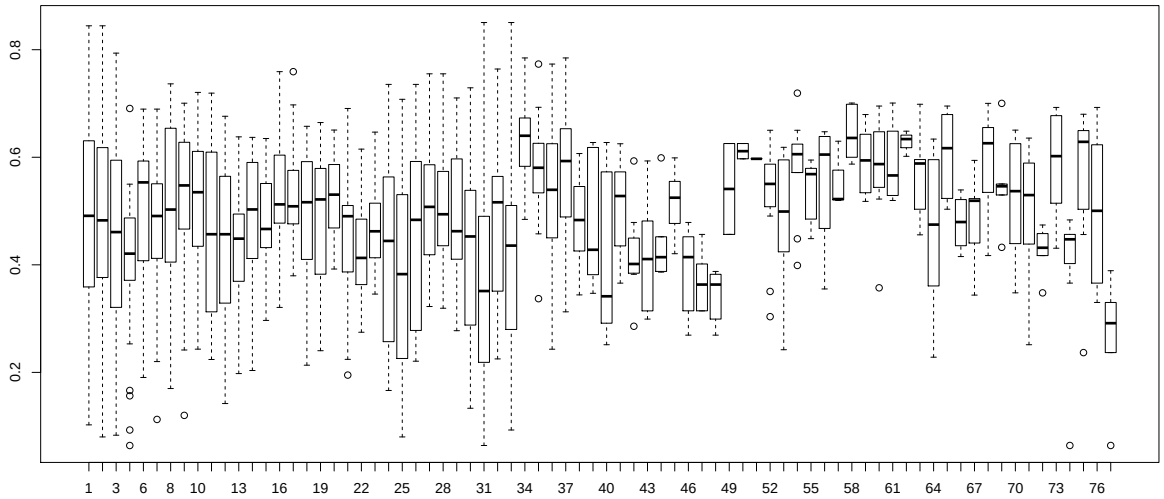


Abbildung 5.10: Boxplot der Kendalls τ für benachbarte PLZ-2-Steller

Auch in diesem Modell wird davon ausgegangen, dass sich die PLZ-2-Steller i mit $J_i > 20$ um ein Startgebiet s mit $J_s > 20$ sammeln und die restlichen Jährlichkeiten wieder

$TPar(1, 1, 20)$-verteilt sind. Dabei erfolgt die Wahl des Startgebietes und der betroffenen Nachbarn analog zur Distanzmatrixmethode, allerdings wurden hier die kritischen Werte durch eine Monte-Carlo-Simulation angepasst.
Da alle J_i von der Jährlichkeit des Startgebietes s abhängen, kann nicht auf den weitverbreiteten LT-Simulationsalgorithmus für Archimedische Copulas zurückgegriffen werden. Dieser verwendet die Eigenschaften der Laplace-Stieltjes-Transformation der Archimedischen Copulas C, um gleichverteilte Zufallsvariablen mit gemeinsamer Copula C zu simulieren (siehe [34], S. 222-224). Dabei ist es aber nicht möglich eine dieser Zufallsvariablen festzuhalten. Hier sollen allerdings die Jährlichkeiten J_i der Nachbarn des Startgebietes s in Abhängigkeit von J_s berechnet werden, d.h. man greift auf die bedingte Copula $C(u_i|u_s)$ zurück, wobei $u_i = 1 - 1/J_i$ und $u_s = 1 - 1/J_s$ $\forall i$ ist.

Setzt man also den Generator der Gumbel-Copula

$$\varphi(u) = (-\log(u))^\theta \qquad \theta \geq 1$$

und dessen Inverse $\varphi^{-1}(s) = \exp\left(-s^{\frac{1}{\theta}}\right)$ in die bedingte Copula-Funktion aus (16) ein, erhält man:

$$C(u_2|u_1) = \frac{-\left[\varphi(u_1) + \varphi(u_2)\right]^{\frac{1}{\theta}-1} \cdot \frac{1}{\theta} \cdot \exp\left(-\left[\varphi(u_1) + \varphi(u_2)\right]^{\frac{1}{\theta}}\right)}{-\left[\varphi(u_1)\right]^{\frac{1}{\theta}-1} \cdot \frac{1}{\theta} \cdot \exp\left(-\left[\varphi(u_1)\right]^{\frac{1}{\theta}}\right)}.$$

Durch Verwendung der Substitution

$$c = -\theta \frac{\partial \varphi}{\partial u_1}(\varphi(u_1)) = \left[\varphi(u_1)\right]^{\frac{1}{\theta}-1} u_1 \qquad \text{und} \qquad x = \varphi(u_1) + \varphi(u_2) \tag{5.4}$$

kann der zweite Schritt des hierarchischen Simulationsalgorithmus vereinfacht werden:

$$\begin{aligned} & v = \frac{1}{c} x^{\frac{1}{\theta}-1} \exp\left(-x^{\frac{1}{\theta}}\right) \\ \iff \qquad & (cv)^{\frac{1}{1-\theta}} = x^{\frac{1}{\theta}} \exp\left(-\frac{x^{\frac{1}{\theta}}}{1-\theta}\right) \\ \iff \qquad & -\frac{(cv)^{\frac{1}{1-\theta}}}{1-\theta} = -\frac{x^{\frac{1}{\theta}}}{1-\theta} \exp\left(-\frac{x^{\frac{1}{\theta}}}{1-\theta}\right). \end{aligned} \tag{5.5}$$

Die obige Gleichung ist nicht analytisch lösbar, kann aber mithilfe der LambertW-Funktion numerisch berechnet werden.

Definition 5.4.1: LambertW-Funktion
Als LambertW-Funktion bezeichnet man die inversen Funktionen von $f(w) = w \exp(w)$, d.h. die möglichen Zweige $W(z)$ erfüllen

$$z = W(z)\, \mathrm{e}^{W(z)}.$$

Für die meisten Werte $z \in \mathbb{R}$ kann $W(z)$ nicht analytisch bestimmt werden. Die effizienteste Methode zur Lösung ist das Halley-Verfahren (siehe [7]). Dabei wird das Newton-Verfahren zu Bestimmung der Nullstelle von $f(x)$ auf die Funktion $g(x) = f(x)/\sqrt{f'(x)}$ angewendet:

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)} = x_n - \frac{2f(x_n)f'(x_n)}{2(f'(x_n))^2 - f(x_n)f''(x_n)} \tag{5.6}$$

Zur Berechnung der LambertW-Funktion wird nun das Halley-Verfahren auf die Funktion $f(w) = w\,\mathrm{e}^w - z$ bzgl. w angewandt. Durch das Einsetzen von

$$f(w) = w\,\mathrm{e}^w - z \qquad f'(w) = \mathrm{e}^w(1+w) \qquad f''(w) = \mathrm{e}^w(2+w)$$

in (5.6) resultiert die bekannte rekursive Formel für $W(z)$

$$w_{j+1} = w_j - \frac{w_j\,\mathrm{e}^{w_j} - z}{e^{w_j}(w_j+1) - \frac{(w_j+1)(w_j\,\mathrm{e}^{w_j}-z)}{2w_j+2}}.$$

Damit kann nun der Simulationsschritt $v = C(u_2|u_1)$ für die Gumbel-Copula C aus (5.5) aufgelöst werden,

$$x = \left[-(1-\theta)W\left(-\frac{[cv]^{\frac{1}{1-\theta}}}{1-\theta}\right)\right]^{\theta},$$

und Rücksubstitution von (5.4) erhält man eine Lösung für u_2:

$$u_2 = \varphi^{-1}\left(\left[-(1-\theta)W\left(-\frac{[(\varphi(u_1))^{\frac{1}{\theta}-1}u_1 v]^{\frac{1}{1-\theta}}}{1-\theta}\right)\right]^{\theta} - \varphi(u_1)\right).$$

5.5 Alternatives Copula-Modell

In der Hochwasser-Studie wird außerdem ein Copula-Modell verwendet, um die Abhängigkeit der Jährlichkeiten zu beschreiben. Es wird angenommen, dass pro Gemeinde i und Bundesland j eine unbekannte Zufallsvariable S_{i,j_i} existiert, sodass die bedingten Jährlichkeiten $J_1|S_{1,j_1}, \ldots, J_d|S_{d,j_d}$ unabhängig sind. Die Abhängigkeit zwischen J_i und S_{i,j_i} wird durch eine Gumbel-Copula dargestellt.
Des Weiteren werden dort Zufallsvariablen R_1,R_2, und R_3 für die Regionen 1 (Niederösterreich, Oberösterreich, Salzburg und Wien), 2 (Burgenland, Kärnten und Steiermark) und 3 (Tirol und Vorarlberg) eingeführt. Diese sollen die Abhängigkeitsstruktur der einzelnen Bundesländer modellieren, sodass $S_{1,j_1}|R_{r_{j_1}}, \ldots, S_{d,j_d}|R_{r_{j_d}}$ unabhängig sind und die Verteilung von S_{i,j_i} und $R_{r_{j_i}}$ bzw. zwischen R_{r_j} und $R_{r_{j'}}$ wieder durch eine Gumbel-Copula beschrieben wird.
Da die Zufallsvariablen, welche die Abhängigkeiten erklären, unbekannt sind, ist es in diesem Modell nicht möglich, die Parameter der Gumbel-Copula zu schätzen. Deshalb wird angenommen, dass sich die Parameter der verschiedenen Copulas nur bzgl. des Bundeslandes und der Region ändern.

Auch in dieser Arbeit wird zum Vergleich der Modelle aus Abschnitt 5.4 eine ähnliche Copula-Methode verwendet. Als zugrundeliegende Copulafunktion wird eine Gumbel-Copula angenommen. In Abbildung 5.11 kann man erkennen, dass die Schätzer der Kendalls

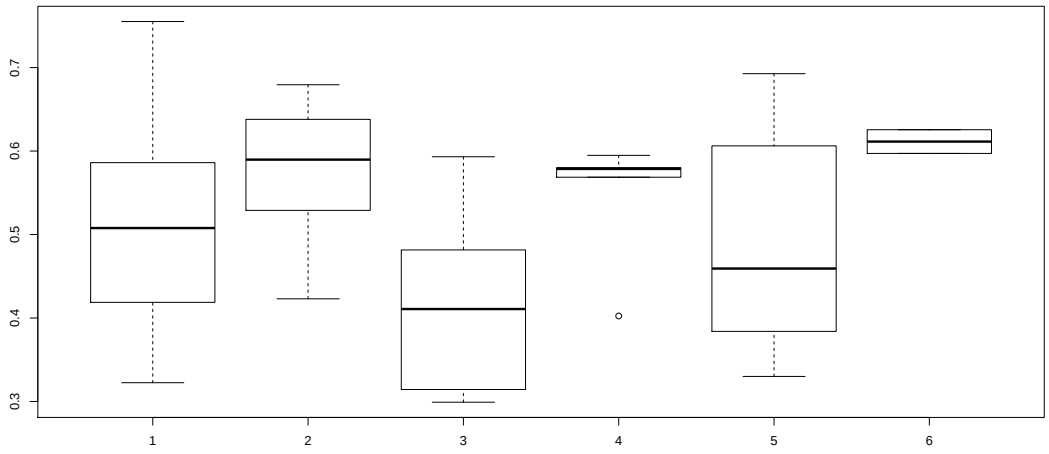


Abbildung 5.11: Boxplot der Kendalls τ der Regionen

τ der Schadensgrade zwischen den PLZ-2-Stellern der verschiedenen Regionen großteils stark variieren. Um ein Sturmereignis zu modellieren, sollten demzufolge für die Abhängigkeiten zwischen den PLZ-2-Stellern verschiedene Parameter der Gumbel-Copula verwendet werden. Dazu wurde für jede der 6 Regionen aus Abschnitt 5.3 jener PLZ-2-Steller d_i be-

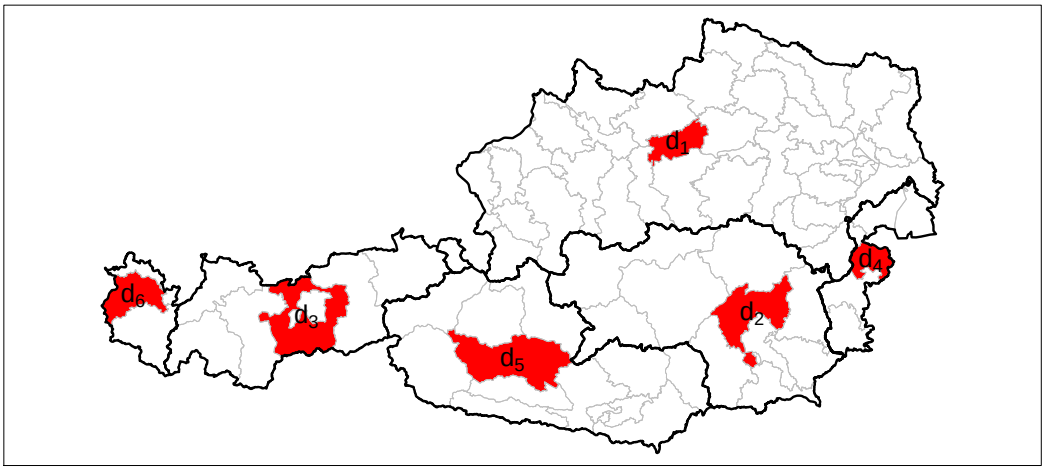


Abbildung 5.12: Mittelpunkte der Regionen

stimmt, der im Mittelpunkt der Region i liegt (vgl. Abbildung 5.12). Die Verteilungen

der Jährlichkeiten der PLZ-2-Steller in derselben Region und dem Mittelpunkt der Region werden durch bivariate Gumbel-Copulas mit separat geschätzten Parametern θ_{i,d_i} dargestellt. Somit kann auch hier der hierarchische Simulationsalgorithmus aus Abschnitt 4.5 verwendet werden.
Der zentrale Unterschied zum Copula-Modell aus Abschnitt 5.4.5 ist, dass hier keine Schranke von 20 Jahren festgelegt wird. Stattdessen sind sämtliche Jährlichkeiten $Par(1)$-verteilt.

6 Ergebnisse

Im folgenden Kapitel werden durch Anwendung der Modelle bzw Simulatiosnmethoden aus Kapitel 5 die beiden wichtigsten versicherungsmathematischen Kennzahlen diskutiert. Zum einen sollen die Einnahmen aus den Nettoprämien einer Versicherung dem erwarteten Schaden entsprechen. Deshalb wird als jährlicher Nettoprämienbedarf der Erwartungswert des Jahresgesamtschadens herangezogen. Auf der anderen Seite geben die europäische Eigenmittelvorschrifte Solvency II vor, dass eine Versicherung über ausreichend Eigenkapital verfügen muss, um auch seltene Ereignisse mit einer Jährlichkeit von 200 Jahren abdecken zu können. Als Kapitalbedarf einer Versicherung kann also das $99,5\%$-Quantil des Jahresgesamtschadens verwendet werden. Im finanzmathematischen Kontext wird dieses Quantil als $99,5\%$-VaR (Value-at-Risk) bezeichnet.
Des weiteren werden die Simulationsergebnisse mit jenen der aktuellen Studie zu Sturmschäden des Joanneum-Research [40] verglichen.

6.1 Ergebnisse für den Schaden eines Sturmereignisses

Bevor der erwartete Nettoprämien- bzw. Kapitalbedarf der Versicherung diskutiert werden, sollen erst die Ergebnisse der verschiedenen Modelle für ein einzelnes Sturmereignis analysiert werden. Diese beruhen auf Monte-Carlo-Simulationen mit jeweils 100.000 Simulationsdurchläufen.
Der mittlere Gesamtschaden der Daten liegt bei ca. 22 Mio. Euro. Da allerdings nur 36 Sturmereignisse betrachtet wurden, ist es durchaus möglich, dass noch weit höhere Gesamtschäden auftreten. Somit muss der Erwartungswert nach oben korrigiert werden.
Wie in Abschnitt 5.2 beschrieben wurde, können die Schadensgrade am besten als Zusammensetzung einer Binomial-verteilten Zufallsvariablen $X,\ X \sim Bin(1,p)$, und einer lognormalverteilten Zufallsvariable $Y,\ Y \sim LN(\mu,\sigma)$ dargestellt werden. Der Erwartungswert dieser Zufallsvariablen ist dann $(1-p)\exp\{\mu+\sigma^2/2\}$. Die Erfolgswahrscheinlichkeiten p, welche durch die Anzahl der nicht-positiven Schadensgrade gegeben sind, sind schon in Tabelle 5.2 gegeben. In Tabelle 6.1 sind nun die Parameter der Lognormalverteilung sowie die Erwartungswerte der Schäden (in Mio. Euro) für die einzelnen PLZ-2-Steller aufgelistet.

Wären die Schäden in den Gebieten unabhängig, so könnte der erwartete Gesamtschaden einfach durch Aufsummieren der einzelnen Erwartungswerte bestimmt werden und würde bei ca. 29 Mio. Euro liegen.
Dieser Erwartungswert beinhaltet aber noch nicht die Tatsache, dass der Schaden das versicherte Volumen nicht überschreiten darf. Allerdings ändert er sich kaum, wenn die Schadensgrade, welche den Anteil des Schadens am Gesamtversicherungsvolumen in Pro-

PLZ_j	μ_j	σ_j	EW	PLZ_j	μ_j	σ_j	EW	PLZ_j	μ_j	σ_j	EW
10	-6,18	1,67	0,37	44	-4,62	2,11	0,98	72	-6,11	1,72	0,02
11	-6,33	1,66	0,53	45	-4,43	2,23	1,22	73	-6,30	2,24	0,07
12	-6,61	1,98	0,44	46	-4,47	2,22	2,62	74	-6,50	2,24	0,09
20	-5,20	1,56	0,13	47	-4,31	2,35	2,22	75	-5,94	2,00	0,07
21	-6,15	1,94	0,18	48	-4,37	2,16	2,29	80	-6,30	2,11	0,58
22	-5,67	1,94	0,25	49	-4,25	2,42	2,06	81	-6,11	2,08	0,20
23	-5,25	1,74	0,43	50	-5,68	2,03	0,47	82	-6,33	2,43	0,39
24	-5,54	1,95	0,34	51	-5,12	2,47	0,85	83	-6,15	1,77	0,08
25	-5,18	1,96	0,49	52	-4,70	2,56	1,91	84	-6,40	2,38	0,27
26	-5,46	2,18	0,37	53	-5,10	2,15	0,35	85	-6,00	1,74	0,12
27	-5,13	1,76	0,21	54	-5,26	2,14	0,24	86	-6,58	2,18	0,16
28	-5,69	1,84	0,09	55	-6,81	2,02	0,04	87	-6,10	2,23	0,42
30	-5,11	1,91	0,28	56	-6,54	1,96	0,04	88	-5,50	1,48	0,03
31	-5,27	1,73	0,28	57	-6,64	2,28	0,11	89	-5,24	2,03	0,31
32	-5,12	1,96	0,30	60	-7,29	1,55	0,03	90	-6,60	1,79	0,09
33	-4,92	2,19	1,21	61	-7,68	1,59	0,01	91	-6,40	2,25	0,15
34	-5,11	1,46	0,21	62	-6,33	2,22	0,14	92	-5,71	1,80	0,03
35	-5,84	1,79	0,11	63	-6,44	2,43	0,40	93	-6,01	2,25	0,15
36	-4,91	1,57	0,10	64	-6,74	1,35	0,02	94	-6,46	2,05	0,07
37	-5,86	1,91	0,05	65	-6,98	1,72	0,01	95	-6,55	2,16	0,18
38	-5,76	2,00	0,10	66	-5,86	1,76	0,02	96	-6,16	2,09	0,04
39	-5,95	2,08	0,13	67	-6,58	1,93	0,04	97	-6,68	1,96	0,02
40	-4,85	1,55	0,84	68	-7,29	2,52	0,30	98	-6,28	1,96	0,07
41	-4,89	1,81	0,32	69	-6,83	2,28	0,12	99	-6,11	2,65	0,22
42	-5,67	2,22	0,46	70	-6,35	1,99	0,09				
43	-5,00	2,08	0,37	71	-5,90	1,94	0,06				

Tabelle 6.1: Parameter der Lognormalverteilung

mille angeben, bei der relativ hohen Grenze von 1000‰ abgeschnitten werden.
Verwendet man hingegen das Distanzmatrixmodell mit multiplikativen Abweichungen, so liegt der erwartete Gesamtschaden eines Sturmereignisses bei 38 Mio. Euro. Dieses hohe Ergebnis kann dadurch erklärt werden, dass die Jährlichkeiten und die Abweichungen in diesem Modell unabhängig sind. Somit können gleichzeitig für beiden Zufallsvariablen hohe Werte generiert werden.
Die Erwartungswerte der beiden Copula-Modelle liegen sehr nahe an jenem der unabhängigen Schadensgrade. Für das Copula-Modell aus Punkt 5.4.5 liegt der erwartete Gesamtschaden mit knapp 28 Mio. Euro sogar darunter.

Insgesamt muss jedoch bemerkt werden, dass die Erwartungswerte dieser Modelle nicht sehr aussagekräftig sind. Durch die Verwendung der Gumbel-Copula bzw. dadurch, dass die Abhängigkeit der Jährlichkeiten erst ab einer gewissen Schranke angenommen wird, sollen vor allem Sturmereignisse mit extrem hohem Schaden dargestellt werden.
Stattdessen ist es sinnvoller, als Nettoprämienbedarf die synthetisch generierten Schäden der Studie des Joanneum Research zu verwenden. Dabei wird für den Logarithmus der Schadenshöhe $S_{i,j}$ ein lineares Regressionsmodell mit Gamma-verteilten absoluten Residuen angepasst, d.h.

$$\log(S_{i,j}) \sim N(sW_{i,j} + I_j^G + \mu_I, \sigma_I^2) + \tilde{S}\Gamma(\nu, \beta).$$

Dabei bezeichnet i die Nummer des Sturmereignisses, j den PLZ-2-Steller und $W_{i,j}$ die Windgeschwindigkeit. Die Parameter s, I_j^G, μ_I und σ_I beschreiben die Steigungen des generalisierten linearen Regressionsmodells. Damit die Abweichung des logarithmierten Schadens von der Windgeschwindigkeit sowohl positiv als auch negativ sein kann, ist $\tilde{S}$ eine Zufallsvariable mit $\mathbb{P}[\tilde{S} = 1] = \mathbb{P}[\tilde{S} = -1] = 1/2$.
Die Erwartungswerte dieses Modells liegen mit 19 Mio. Euro bzw. 26 Mio. Euro, wenn nur Gesamtschäden über 2.5 Mio Euro betrachtet werden, natürlich nahe am Mittelwert der Daten. Der Effekt, dass extrem große Sturmschäden in den Daten möglicherweise noch nicht vorkommen, wird berücksichtigt, indem dieses Modell auf die Windgeschwindigkeiten großer Sturmereignisse seit 1974 angewendet wird. Dadurch erhöhen sich die Erwartungswerte auf 22 bzw. 30 Mio. Euro Schaden.

Interessant gestaltet sich auch der Vergleich der Quantile zu den Jährlichkeiten von 20, 50, 100, 150, 200 und 250 Jahren. Der bisher größte Schaden von 175 Mio. Euro (auf das Jahr 2009 normiert) wurde durch den Sturm Kyrill verursacht. Die Quantile der Daten (siehe Tabelle 6.2) können diese Grenze überschreiten. Dennoch ist es möglich, dass höhere Schäden als die bisher beobachteten auftreten. Die Simulationsergebnisse sollten demnach größer sein.

In Tabelle 6.3 sind diese Quantile für die verschiedenen Modelle aus Kaptiel 5 angegeben. Auch hier sind die Werte des Distanzmatrixmodells mit multiplikativen Abweichungen aus Punkt 5.4.4 größer als jene der anderen beiden Modelle. Wie man erkennen kann, gibt es einen Schnittpunkt der Verteilungsfunktionen des Gesamtschadens aus Punkt 5.4.5 und 5.5. Dieser Effekt kann dadurch erklärt werden, dass im Distanzmatrixmodell mit Copula-Funktionen kleinere Gebiete voneinander abhängig sind als im alternativen Copula-Modell.

Jährlichkeit	20	50	100	150	200	250
Quantil	0,95	0,98	0,99	0,9933	0,995	0,996
Daten	77	132	153	161	164	166

Tabelle 6.2: Quantile des Gesamtschadens der gegebenen Daten

Dort werden durch die Gumbel-Copula in einer gesamten Region relativ hohe Schäden erzeugt, wenn im Mittelpunkt der Region ein Schaden mit großer Jährlichkeit auftritt.

Jährlichkeit	20	50	100	150	200	250
Quantil	0,95	0,98	0,99	0,9933	0,995	0,996
Mult. Abw.	127	270	446	603	741	846
Copula (Nachbarn)	105	208	329	419	503	583
Copula (Regionen)	88	186	319	433	519	618

Tabelle 6.3: Quantile des Gesamtschadens eines Sturmereignisses für die verschiedenen Modelle aus Kapitel 5

Die Wahl einer Schranke von 1000‰, d.h. das gesamte versicherte Volumen im PLZ-2-Steller wird zerstört, hat auch auf diese Ergebnisse keine Auswirkungen. Allerdings ist davon auszugehen, dass der Endpunkt der Verteilung der Schadensgrade nicht so groß ist. Laut [14] würde bei einer Windgeschwindigkeit von 135km/h ein Schaden an Massivbauten in der Höhe von ca. 1‰ des Neuwertes auftreten. Dieser Wert bezieht sich allerdings auf die spezielle Winderscheinung von Tornados. Generell sind solche Einschätzungen stark von dem Standort des Gebäudes und seiner Bauweise abhängig. In [23] wird außerdem darauf hingewiesen, dass auch die Reparaturkosten geografisch große Unterschiede aufweisen. Eine starke Beschädigung des Daches würde in Europa Reparaturkosten von 8‰ bis 30‰ verursachen. Um solche Schranken in die Analyse einfließen zu lassen, müssten detaillierte Daten über den Gebäudebestand und die Art der Versicherung gegeben sein.
Dennoch wurden die Auswirkungen einer Begrenzung des maximalen Schadensgrad überprüft. In Tabelle 6.4 sind die Quantile der verschiedenen Modelle für die Schranken von 10‰ und 100‰ angegeben. Außerdem wurden die Gesamtschäden der Modelle auch unter

Jährlichkeit	20	50	100	150	200	250
Quantil	0,95	0,98	0,99	0,9933	0,99.5	0,996
Mult. Abw., Schranke: 10‰	127	259	400	518	613	685
Mult. Abw., Schranke: 100‰	127	270	446	603	741	846
Copula (Nachbarn), Schranke: 10‰	105	205	314	394	458	528
Copula (Nachbarn), Schranke: 100‰	105	207	328	419	502	582
Copula (Regionen), Schranke: 10‰	88	184	310	416	498	576
Copula (Regionen), Schranke: 100‰	88	186	319	433	519	618

Tabelle 6.4: Quantile der Gesamtschäden mit beschränkten Schadensgraden

der Annahme einer Schranke von 300‰ simuliert. Da aber nur 18 bzw. 4 Schadensgrade über 300‰ im Modell der multiplikativen Abweichung bzw. in den beiden Copula-Modellen auftreten, ändern sich die Werte gegenüber einer Schranke von 100‰ nicht. Bei einer Wahl von 10‰ gibt es hingegen 2758 bzw. ca. 1400 Überschreitungen.
Die Einführung von Schranken hat in den Copula-Modellen keine großen Auswirkungen. Im ursprünglichen Modell der multiplikativen Abweichungen sind aber für die Wahl einer kleinen Schranke starke Änderungen am rechten Rand der Verteilung des Gesamtschadens erkennbar. Auch der Erwartungwert dieses Modells beträgt bei einer Schranke von 1‰ nur mehr 32 Mio. Euro ist ist somit um 4 Mio. Euro höher als die Erwartungswerte der anderen Modelle ohne Schranken. Durch die Wahl einer passenden Obergrenze könnte der Effekt ausgeglichen werden, dass die Höhe der Jährlichkeiten und die multiplikativen Abweichungen als unabhängig angenommen werden. Dazu müsste aber eine ausführliche Analyse der möglichen Maximal-Schäden durchgeführt werden.

Um diese Auswirkung der Abhängigkeiten der PLZ-2-Steller zu verdeutlichen, sind in Tabelle 6.5 die Quantile der Gesamtschäden für einfache Abhängigkeitsstrukturen angegeben. Die Quantile der unabhängigen PLZ-2-Steller bewegen sich hier ungefähr im Rahmen

Jährlichkeit	20	50	100	150	200	250
Quantil	0,95	0,98	0,99	0,9933	0,99.5	0,996
Co-Monotone PLZ-2-Steller	97	239	429	586	739	875
Co-Monotone Regionen	95	216	372	501	633	762
Unabhängige PLZ-2-Steller	73	122	183	229	279	322

Tabelle 6.5: Quantile der Gesamtschäden eines Sturmeregnisses mit einer einfachen Abhängigkeitsstruktur zwischen den PLZ-2-Stellern

jener des Regressionsmodells der Sturmstudie des Joanneum Research. Erhöht man den Grad der Abhängigkeit, indem man annimmt, dass die Jährlichkeit für alle PLZ-2-Steller dieselbe ist (siehe Co-Monotone Regionen in Tabelle 6.5), so steigt auch die Höhe extremer Sturmschäden stark an. Dies liegt vor allem am sogenannten Diversifikationseffekt. Bei unabhängigen PLZ-2-Stellern wird das Risiko eines hohen Gesamtschadens dadurch verringert, dass sich hohe Schäden in einem PLZ-2-Steller durch relativ kleine Schäden in den anderen PLZ-2-Stellern ausgleichen. Dieser Effekt wird auch übertrieben dargestellt (siehe Co-Monotone PLZ-2-Steller in Tabelle 6.5). Hier wird für alle PLZ-2-Steller diesselbe Jährlichkeit angenommen. Man kann auch deutlich erkennen, dass sich diese Erweiterung der Co-Monotonie der PLZ-2-Steller vor allem auf die Quantile der hohen Jährlichkeiten auswirkt.

Die Bedeutung der Abhängigkeitsstruktur kann außerdem durch eine andere Copula-Funktion überprüft werden. Dazu wurde die Simulationen der beiden Copula-Modelle auch mit einer Clayton-Copula, welche vor allem eine Abhängigkeit der kleinen Schäden beschreibt, durchgeführt. Die hohen Quantile des Distanzmatrix-Modells ändern sich bei

dieser Wahl kaum. Das kann dadurch erklärt werden, dass Schäden bis zu einer Jährlichkeit von 20 unabhängig sind und erst ab dieser Grenze eine starke Abhängigkeitsstruktur aufweisen. Hohe Schäden sind wiederum beinahe unabhängig.
Verwendet man auch für die Beschreibung der Abhängigkeiten innerhalb einer Region

Jährlichkeit	20	50	100	150	200	250
Quantil	0,95	0,98	0,99	0,9933	0,99.5	0,996
Copula (Nachbarn)	106	264	384	465	521	561
Copula (Regionen)	87	140	207	261	302	344

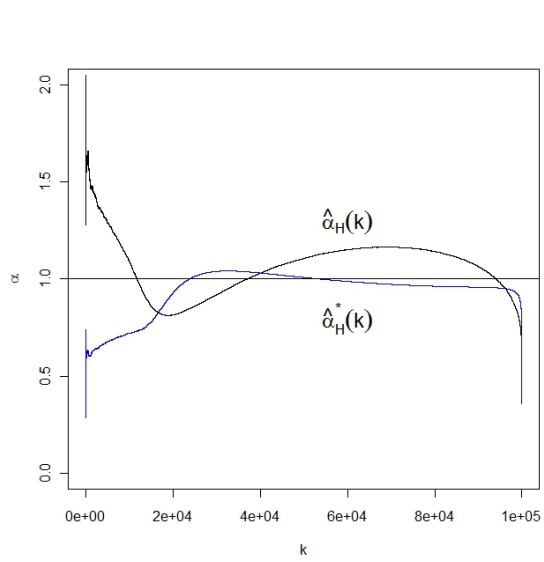
Tabelle 6.6: Quantile der Gesamtschäden eines Sturmeregnisses bei Verwendung einer Clayton-Copula

eine Clayton-Copula, ändern sich die Ergebnisse aus Tabelle 6.3 stark. Bei der Wahl einer Clayton-Copula weisen vor allem kleine Schäden eine hohe Abhängigkeit auf. Dadurch können in der Simulation große Gebiete mit sehr geringem Schaden entstehen. Die hohen Schäden treten allerdings eher vereinzelt auf.
Diese Auswirkungen der unterschiedlichen Copula-Funktionen sind eher am Verteilungsrand des Gesamtschadens bemerkbar. Die Erwartungswerte der Simulationen bleiben durch die Verwendung einer anderen Copula beinahe unverändert.

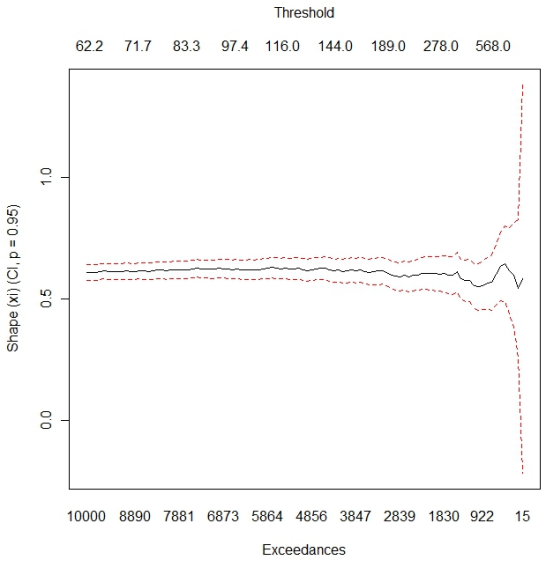
Neben dem loglinearen Modell zur Generierung von synthetischen Sturmschäden wurde von dem Joanneum Research auch eine Extremwertanalyse des Gesamtschadens durchgeführt. Dabei stellte sich heraus, dass dieser aus dem maximalen Anziehunggebiet einer Frechet-Verteilung stammt. Mit einem Tail-Index von $\alpha = 1,25$ bis $1,38$ pro Schätzmethode besitzt der Gesamtschaden einen endlichen Erwartungswert.
Durch Anpassung einer Generalisierten Pareto-Verteilung auf die 10 bis 16 größten Datenpunkte sind auch Schätzer für die Quantile des Gesamtschadens mittels Extremwertanalyse verfügbar. Diese ähneln jenen der beiden Copula-Modelle aus Kapitel 5. Für den Value-at Risk schwanken die Schätzungen der Studie zwischen 400 und 500 Mio. Euro pro Sturmereignis und liegen somit nahe jenen aus Tabelle 6.3.

6.2 Extremwertanalyse der Simulationsergebnisse

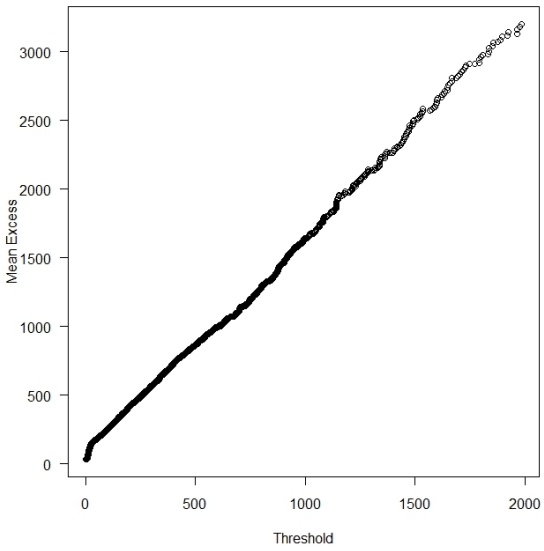
Der Tail-Index der Modelle aus [40] beträgt 1,25 bis 1,38 bei Anpassung einer Generalisierten Pareto-Verteilung und 1,83 für das log-lineare Modell. In allen Fällen liegt hier also ein endlicher Erwartungswert vor, womit die Versicherbarkeit der zugrundeliegenden Risiken gegeben ist.
Um diese Werte vergleichen zu können, werden im Folgenden auch die Simulationsergebnisse der verschiedenen Modelle einer Extremwertanalyse unterzogen. Dabei soll vor allem der Formparameter $\gamma = 1/\alpha$ bestimmt werden, um auch hier eine Aussage über den Erwartungswert treffen zu können.



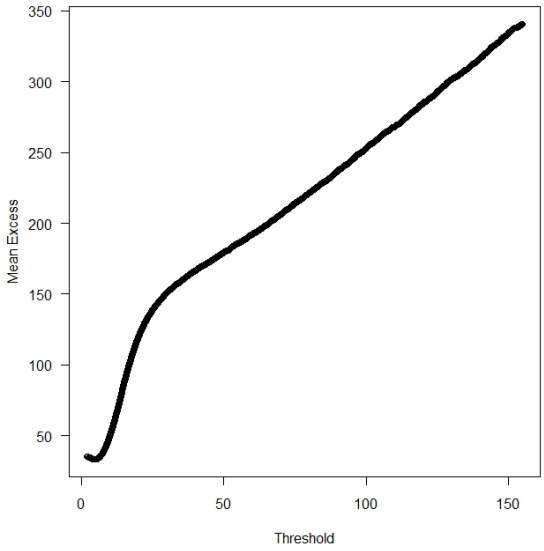
(a) Hill-Plot



(b) Maximum-Likelihood



(c) Mean-Excess-Plot



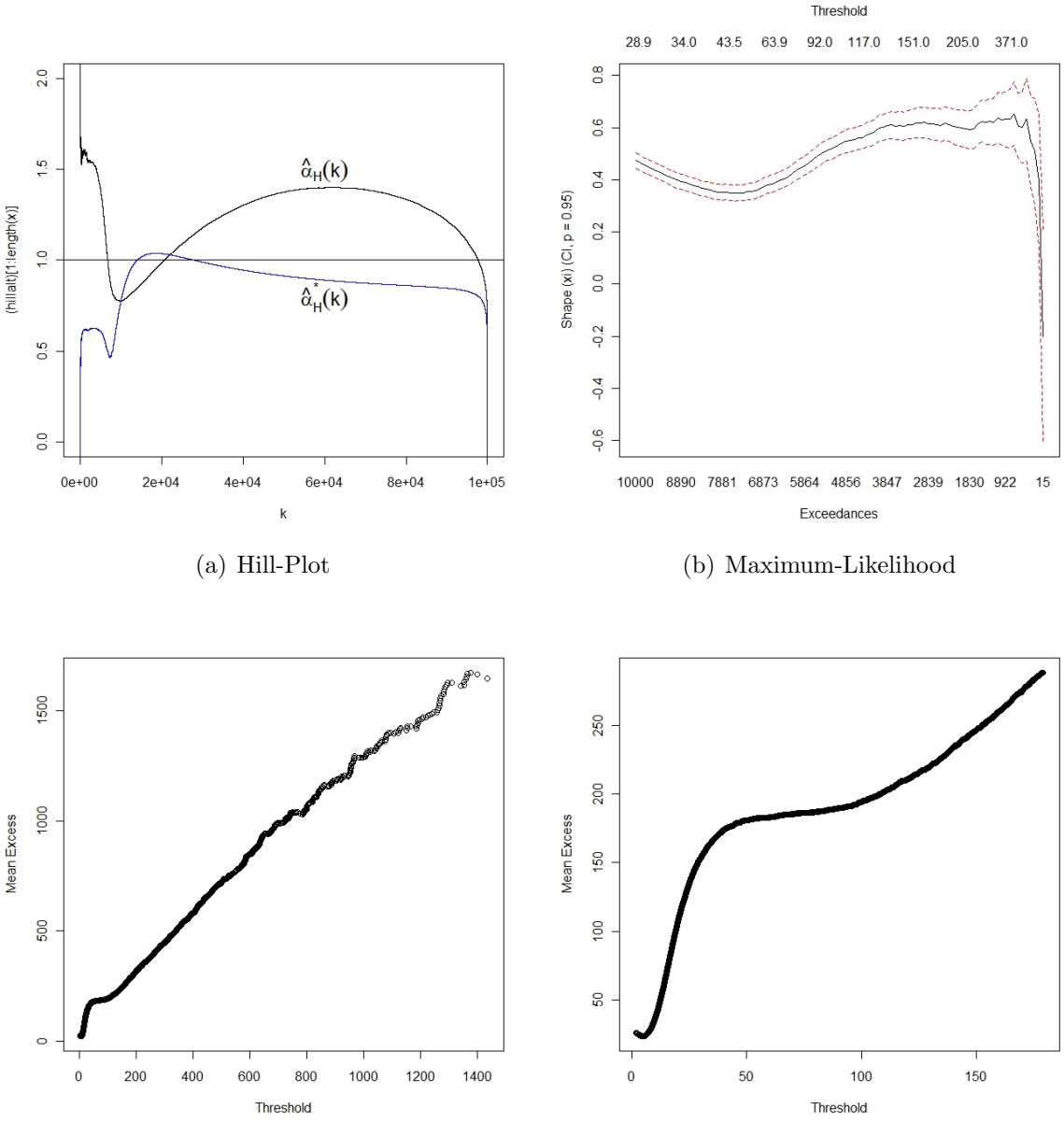
(d) Mean-Excess-Plot (linker Rand)

Abbildung 6.1: Graphische Methoden für Modell 1

In Abbildung 6.1 sind verschiedene Plots zur Analyse des Modells mit multiplikativen Abweichungen dargestellt. In Abbildung 6.1(b) sind die Schätzer des Parameter γ von 100 Maximum-Likelihood-Schätzungen sowie deren Konfidenzintervalle eingezeichnet. Für eine Schranke u von 2, 50, 100 und 150 Mio. Euro beträgt der geschätzte Wert für γ in diesem Modell 0,55, 0,59, 0,62 und 0,61. Eine Maximum-Likelihood-Schätzung würde in diesem Modell also einen Tail-Index von 1,6 bis 1,8 ergeben. Außerdem scheint diese Methode

auch zu konvergieren. Der Hill-Plot in Abbildung 6.1(a) ist jedoch von diesen Werten weit entfernt. Während der Hillschätzer $\hat{\alpha}$ sehr selten auf einer Höhe bleibt, pendelt sich $\hat{\alpha}^*$ ca. bei 1 ein. In diesem Fall kann also nicht mit Sicherheit ausgesagt werden, ob ein endlicher Erwartungswert vorliegt.

Der Grund für dieses Verhalten könnte in 6.1(d) gegeben sein. Während der Mean-Excess-



(a) Hill-Plot

(b) Maximum-Likelihood

(c) Mean-Excess-Plot

(d) Mean-Excess-Plot (linker Rand)

Abbildung 6.2: Graphische Methoden für Modell 2

Plot wirklich die Form einer geraden Linie annimmt (siehe 6.1(c)), schwankt er am Anfang

stark. Vor einem Gesamtschaden von ca. 30 Mio. Euro scheint er sich schon einer Gerade mit relativ starker Steigung anzunähern, danach wird diese aber schwächer. Dieses Verhalten kann durch die plötzliche Veränderung bei einer Jährlichkeit der Abhängigkeitesstruktur von 20 Jahren erklärt werden.
Diese Eigenschaft ist in den Plots der Ergebnisse des Nachbarschafts-Modells mit Copula-Funktionen in Abbildung 6.2 noch ausgeprägter zu erkennen. Hier dauert es auch relativ lang, bis sowohl die beiden Hill-Schätzer als auch die Maximum-Likelihood-Schätzer konvergieren. Wird die Schranke u groß genug gewählt, sodass der Strukturbruch der Daten nicht mehr ausschlaggebend ist, liegen der Hill- und der Maximum-Likelihood-Schätzer für γ bei ca. 0,6, d.h. der Tail-Index α lässt mit einem Wert von ca. $1,6$ auf einen endlichen Erwartungswert schließen. Der Schätzer $\hat{\alpha}^+$ nimmt aber andere Wert an. Dieser Schätzer konvergiert im Allgemeinen aber langsamer als der ursprüngliche Hill-Schätzer.
Auch für das Modell, das eine Copula-Funktion zur Beschreibung der Abhängigkeiten innerhalb der Regionen verwendet, können die unterschiedlichen Schätzer des Tail-Index α keine sichere Aussage über einen endlichen Erwartungswert treffen. Sowohl der Hill-Schätzer als auch der Schätzer von Dekkers, Einmahl und de Haan schwanken um den Wert von 1 (siehe Abbildung 6.3(a)). Der Maximum-Likelihood-Schätzer (Abbildung 6.3(b)) für γ nimmt hier im Vergleich zu den anderen beiden Modellen leicht größere Werte an, die aber unter 1 liegen. Somit würde aufgrund dieser Schätzmethode ein endlicher Erwartungswert vorliegen. Allerdings kann man in den Abbildungen 6.3(d) und 6.3(c) erkennen, dass der Mean-Excess-Plot eine leichte Krümmung aufweist.

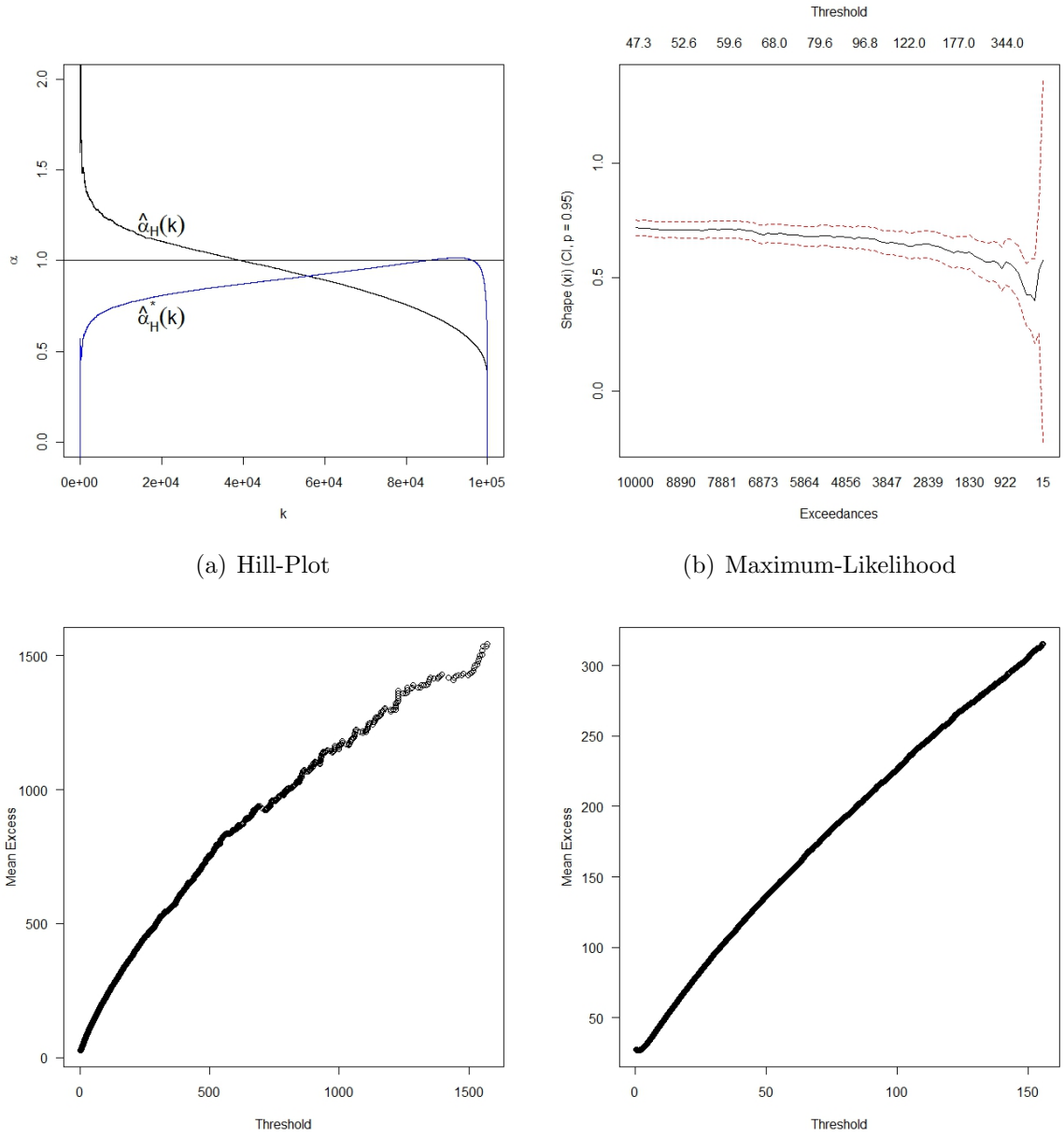
Insgesamt kann für die verschiedenen Modelle der Tail-Index α nicht mit Sicherheit bestimmt werden. Dies liegt vor allem an der Änderung der Verteilung in den Distanzmatrixmodellen und an der Abhängigkeit zwischen den PLZ-2-Stellern im Copula-Modell.

6.3 Ergebnisse für den Jahresgesamtschaden

Um den benötigten Nettoprämien- und Kapitalbedarf für Versicherungen zu bestimmen, wird nun in Folge der Jahresgesamtschaden abgeschätzt. In Abschnitt 5.1 wurde festgestellt, dass ein positiver Trend in der erwarteten Anzahl von Sturmereignissen pro Jahr vorliegt. Da dieser Trend allerdings nicht sehr stark ist und um die Ergebnisse des Jahresgesamtschadens mit jenen des Joanneum Research vergleichbar zu machen, wurde er nicht berücksichtigt.
Pro Jahr treten im Durchschnitt 3 Sturmereignisse mit einem Gesamtschaden von über 2 Mio. Euro auf. Die Varianz dieser Anzahl liegt bei $2,18$. Wenn die Anzahl der Stürme pro Jahr, $X_i,\ i = 1, \ldots, n$, unabhängig und Poisson-verteilt mit Parameter λ ist, sollte

$$D = \frac{\sum(X_i - \bar{X})^2)}{\bar{X}} \sim \chi^2_{n-1}$$

gelten, wobei $\bar{X}$ die durchschnittliche Anzahl der Sturmereignisse pro Jahr bezeichnet. Für die hier verwendeten Daten ist $D = 8$ und somit kann die Nullhypothese einer Poisson-Verteilung mit Parameter $\lambda = 3$ nicht verworfen werden.



(a) Hill-Plot

(b) Maximum-Likelihood

(c) Mean-Excess-Plot

(d) Mean-Excess-Plot (linker Rand)

Abbildung 6.3: Graphische Methoden für Modell 3

Im Allgemeinen sind die Charakteristika von Winter- und Sommerstürmen unterschiedlich. Da allerdings bei 36 Sturmereignissen in 12 Jahren nicht genügend Daten vorliegen, um eine saisonale Modellierung der Schäden vorzunehmen, werden in dieser Analyse Winter- und Sommerstürme gleich behandelt. Deshalb wird zur Ermittlung der Jahresgesamtschäden pro Simulationsdurchlauf eine Poisson-verteilte Zufallsvariable N mit Parameter $\lambda = 3$ generiert, welche die Anzahl der Sturmereignisse im gerade simulierten Jahr beschreibt.

Die Schäden dieser Sturmereignisse werden nach den Modellen in Kapitel 5 erzeugt, eine Analyse findet sich in Abschnitt 6.1.

Dieser Ansatz der Poisson-verteilten Anzahl an Sturmereignissen pro Jahr wurde auch in der Studie des Joanneum Research verwendet, um den Erwartungswert bzw. die Quantile des Jahresgesamtschadens zu bestimmen.
Die Erwartungswerte des loglinearen Modells liegen je nach Schätz- und Extrapolationsmethode bei 55 bis 91 Mio Euro. Die Ergebnisse dieser Arbeit sind teilweise mit diesen Mittelwerten vergleichbar. Während das Distanzmatrixmodell mit multiplikativen Abweichungen, dessen Mittelwert bei 112 Mio Euro liegt, diese Ergebnissen übersteigt, ergeben auch die beiden Copula-Modelle dieser Arbeit einen durchschnittlichen Jahresgesamtschaden von 86 bzw. 90 Mio. Euro.

Um die Quantile der Jahresgesamtschäden im generalisierten loglinearen Modell zu bestimmen, werden die Windgeschwindigkeiten in [40] extrapoliert. Diese Methode führt auf ein $99,5\%$-Quantil von ca. 800 Mio. Euro. Auch die Extremwertanalyse dieser Studie ist für Jahresgesamtschäden gegeben. Hier schwanken die Ergebnisse des $99,5\%$-VaR zwischen rund 1000 und 1250 Mio Euro.

Jährlichkeit	20	50	100	150	200	250
Quantil	0,95	0,98	0,99	0,9933	0,99.5	0,996
Mult. Abw.	378	680	1067	1344	1619	1856
Copula (Nachbarn)	296	505	771	966	1149	1305
Copula (Regionen)	280	529	846	1085	1300	1578

Tabelle 6.7: Quantile des Jahresgesamtschadens für die verschiedenen Modelle aus Kapitel 5

In Tabelle 6.7 sind die Quantile der Simulationsergebnisse des Jahresgesamtschadens dieser Arbeit aufgelistet. Auch hier ähneln vor allem die Copula-Modelle am ehesten jenen der Sturmstudie des Joanneum Research. Allerdings ist zu bemerken, dass die Schätzer des $99,5\%$-VaR für das Modell der multiplikativen Abweichungen sowie für das Copula-Modell der Regionen größer sind als die Ergebnisse aus [40]. Das Distanzmatrixmodell mit Copula-Funktionen liegt aber hier im Bereich der geschätzten Value-at-Risk der Extremwertanalyse des Jahresgesamtschadens des Joanneum Research.

Zum Vergleich sind auch in diesem Abschnitt die Simulationsergebnisse des Jahresgesamtschadens für einfache Abhängigkeitsstrukturen angegeben. Obwohl hier das Zusammenfassen der Schäden mehrerer Sturmereignisse zum Jahresgesamtschaden einen gewissen Diversifikationseffekt bietet, kann dieser die Auswirkungen der Abhängigkeit zwischen den PLZ-2-Stellern nicht mindern. Wie schon im vorherigen Abschnitt beschrieben wurde, nimmt auch der Jahresgesamtschaden mit dem Grad der Abhängigkeit zwischen den PLZ-2-Stellern zu.

Jährlichkeit	20	50	100	150	200	250
Quantil	0,95	0,98	0,99	0,9933	0,995	0,996
Co-Monotone PLZ-2-Steller	317	644	1055	1434	1762	2026
Co-Monotone Regionen	300	566	911	1176	1411	1640
Unabhängige PLZ-2-Steller	229	334	463	563	649	720

Tabelle 6.8: Quantile des Jahresgesamtschadens bei einfachen Abhängigkeitsstrukturen zwischen den PLZ-2-Stellern

6.4 Fazit

Im Rahmen dieser Arbeit konnte die zentrale Komponente des Distanzmatrix-Modells nach Url für Versicherungsschäden durch extreme Hochwasser, die Beschreibung der Abhängigkeit einzelner Gebiete durch eine Nachbarschaftsrelation, auf jene extremer Sturmereignisse angepasst werden. Die einfache Version der multiplikativen Abweichungen kann aber nur bedingt verwendet werden. Hier besteht Verbesserungsbedarf durch das Finden einer Abhängigkeitsstruktur zwischen den Abweichungen und der Höhe der betroffenen Jährlichkeiten. Die Erweiterung des Modells durch eine Copula-Funktion führte hingegen zu Ergebnissen, welche auch durch die Studie der Forschungsgesellschaft Joanneum Research erzielt wurden.
Sowohl die Ablehnung einer Generalisierten Pareto-Verteilung für die Schadensgrade in den einzelnen Gebieten als auch der Umstand, dass die Simulationen der Modelle dieser Arbeit höhere Schäden als jene der Sturmstudie der Joanneum Research hervorbringen, zeigen, dass die extreme Höhe der gegebenen Gesamtschäden auf das vermehrte Auftreten hoher Jährlichkeiten in benachbarten PLZ-2-Steller zurückzuführen ist.
Im Gegensatz zu dem Distanzmatrix-Modell der Hochwasser-Studie konnte die Windgeschwindigkeit in dieser Arbeit nicht zur Beschreibung der Abhängigkeitsstruktur der Jährlichkeiten der Schadensgrade verwendet werden. Dies ist ein Nachteil, weil die Jährlichkeiten der Sturmereignisse somit nicht mit den allgemein verwendeten Jährlichkeiten der Windgeschwindigkeiten verglichen werden können. Es muss jedoch betont werden, dass solche Einteilungen aufgrund der örtlichen Variabilität von Windgeschwindigkeiten generell mit Vorsicht zu genießen sind.
Diese können nur dann in das Distanzmatrix-Modell miteinbezogen werden, wenn sowohl die aufgetretenen Schäden als auch Windrichtungen und -geschwindigkeiten in sehr hoher räumlicher und zeitlicher Auflösung gegeben sind. In diesem Fall könnten, wenn auch detaillierte Daten über den Gebäudebestand und das Versicherunngsvolumen gegeben sind, geeignete Schranke für die Schadensgrade der einzelnen Modelle gefunden werden.

Literaturverzeichnis

[1] Bivand R., Pebesma E., Gómez-Rubio V. (2008) *Applied Spatial Data Analysis with R*, Springer ISBN 978-0-387-78170-9.

[2] Austrian Wind Potential Analyses - Windatlas und Windpotentialstudie Österreich (2009-2011), `http://www.windatlas.at/index.html` (aufgerufen am: 31.08.2011)

[3] Brodin E., Rootzen H. (2009) *Univariate and bivariate GPD methods for predicting extreme wind storm losses*, Insurance: Mathematics and Economics, Vol. 44, Nr. 3, S. 345 – 356.

[4] Burroughs S., Tebbens S. (2000) *Upper-Truncated Power Law Distributions*, Fractals, Vol. 9, Nr. 2, S. 209-222.

[5] Charpentier A., Sibaï (2009) *Dynamic flood modeling: combining Hurst and Gumbel's approach* Environmetrics, Vol. 20, Issue 1, S. 32-52.

[6] Coles S. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics, Springer-Verlag London, Ltd., London, ISBN 1-85233-459-2.

[7] Corless R. M., Gonnet G. H., Hare D. E. G., Jeffrey D. J., Knuth D. E. (1996) *On the LambertW function*, Advances in Computational Mathematics, Vol. 5, No. 1, S. 329-359.

[8] Couturier D.-L., Victoria-Feser M.-P. (2009) *Zero-Inflated Truncated Generalized Pareto Distribution for the Analysis of Radio Audience Data*, Ann. Appl. Stat., Vol.4, No. 4, S. 1824-1846.

[9] Daley D.J., Vere-Jones, D. (2003) *An Introduction to the theory of point processes, Volume I: Elementary Theorie and Methods*, Second Edition, Springer ISBN 0-387-95541-0.

[10] Daley, D.H., Vere-Jones, D. (2007) *An Introduction to the theory of point processes, Volume II: General theory and structure*, Springer ISBN 0-387-21337-6.

[11] Dekkers A. L. M., Einmahl J. H. J., de Haan L. (1989) *A Moment Estimator for die Index of an Extreme-Value Distribution* The Annals of Statistics, Vol. 17, No. 4, S. 1833-1855.

[12] Donat M. G., Leckebusch G. C., Pinto J. G., Ulbrich U. (2010) *European storminess and associated circulation weather tyes: future changes deduced from a multi-model ensemble of GCM simulations* Climate Research, Vol. 42, No. 1, S. 27-43.

[13] Dorland, C., Tol R. S. J., Palutikof J. P. (1999) *Vulnerability of the Netherlands and Northwest Europe to Storm Damage under Climate Change* Springer Netherlands, Climatic Change, Vol. 43, Issue 3, S. 513-535.

[14] Dotzek N., Berz G., Rauch E., Peterson R. E. (2000) *Die Bedeutung von Johannes P. Letzmanns „Richtlinien zur Erforschung von Tromben, Tornados, Wasserhosen und Kleintromben"für die heutige Tornadoforschung*, Meteorologische Zeitschrift, Vol. 9, No. 3, S. 165 – 174.

[15] Drees H., de Haan L., Resnick S. (2000) *How to Make a Hill Plot* The Annals of Statistics, Vol. 28, No. 1, S. 254-274.

[16] Elandt R.C. (1961) *The Folded Normal Distribution: Two Methods of Estimating Parameters from Moments* Technometrics, Vol. 3, No. 4, S. 551-562.

[17] Embrechts P., Klüppelberg C., Mikosch T. (1997) *Modelling Extremal Events for Insurance and Finance*, Springer, ISBN 3-540-60931-8.

[18] Etienne C., Lehmann A., Goyette S. (2010) *Spatial Predictions of Extreme Wind Speeds over Switzerland Using Generalized Additive Models* Journal of Applied Meteorology and Climatology, Vol. 49, Issue 9, S. 1956–1970

[19] Frahm G., Junker M., Schmidt R. (2005) *Estimating the tail-dependence coefficient: Properties and pitfalls* Insurance: Mathematics and Economics, Vol. 37, No. 1, S. 80-100.

[20] Fusai G., Roncoroni A. (2008) *Implementing Models in Quantitative Finance: Methods and Cases*, Springer, ISBN 987-3-540-22348-1.

[21] Haan L., Ferreira A. (2006) *Extreme value theory: an introduction*, Springer series in operations research and financial engineering, Springer, ISBN 978-0-387-23946-0.

[22] Häckel H. (2008) *Meteorologie*, 6. Auflage, UTB, Stuttgart, ISBN 978-3-8252-1338-1.

[23] Heneka P., Ruck B. (2008) *A damage model for the assessment of storm damage to buildings* Engineering Structures, Vol. 30, No. 12, S 3603-3609.

[24] Holden L., Sannan S., Bungum H. (2003) *A stochastic marked point process model for earthquakes* Natural Hazards and Earth System Sciences, Vol. 3, No. 1/2, S. 95-101.

[25] Holmes J. D. (2007) *Wind loading of structures* Taylor & Francis, ISBN 978-0-415-40946-9.

[26] Klawa M., Ulbrich U. (2003) *A model for the estimation of storm losses and the identification of severe winter storms in Germany* Natural Hazards and Earth System Science, Vol. 3, No. 6, S. 725–732.

[27] Knippertz P., Ulbrich U., Speth P. (2000) *Changing cyclones and surface wind speeds over the North Atlantic and Europe in a transient GHG experiment* Climate Research, Vol. 15, No. 2, S. 109-122.

[28] Kompetenzzentrum für lokale Unwetter in Deutschland, Österreich und der Schweiz, `http://www.tordach.org/at/` (aufgerufen am 1.9.2011)

[29] König D., Schmidt V. (1992) *Zufällige Punktprozesse* B.G. Teubner, Stuttgart, ISBN 3-519-02733-X.

[30] Leckebusch G. C., Koffi B., Ulbrich U., Pinto J. G., Spangehl T., Zacharias S. (2006) *Analysis of frequency and intensity of European winter storm events from a multi-model perspective, at synoptic and regional scales*, Climate Research, Vol. 31, No. 1, S. 59-74.

[31] Leone F. C., Nelson L. S., Nottingham R. B. (1961) *The Folded Normal Distribution* Technometrics, Vol. 3, No. 4, S. 543-550.

[32] Lescourret L., Robert C. (2006) *Extreme dependence of multivariate catastrophic losses* Scandinavian Actualial Journal, 2006:4, S. 203-225.

[33] McNeil A.J. (1997) *Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory* Astin Bulletin, Vol 27, No. 1, S. 117-137.

[34] McNeil A.J., Frey R., Embrechts P. (2005) *Quantitative risk management: concepts, techniques and tools* Princeton series in finance, Princeton University Press, ISBN 978-0691-12255-7.

[35] McNeil A.J., Nešlehová J. (2009) *Multivariate Archimedian Copulas, d-monotone functions and ℓ_1-norm symmetric distributions*, Annals of Statistics, Vol. 37, Nr. 5b, S. 3059-3097 (preprint version, `http://www.ma.hw.ac.uk/~mcneil/ftp/McNeil-Neslehova-07.pdf`, aufgerufen am 2.8.2011).

[36] Natural Hazard Overview Risk Assessment Austria, `http://www.hora.gv.at/` (aufgerufen am 31.08.2011).

[37] Nelsen R. B. (2006) *An Introduction to Copulas*, Second edition, Springer Series in Statistics. Springer, New York, ISBN 978-0387-28659-4.

[38] Pinto J. G., Fröhlich E. L., Leckebusch G. C., Ulbrich U. (2007) *Changing European storm loss potentials under modified climate conditions according to ensemble simulations of the ECHAM5/MPI-OM1 GCM*, Natural Hazards and Earth System Sciences, Vol. 7, No. 1, S. 165–175.

[39] Prettenthaler F., Albrecher H. (2009) *Hochwasser und dessen Versicherung in Österreich*, Studien zum Klimawandel Bd. III, S. 91–103.

[40] Prettenthaler F., Albrecher H., Köberl J., Kortschak D. *Risk and insurability of storm damages to residential buildings in Austria*, Preprint, zur Publikation eingereicht.

[41] Probst O., Cárdenas D. (2010) *State of the Art and Trends in Wind Resource Assessment*, Energies, Vol. 3, No. 6, S. 1087–1141.

[42] Reiss T., Thomas M. (2001) *Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology and Other Fields*, Birkhauser, ISBN 3-7643-6487-4.

[43] Resnick S. (2007) *Heavy-tail phenomena: probabilistic and statistical modeling* Springer, ISBN 978-0387-24272-9.

[44] Rinne H. (2008) *Taschenbuch der Statistik*, Wissenschaftlicher Verlag Harri Deutsch Gmbh, ISBN 978-3817-11827-4.

[45] Rootzén H., Tajvidi N. (1997) *Extreme Value Statistics and Wind Storm Losses: A Case Study* Scandinavian Actuarial Journal, Vol. 1997, Issue 1, S. 70-94.

[46] Rootzén H., Tajvidi N. (2001) *Can Losses Caused by Wind Storms be Predicted from Meteorological Observations?* Scandinavian Actuarial Journal, Vol. 2001, Issue 2, S. 162-175.

[47] Salvadori G., De Michele C., Kottegoda N. T., Rosso R. (2007) *Extremes in Nature, An Approach Using Copulas*, Springer ISBN 978-1-4020-4414-4.

[48] Schmidt R., Stadtmüller U. (2006) *Nonparametric estimation of tail dependence* The Scandinavian Journal of Statistics, Vol. 33, No. 2, S. 307-335.

[49] Schönwiese C.-D. (2008) *Klimatologie*, 3. Auflage, UTB, Stuttgart, ISBN 978-3-8252-1793-8.

[50] Simiu E., Heckert N. A., Filliben J. J., Johnson S. K. (2001) *Extreme wind load estimates based on the Gumbel distribution of dynamic pressures: an assessment* Structural Safety, Vol. 23, Issue 3, S. 221–229

[51] SwissRe (2006) *The effects of climate change: Storm damage in Europe on the rise* `http://media.swissre.com/documents/storm_damage_in_europe_on_the_rise_en.pdf` (aufgerufen am 12.7.2011).

[52] SwissRe (2008) *Natural catastrophes and man-made disasters in 2007: high losses in Europe* SwissRe Sigma Report No. 1/2008 `http://media.swissre.com/documents/sigma1_2008_en.pdf` (aufgerufen am 12.7.2011).

[53] Url T. (2008) *Wahrscheinlichkeits-Überschreitungskurven für Hochwasserkatastrophen in Österreich*, WIFO-Studie. http://www.wifo.ac.at/wwa/jsp/index.jsp?fid=23923 &id=34140&typeid=8&display_mode=2 (aufgerufen am: 1.12.2010).

[54] VVÖ (2001) *Sturmversicherung - Allgemeine Bedingungen, Musterbedingungen AStB 2001*, http://www.vvo.at/sturmversicherung/index.php(aufgerufen am: 14.2.2011).

[55] Xu Z. (1993) *Homogeneous stochastic point process model for flood risk analysis*, Extreme Hydrological Events: Precipitation, Floods and Droughts, Proceedings of the International Symposium held at Yokohama, Japan, 20-23 July 1993. ISBN 0-947571-58-2, S. 267-275.

[56] ZAMG – Projekt BEAUFORT, http://www.zamg.ac.at/forschung/klimatologie/klimamodellierung/beauvort/index.php (aufgerufen am 1.9.2011).

Kurzfassung

In Europa wird der Großteil der durch Naturkatastrophen verursachten Versicherungsschäden durch Sturmereignisse verursacht. Zur statistischen Analyse dieser eher seltenen Ereignisse
wird häufig Extremwerttheorie verwendet. In einem Teil dieser Arbeit wird deshalb ein Überblick über diese mathematischen Methoden gegeben. Allerdings müssen die Stürme, welche große Gesamtschäden verursachen, nicht unbedingt extreme Windgeschwindigkeiten besitzen, sondern erstrecken sich meist über eine weite Fläche. Aus diesem Grund werden in dieser Arbeit die verbreiteten Ansätze in der Modellierung von Abhängigkeiten, die Copula-Funktionen, beschrieben. Des Weiteren wird versucht das Distanzmatrix-Modell nach Url, welches im Jahr 2009 bereits von der Forschungsgesellschaft Joanneum Research zur Darstellung der Abhängigkeitsstruktur extremer Hochwasser verwendet wurde, auf Sturmereignisse anzupassen bzw. zu erweitern. Dieses Modell bzw. seine Erweiterungen werden auf die Daten von Versicherungsschäden in • Osterreich der Jahre 1998-2009, welche eine Grenze von 2 Mio. Euro überschreiten, angewendet. Diese Ergebnisse werden anschließend mit jenen der aktuellen Studie zu Sturmschäden der Joanneum Research verglichen.

Abstract

In Europe losses due to storm events make up the major part of insurance losses which are caused by natural catastrophes. Usually extreme value theory is used to describe such rare events. One part of work summarized these mathematical methods. Often the storm events leading to extraordinary insurance losses are not characterized by strong wind speeds but by affecting a large area. Therefore this work describes Copula functions; these are a helpful tool for modeling dependence structures. In addition the "Distanzmatrix-Modell" (distance matrix model) by Url, which was used by the research cooperation Joanneum Research in the year 2009 to build a model for insurance losses cause by extreme floods, is adapted
to storm losses. Then the resulting models are applied to the loss data of the years 1998- 2009 which cumulative insurance loss is larger than 2 Mio. Euro. Finally these results are compared to the current study of storm events by Joanneum Research.

Wegener Center for Climate and Global Change
University of Graz
Leechgasse 25
A-8010 Graz, Austria
www.wegcenter.at
ISBN 978-3-9503112-1-1