



# Diplomarbeit

## Parametrisierung einer Schätzfunktion für die Schadenanzahl eines Sturmereignisses durch meteorologische Daten

Von der Wetterprognose zur Schadenprognose

eingereicht von: Lisa Strasser, BSc  
Matrikelnummer: 0826437  
Studiengang: 066 405 Finanz- und Versicherungsmathematik  
TU Wien

betreut durch: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Peter Grandits (TU Wien)  
Dipl.-Ing. Paul Brandstätter (UNIQA Insurance Group AG)

---

# Kurzfassung

In der vorliegenden Diplomarbeit wird sich das Ziel gesetzt eine geeignete Schätzfunktion für die Anzahl der Versicherungsschäden eines Sturmereignisses zu bestimmen. Um ein besseres Verständnis für die verwendeten Daten zu bekommen, werden kurze Einblicke in die Meteorologie und in die Geodäsie gegeben. Im Hauptteil der Diplomarbeit wird eine Regressionsanalyse mit linearen und generalisierten linearen Modellen durchgeführt. Es folgen die Interpretation und der Vergleich der verschiedenen Ergebnisse und ein kurzes Ausblickskapitel über Möglichkeiten wie das Thema weitergeführt werden könnte. Zum Schluss werden einige im Rahmen dieser Diplomarbeit entstandene Programmierbeispiele zur Verfügung gestellt.

# Abstract

The aim of this master thesis is to define an analytical function which estimates the number of insurance claims caused by a storm event. For a better understanding of the used datasets there are given short insights into meteorology and geodesy. The main part of the master thesis contains a regression analysis with linear and generalized linear models. The following part includes the interpretation and comparison of the different results and suggestions on ways to bring the topic further. To conclude the master thesis there are a few programming examples given.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Einführung in die Meteorologie</b>	<b>3</b>
2.1	Wind . . . . .	4
2.1.1	Entstehung des Windes . . . . .	5
2.2	Sturm in der Versicherung . . . . .	7
<b>3</b>	<b>Geografische Koordinaten und GPS</b>	<b>9</b>
3.1	Geografisches Koordinatensystem . . . . .	9
3.1.1	Geoid . . . . .	10
3.1.2	Rotationsellipsoid . . . . .	11
3.1.3	Rechnen mit geografischen Koordinaten . . . . .	11
3.2	World Geodetic System 1984 . . . . .	15
3.3	GPS - Grundlagen der Ortung mit Satelliten . . . . .	18
<b>4</b>	<b>Die Stürme Paula und Emma</b>	<b>21</b>
4.1	Paula . . . . .	21
4.2	Emma . . . . .	21
4.3	Daten von Ubimet . . . . .	22
4.3.1	Beschreibung der Variablen . . . . .	22
4.3.2	Punkt-in-Polygon-Test von Jordan . . . . .	25
4.4	Sturmkarten . . . . .	27
<b>5</b>	<b>Regressionsmodelle</b>	<b>29</b>
5.1	Datenanalyse . . . . .	29
5.1.1	Zusammenhangsanalyse . . . . .	31
5.2	Lineare Regressionsmodelle . . . . .	33

5.2.1	Das einfache lineare Regressionsmodell . . . . .	33
5.2.2	Methode der kleinsten Quadrate . . . . .	36
5.2.3	Das multiple lineare Regressionsmodell . . . . .	37
5.2.4	Polynomiale Regression . . . . .	38
5.3	Generalisierte lineare Modelle (GLM) . . . . .	40
5.3.1	GLM für Zählraten - Poisson Regression . . . . .	40
5.3.2	Maximum-Likelihood-Schätzung . . . . .	42
5.3.3	Poisson-Regression mit Polynomen . . . . .	43
<b>6</b>	<b>Anwendung der Regressionsanalyse</b>	<b>44</b>
<b>7</b>	<b>Fazit und Ausblick</b>	<b>49</b>
<b>8</b>	<b>Anhang</b>	<b>51</b>
8.1	SAS-Code für ein multiples lineares Modell . . . . .	51
8.2	SAS-Code für den Punkt-in-Polygon-Test . . . . .	53
	<b>Literaturverzeichnis</b>	<b>57</b>

# Kapitel 1

## Einleitung

Seit den 1960er-Jahren hat sich die Anzahl der Wetterkatastrophen in Österreich fast verdoppelt und die volkswirtschaftlichen Schäden sogar versiebenfacht. Um als Versicherungsunternehmen bereits im Vorfeld auf ein herannahendes Sturmereignis reagieren zu können, werden im Rahmen dieser Diplomarbeit verschiedene Regressionsmodelle miteinander verglichen, mit denen die Schadenanzahl eines Sturmes geschätzt werden kann. Im Idealfall können so schon bevor ein Extremereignis überhaupt eingetreten ist, operative Ressourcen gebündelt und gezielt eingesetzt werden.

Das Hauptziel dieser Arbeit liegt in der Untersuchung der Korrelation zwischen Windgeschwindigkeit und Schadenanzahl und eine damit parametrisierte Schätzfunktion, die aus den prognostizierten Wetterdaten sowie zusätzlichen Daten aus dem Versicherungsbestand eine Schadenprognose errechnet.

### Aufbau

Im ersten Teil der Arbeit wird der meteorologische Hintergrund eines Sturmes und dessen Auswirkungen auf das Versicherungswesen erläutert.

Das darauf folgende Kapitel *Geografische Koordinaten und GPS* versucht einen Überblick über die wichtigsten Begriffe der Geodäsie zu geben. Da die Verteilung der Wetterstationen in Österreich von Bundesland zu Bundesland sehr unterschiedlich ist und dies für Berechnungen ungeeignet ist, werden Wetterprognosen auf äquidistanten Gitterpunkten angefertigt. Eine erste Herausforderung stellt die Zuordnung der Versicherungspolizen zu diesen gegebenen Gitterpunkten dar, da jede Polizzae genau dem nächsten Gitterpunkt und dessen Windprognose zugeordnet werden muss. Deshalb ist auch ein großer Teil des Kapitels *Abstandsberechnungen zwischen geografischen Koordinaten* gewidmet.

Im vierten Kapitel erfolgt eine genaue Beschreibung der verwendeten Daten. Die Wetterprognosedaten für die beiden Stürme Emma und Paula wurden von der Wetterfirma *Ubimet* zur Verfügung gestellt. Abgeglichen mit dem Versicherungsbestand von *UNIQA* wurden sie dann mit dem Punkt-in-Polygon-Test von Jordan für die weitere Verwendung aufbereitet.

Der darauf folgende Hauptteil der Arbeit analysiert die Daten in Anlehnung an verschiedene Regressionsmodelle. Die mathematische Herausforderung liegt dabei in der Auswahl einer geeigneten Funktion und deren Parameter. In Kapitel 5 erfolgt die Schätzung dieser Funktionen mit linearen und generalisierten linearen Regressionsmodellen für den Sturm Emma.

Die in diesem Teil berechneten Gleichungen werden anschließend in Kapitel 6 auf die Prognosedaten des Sturmes Paula angewendet. Anhand dieses Praxistests können die verschiedenen Regressionsmodelle verglichen und interpretiert werden.

Darauf folgt ein kurzes Fazit und ein Ausblick auf mögliche Weiterführungen des Themas. Abschließend werden noch ausgewählte Programmiercodes dargestellt.

## **Software**

Für die Berechnungen wird die Computersoftware *SAS* (Statistical Analysis System) verwendet, wobei *SAS* sowohl die Software als auch das herstellende Unternehmen bezeichnet. *SAS* entstand 1966 an der North Carolina State Universität als Projekt für Forschungsanalysen in der Landwirtschaft. Heute wird es in vielen Bereichen verwendet, besonders für Analysen, Datenmanagement und Business Intelligence.

# Kapitel 2

## Einführung in die Meteorologie

In diesem Kapitel wird ein kurzer Überblick über die wichtigsten Begriffe der Meteorologie gegeben. Es folgt eine Beschreibung zur Entstehung des Windes und eine grundsätzliche Betrachtung von Sturmschäden im Versicherungswesen. Für dieses Kapitel werden die im Literaturverzeichnis angegebenen Quellen [6] und [16] verwendet.

Der Begriff *Meteorologie* kommt aus dem Altgriechischen und beschreibt die Lehre der physikalischen und chemischen Vorgänge in der Atmosphäre. Neben der Wettervorhersage beschäftigen sich Meteorologen also auch mit dem Klima, der Umwelt und der Geophysik. Der staatliche meteorologische Dienst Österreichs ist die *Zentralanstalt für Meteorologie und Geodynamik* (ZAMG). Mit ihrem Hauptsitz auf der Hohen Warte (Wien-Döbling) ist die 1851 gegründete ZAMG der älteste Wetterdienst der Welt. Eines der vielen Untergebiete der Meteorologie ist die *Synoptische Meteorologie*, die sich mit dem Wetter (von *Vetor*, indoeuropäisch für Wind) beschäftigt, also mit dem sich verändernden Zustand der Atmosphäre. Es werden verschiedene Prognosehorizonte unterschieden:

Nowcasting:	0	-	2	Stunden
Kürzestfristprognose:	2	-	12	Stunden
Kurzfristprognose:	12	-	72	Stunden
Mittelfristprognose:	3	-	10	Tage

Die Vorhersagen für das *Nowcasting* werden direkt aus Wetterstationsmeldungen und Satellitenbildern abgelesen. Da bei dieser Art der Vorhersage keine Zeitreihen erstellt werden müssen, ist sie die zuverlässigste Form der Wettervorhersage. Die Zuverlässigkeit

der Wetterprognose nimmt ab, je größer der Prognosehorizont ist. Trotzdem ist heutzutage eine Prognose für die kommende Woche in etwa so zuverlässig, wie sie es vor 30 Jahren für den nächsten Tag war.

Die Qualität und Prognosegüte der Wettervorhersage hängt mittlerweile nicht mehr nur von der Stabilität der Wetterlage ab, sondern auch von der Leistungsfähigkeit der Computersysteme, da die Bewältigung der Datenmenge für die Modellberechnungen eine große Herausforderung darstellt.

„Das *Europäische Zentrum für mittelfristige Wettervorhersagen* (ECMWF) in England betreibt einen der stärksten Supercomputer Europas. Seine Spitzenleistung liegt derzeit bei 3593 Teraflops. Das sind 3593 Billionen Rechenoperationen pro Sekunde.“<sup>1</sup>

## 2.1 Wind

Mit *Wind* wird bewegte Luft relativ zur rotierenden Erde bezeichnet. Benannt wird der Luftstrom nach der Richtung aus der er weht. So kommt ein Seewind von einem See und ein Südwind weht aus dem Süden.

Wind trägt in Österreich wesentlich zur Energieversorgung bei, denn es werden ca. 40% der Haushalte mit Strom aus Windkraftanlagen versorgt. Von großer Bedeutung ist der Wind auch im Luft- und Seeverkehr und für Sturmwarnungen, welche herausgegeben werden um die mit hohen Windgeschwindigkeiten verbundenen Schäden zu verringern. Die Windgeschwindigkeit ist in den letzten 20 Jahren in den mittleren Breiten um etwa 5-10% angestiegen, in den Tropen sogar um 20%. Das liegt daran, dass sich das Druckgefälle zwischen dem Äquator und den Polen wegen den allgemein höheren Unterschieden in der Lufttemperatur verstärkt hat. So wird es immer wichtiger die Windgeschwindigkeiten exakt zu prognostizieren und Stürme konkret vorherzusagen.

Die Windgeschwindigkeit ist eine gerichtete Größe und wird als Vektor mit einer horizontalen und einer vertikalen Komponente definiert. Obwohl die vertikale Komponente eine relevante Rolle in der Luftfahrt oder in der Gewitterbildung spielt, ist der Vektor in der Praxis meist nur auf die horizontale Luftströmung beschränkt.

Das horizontale Windfeld wird hauptsächlich von der Bodenreibung und der Corioliskraft bestimmt, so erfolgt ein Einströmen des Windes in das Tiefdruckgebiet und ein Ausströmen aus dem Hochdruckgebiet. Diese Eigenschaft des Windfeldes ist die

---

<sup>1</sup>Quelle: [www.zamg.at](http://www.zamg.at)



Grundlage der BUYS-BALLOT<sup>2</sup>-Regel:

„Steht man mit dem Rücken zum Wind, so liegt auf der Nordhalbkugel (Südhalbkugel) das Tiefdruckgebiet vorne links (rechts) und das Hochdruckgebiet hinten rechts (links).“  
(Klose 2008, S.147)

### 2.1.1 Entstehung des Windes

Der Wind entsteht im Großen und Ganzen durch das Zusammenspiel von folgenden natürlichen Faktoren:

1. den Windsystemen der Erde
2. den Gesetzen der Thermik
3. den topografischen Gegebenheiten vor Ort

Hauptsächlich entsteht Wind jedoch durch die verschiedenen Luftdruckgebiete der Erde. Es werden Hoch- und Tiefdruckgebiete unterschieden, die aus der Einstrahlung der Sonne resultieren. Am Äquator treffen die Sonnenstrahlen annähernd senkrecht auf den Boden auf, wodurch eine Erwärmung der dortigen Lufttemperaturen erfolgt. An den Polspitzen treffen die Sonnenstrahlen in einem flachen Winkel auf die Erde auf, die Luft wird dort deshalb nicht annähernd so stark erwärmt wie am Äquator.

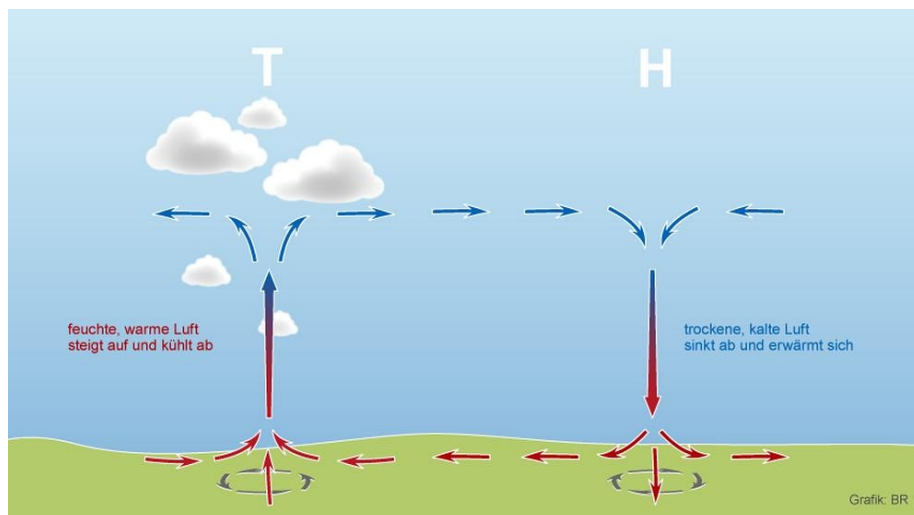


Abb. 2.1: Luftbewegung zwischen Hochdruck- und Tiefdruckgebiet<sup>3</sup>

Je mehr sich die Luft durch die Sonneneinstrahlung aufheizt, desto mehr dehnt sie sich aus, wird leichter und steigt nach oben. Im Gegensatz dazu zieht sich kalte Luft

<sup>2</sup>Christophorus H. D. BUYS-BALLOT (\*1817 in Zeeland; †1890 in Utrecht) war ein niederländischer Wissenschaftler und gilt als einer der Begründer der Meteorologie.

<sup>3</sup>Bildquelle: <http://www.br.de/themen/wissen> am 30.03.2017

zusammen, wird schwerer und konzentriert sich am Boden. Die Luftmoleküle bewegen sich aus dem Gebiet mit dem höheren Druck in das Gebiet mit dem tieferen Druck und zwar so lange, bis der Luftdruck ausgeglichen ist.

Je größer der Druckunterschied zwischen den Gebieten ist, desto heftiger ist auch der daraus entstehende Luftstrom. Da gewisse Ursachen des Druckunterschiedes, wie zum Beispiel die ungleiche Sonneneinstrahlung auf die Erde, nicht ausgeglichen werden können gibt es nicht nur kurzfristige Winde, sondern auch länger andauernde Windsysteme.

Die Luftmassen am Äquator werden zum Beispiel täglich so sehr erwärmt, dass sich dort eine permanente Tiefdruckzone entwickelt hat, die *äquatoriale Tiefdruckrinne*. In Folge der Abkühlung sinken die Luftmassen bei 30° nördlicher und südlicher Breite wieder ab und verursachen dort zwei Gürtel mit hohem Druck und großen Wüsten, wie zum Beispiel der Sahara. Von dort aus strömt die Luft wieder zurück zum Äquator. Durch die Erdrotation werden die Luftströmungen abgelenkt und verlaufen deshalb nicht direkt sondern spiralförmig. Diese dauernden Luftbewegungen werden *Passatwinde* genannt.

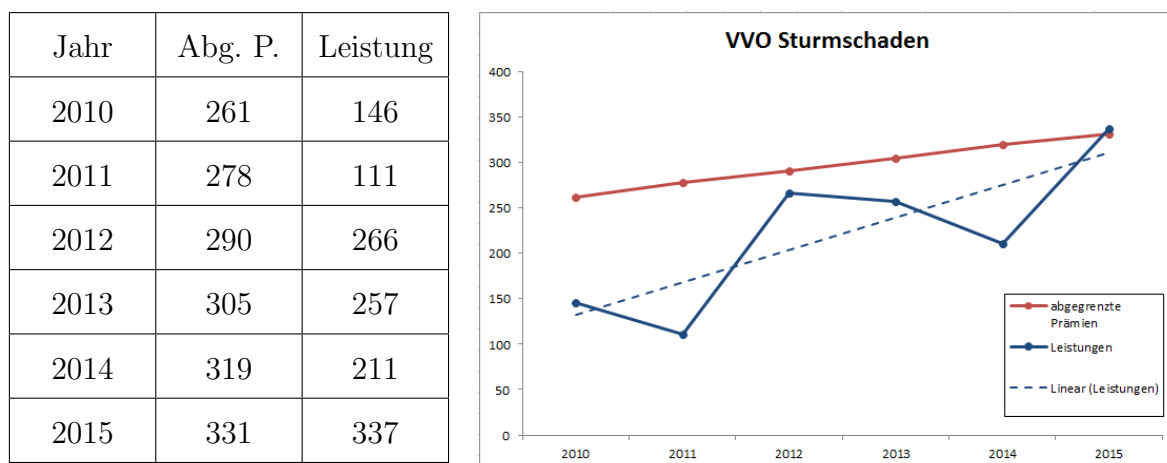
### Windnamen

Verschiedene bekannte Windsysteme, wie z. B. der Passat, die Bora oder der Monsun werden zur besseren Unterscheidung der bestimmten meteorologischen bzw. klimatischen Eigenheiten mit Namen versehen. Neben diesen Windsystemen ist es auch für kurzfristigere Wetterereignisse üblich Namen zu vergeben.

Für mitteleuropäische Hoch- und Tiefdruckgebiete werden diese Namen vom *Meteorologischen Institut der Freien Universität Berlin* vergeben. Der Einfachheit halber werden in Österreich die Namen aus Deutschland übernommen, deshalb heißen auch die zwei in dieser Diplomarbeit betrachteten Stürme *Emma* und *Paula*.

## 2.2 Sturm in der Versicherung

In den letzten Jahren sind Versicherungsschäden bedingt durch starken Wind enorm angestiegen. Die folgende Tabelle enthält die abgegrenzten Prämien und Leistungen (in Mio. €) der letzten sechs Jahre für Sturmschäden in Österreich. Als Grundlage für diese Tabelle dienen die Geschäftsberichte des *Versicherungsverbandes Österreichs* (VVO).



Im Versicherungswesen werden Schäden als Sturmschäden bezeichnet, wenn sie durch einen Wind mit mehr als 60 km/h verursacht werden. Die Feststellung der Sturmgeschwindigkeit übernimmt im Einzelfall die ZAMG. Grundsätzlich werden drei Arten von Sturmschäden unterschieden:

1. Schäden, die durch die unmittelbare Sturmeinwirkung entstehen, also zum Beispiel wenn versicherte Sachen durch den Druck aufprallender Luft beschädigt oder zerstört werden.
2. Schäden, die hervorgerufen werden, weil Gegenstände (wie Bäume, Masten, usw.) durch das Sturmereignis auf die versicherten Sachen geworfen werden.
3. Schäden, die durch die direkte Folge eines Sturmes entstehen, wie zum Beispiel Niederschlagswasser, dass durch den Sturm beschädigte Dach- oder Mauerteile eindringt.

Aber auch die indirekten Folgeschäden von Stürmen sind nicht zu unterschätzen. Beispielsweise gehört die Aufarbeitung von Sturmholz zu den gefährlichsten Tätigkeiten bei der Waldarbeit. Obwohl im Jahr 2007 zum Zeitpunkt des Orkans *Kyrill* in Österreich keine Todesopfer zu beklagen waren, kam beim späteren Schneiden der verspannten Bäume ein Landwirt ums Leben.

Hauptsächlich betreffen Sturmschäden aber abgedeckte Dächer und kaputte Rauchfänge. Diese Art von Schäden wird meist im Rahmen einer Eigenheimversicherung gedeckt. Für durch Sturm beschädigte Fahrzeuge reicht eine Kfz-Haftpflichtversicherung meist nicht aus, hier kommt nur die Voll- oder Teilkaskoversicherung für Schäden auf. Speziellere Risiken, wie zum Beispiel Solaranlagen oder großflächige Verglasungen sind durch Zusatzversicherungen gedeckt.

Schäden an beweglichen Gegenständen im Freien, wie Gartenmöbel oder Sonnenschirme werden von den meisten Haushaltsversicherungen gedeckt. Da sich in den letzten Jahren ein Trend zu immer wertvolleren Gartenmöbeln, Grillgeräten und elektrischen Markisen bzw. Sonnensegel entwickelt hat, steigt tendenziell die Schadenhöhe pro Schadensfall. Bei einem Sturmereignis sind für ein Versicherungsunternehmen weniger die einzelnen Großschäden von Bedeutung, sondern eher die vielen Kleinschäden. Einige dieser Schäden würden sich durch vorzeitige Maßnahmen des Versicherungsnehmers verhindern lassen, deshalb bieten immer mehr Versicherer SMS-Warnsysteme oder Apps zur Schadensminimierung von Sturmschäden an.



Die geografische Breite nimmt am Äquator den Wert  $0^\circ$  an. Es werden jeweils 90 Grade in Richtung Norden bzw. Süden gezählt, sodass sich  $90^\circ$  Nord ( $90^\circ\text{N}$ ) am Nordpol und  $90^\circ$  Süd ( $90^\circ\text{S}$ ) am Südpol befindet. Fehlt die Angabe  $^\circ\text{N}$  bzw.  $^\circ\text{S}$ , so stehen positive Werte für nördliche und negative für südliche Breitengrade.

Die geografische Länge ist ein Winkel, der ausgehend vom Nullmeridian ( $0^\circ$ ) in Greenwich gemessen wird. Mit jeweils  $180^\circ$  in östlicher und westlicher Richtung wird die gesamte Erde umspannt.

Die geografischen Koordinaten werden im Sexagesimalsystem angegeben, das heißt dass ein Grad in 60 Minuten unterteilt werden kann. Eine Minute entspricht wiederum 60 Sekunden. Typografisch werden die Minuten mit einer Prime ( $'$ ) und die Sekunden mit einer Doppelprime ( $''$ ) gekennzeichnet.

*Beispiel:*  $16^\circ 22' 23,294''$  und  $48^\circ 12' 30,582''$  sind die geografischen Koordinaten vom Stephansdom in Wien.

Mittlerweile wird auch immer häufiger die Dezimalschreibweise verwendet, über die durch beliebig viele Nachkommastellen die gewünschte Genauigkeit erreicht werden kann.

*Beispiel:*  $16,373056^\circ$  und  $48,208333^\circ$  beschreibt wieder den Stephansdom.

### 3.1.1 Geoid

Um die Funktionsweise von Satellitennavigationssystemen wie GPS zu verstehen, ist es notwendig, sich vorher mit den Grundzügen der Geodäsie zu beschäftigen.

Die Geodäsie ist laut der Definition von F. R. HELMERT<sup>5</sup> (1880) die Wissenschaft der Ausmessung und Abbildung der Erdoberfläche. Dies umfasst unter anderem die Bestimmung der Figur der Erde. Schon in der griechischen Antike gingen Gelehrte wie PYTHAGORAS und PLATON von einer kugelförmigen Gestalt der Erde aus. Allerdings erhoben einige Autoren der Spätantike Einspruch gegen diese Hypothese. Zwar war der Einfluss dieser Autoren sehr gering, trotzdem wurde die Kugelgestalt der Erde erst zu Zeiten von KOLUMBUS nicht mehr in Frage gestellt.

Die wirkliche Form der Erde exakt zu beschreiben bleibt trotzdem eine schwierige Wissenschaft. Eine Annäherung an die wahre Körpergestalt der Erde wird *Geoid* genannt.

---

<sup>5</sup>Friedrich Robert HELMERT (\*1843 in Freiberg; †1917 in Potsdam) war ein deutscher Geodät und Mathematiker.

### 3.1.2 Rotationsellipsoid

Nur sehr schwer können Berechnungen auf dem Geoid durchgeführt werden, da es keine analytische Fläche ist. Um eine mathematisch definierte Annäherung zu erhalten wird eine Achse zwischen Nord- und Südpol definiert. Lässt man die Fläche einer Ellipse um diese Symmetrieachse rotieren, erhält man ein *Rotationsellipsoid*. Wegen ihrer Einfachheit werden solche Ellipsoide auch heute noch für Berechnungen auf der Erdoberfläche benutzt.

Ein Rotationsellipsoid wird durch eine große Halbachse  $a$  (Äquatorradius) und einer kleinen Halbachse  $b$  bestimmt. Als Maß für die Abweichung von der idealen Form der Kugel wird die Abplattung  $f$  durch  $f = \frac{a-b}{a}$  berechnet.

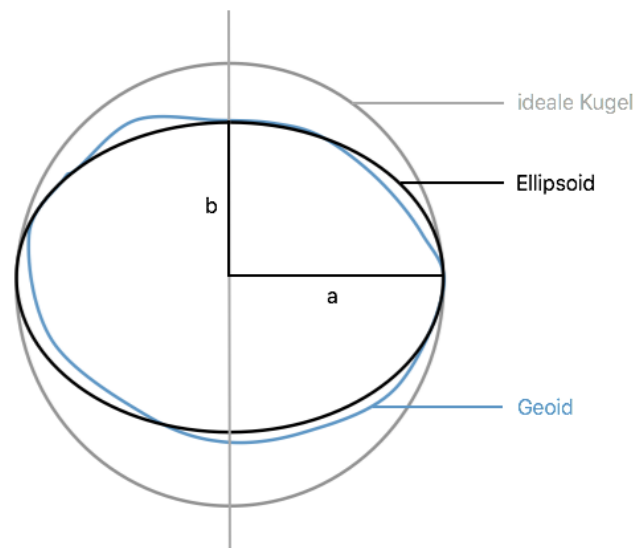


Abb. 3.2: Gestalt der Erde

Jedes geodätische System besitzt ein bestimmtes Rotationsellipsoid, das bestmöglich an die Erdoberfläche angepasst ist, es wird *Referenzellipsoid* genannt.

Das derzeit am häufigsten verwendete globale Referenzsystem ist das *World Geodetic System 1984* (siehe Kapitel 3.2).

### 3.1.3 Rechnen mit geografischen Koordinaten

Durch die Erdkrümmung sind die Abstände zwischen zwei Breitengraden bzw. zwei Längengraden unterschiedlich groß.

Der Abstand zwischen zwei Breitengraden  $\Delta_{lat}$  variiert von 110,57 km bis 111,69 km und wird in ungenaueren Berechnungen gerne mit 111,3 km definiert.

Der Abstand zwischen zwei Längengraden  $\Delta_{long}$  konvergiert gegen 0, am Äquator beträgt er 111,32 km und verschwindet an den Polspitzen (siehe Tabelle 3.1).

Um mit geografischen Koordinaten rechnen zu können, wird die Form der Erde einem Rotationsellipsoid angenähert (siehe Kapitel 3.1.2).

Bei einer Ellipse wird der Abstand der Brennpunkte vom Mittelpunkt als *lineare Exzentrizität*  $\epsilon$  bezeichnet. Das Verhältnis der linearen Exzentrizität zur großen Halbachse  $a$  ist als *numerische Exzentrizität*  $e$  bekannt.

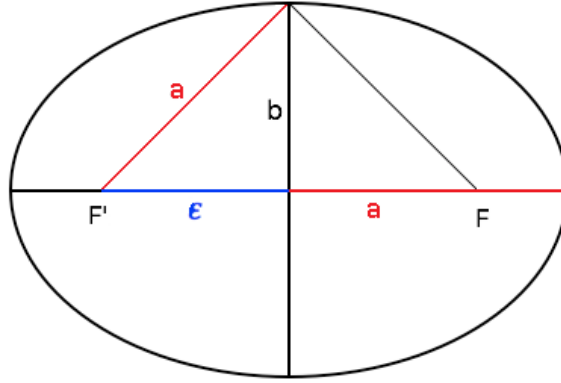


Abb. 3.3: Lineare Exzentrizität einer Ellipse

Mathematisch beschreibt sich der Zusammenhang zwischen  $\epsilon$  und  $e$  folgendermaßen:

$$\epsilon = \sqrt{a^2 - b^2} \qquad e = \frac{\epsilon}{a} = \frac{\sqrt{a^2 - b^2}}{a}$$

Nun kann die Mediankrümmung  $M$  an der Stelle  $\alpha$  mit

$$M(\alpha) = \frac{a \cdot (1 - e^2)}{(1 - e^2 \cdot \sin^2 \alpha)^{3/2}} \quad (3.1)$$

angegeben werden.

Aus Formel (3.1) können folgende Abstandsformeln für  $\Delta_{lat}^{1^\circ}$  und  $\Delta_{long}^{1^\circ}$  abgeleitet werden:

$$\Delta_{lat}^{1^\circ}(\alpha) = \frac{\pi}{180} \cdot \frac{a \cdot (1 - e^2)}{(1 - e^2 \cdot \sin^2 \alpha)^{3/2}} \quad (3.2)$$

$$\Delta_{long}^{1^\circ}(\alpha) = \frac{\pi}{180} \cdot \frac{a \cdot \cos \alpha}{\sqrt{1 - e^2 \cdot \sin^2 \alpha}} \quad (3.3)$$

$$e^2 = \frac{a^2 - b^2}{a^2} \quad (3.4)$$



**Beispiel 3.1.**

Im Fokus dieses Beispiels steht die Berechnung der Entfernung zwischen zwei Breiten-graden am Äquator. Mit den Werten  $a = 6378137$  Meter und  $b = 6356752,31$  Meter aus dem Referenzsystem WGS84 (siehe Kapitel 3.2) wird zuerst  $e^2$  aus der Formel 3.4 berechnet. Um die Entfernungen am Äquator zu erhalten wird  $\alpha = 0^\circ$  definiert und in die Formel 3.2 eingesetzt:

$$\Delta_{lat}^{1^\circ}(0^\circ) = 0,0175 \cdot \frac{6378137 \cdot (1 - 0,0067)}{(1 - 0,0067 \cdot 0)^{3/2}} = 110574,27 \text{ Meter}$$

$$\Delta_{lat}^{1'}(0^\circ) = 1842,90 \text{ Meter}$$

$$\Delta_{lat}^{1''}(0^\circ) = 30,72 \text{ Meter}$$

Ursprünglich wurde eine Seemeile durch eine Winkelminute auf einem Breitengrad definiert, das ist am Äquator 1842,90 Meter (siehe obiges Beispiel).

Da nach dieser Definition eine Seemeile auf den unterschiedlichen Breitengraden verschiedene Werte annehmen würde, wurde sie später bei 1852,21 Meter fixiert.

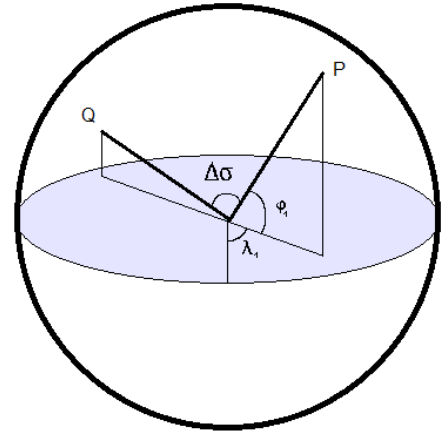
In der folgenden Tabelle sind die unterschiedlichen Abstände zwischen Breitengraden und Längengraden in Kilometern ausgerechnet.

$\alpha$	$\Delta_{lat}^{1^\circ}(\alpha)$	$\Delta_{long}^{1^\circ}(\alpha)$
$0^\circ$	110,57 km	111,32 km
$15^\circ$	110,65 km	107,55 km
$30^\circ$	110,85 km	96,49 km
$45^\circ$	111,13 km	78,85 km
$60^\circ$	111,41 km	55,80 km
$75^\circ$	111,62 km	28,90 km
$90^\circ$	111,69 km	0,00 km

Tabelle 3.1: Abstände zwischen Breiten- bzw. Längengraden

Bis jetzt wurden nur Distanzen berechnet die sich auf dem selben Längen- bzw. Breitengrad befanden. Um nun den Abstand zwischen zwei beliebig liegenden Punkten auf der Erde zu bestimmen, wird die *Orthodrome* (bzw. *Great-Circle Distance*) berechnet. Die Orthodrome beschreibt die kürzeste Verbindung zweier Punkte auf einer Kugeloberfläche und wird daher umgangssprachlich *Luftlinie* genannt.

Es gibt verschiedene Arten die Orthodrome zu berechnen, die einfachste Methode wird aus der sphärischen Trigonometrie abgeleitet. Dazu wird der Abstand der zwei geografischen Koordinaten  $P = (\varphi_1, \lambda_1)$  und  $Q = (\varphi_2, \lambda_2)$  mit  $d$  bezeichnet. Die absolute Differenz von  $\lambda_1$  und  $\lambda_2$  ist mit  $\Delta\lambda$  definiert. Der Winkel zwischen den Koordinaten  $\Delta\sigma$  (siehe Abb. 3.4) wird folgendermaßen berechnet:



$$\Delta\sigma = \arccos[\sin \varphi_1 \cdot \sin \varphi_2 + \cos \varphi_1 \cdot \cos \varphi_2 \cdot \cos(\Delta\lambda)]$$

Abb. 3.4: Winkel  $\Delta\sigma$

Um die Distanz  $d$  zu erhalten muss der Winkel  $\Delta\sigma$  noch mit dem mittleren Erdradius  $r = 6371000,8$  Meter multipliziert werden, also  $d = r \cdot \Delta\sigma$ .

### Beispiel 3.2.

Der Abstand zwischen dem Stephansdom  $P = (48,2083333; 16,3730560)$  und dem Schloss Schönbrunn  $Q = (48,1848648; 16,3122398)$  wird berechnet.

Zuerst werden alle Winkel mit dem Faktor  $\frac{\pi}{180}$  multipliziert um ins Bogenmaß zu wechseln. Dann wird der Winkel  $\Delta\sigma$  berechnet:

$$\Delta\sigma = \arccos[\sin \varphi_1 \cdot \sin \varphi_2 + \cos \varphi_1 \cdot \cos \varphi_2 \cdot \cos(\Delta\lambda)] = 0,0008175435$$

Nun kann der Abstand  $d$  ermittelt werden:

$$d = r \cdot \Delta\sigma = 6371000,8 \cdot 0,000817543 = 5208,5703041 \text{ Meter}$$

In SAS wird für Abstandsberechnungen die vordefinierte Funktion `geodist` verwendet. Diese Funktion basiert auf der Abstandsformel von VINCENTY<sup>6</sup> und ist genauer als die vorher gezeigte Version. Bei der Formel von Vincenty ändert sich nur die Berechnung des Winkels  $\Delta\sigma$  in:

$$\Delta\sigma = \arctan \frac{\sqrt{[\cos \varphi_2 \cdot \sin(\Delta\lambda)]^2 + [\cos \varphi_1 \cdot \sin \varphi_2 - \sin \varphi_1 \cdot \cos \varphi_2 \cdot \cos(\Delta\lambda)]^2}}{\sin \varphi_1 \cdot \sin \varphi_2 + \cos \varphi_1 \cdot \cos \varphi_2 \cdot \cos(\Delta\lambda)} \quad (3.5)$$

<sup>6</sup>Thaddeus VINCENTY (\*1920 in Polen; †2002 in Maryland) war ein polnisch-amerikanischer Geodät. Von ihm stammt unter anderem ein Algorithmus zur Lösung der ersten und zweiten Hauptaufgabe der Geodäsie.

**Beispiel 3.3.**

Wieder wird der Abstand zwischen dem Stephansdom  $P = (48,2083333; 16,3730560)$  und dem Schloss Schönbrunn  $Q = (48,1848648; 16,3122398)$  berechnet. Dieses Mal mit der Abstandsformel von Vincenty.

Für  $\Delta\sigma$  ergibt Formel 3.5 den Wert 0,0008175435. Damit errechnet sich für die Distanz  $d = 5208,5703049$  Meter.

Im Vergleich zum vorherigen Beispiel ändert sich erst die siebente Kommastelle, das heißt für kurze Abstände ändert sich das Ergebnis kaum merklich. Jedoch nimmt der Fehler mit der Länge der Distanz zu, somit ist die Abstandsformel von Vincenty zwar komplizierter zum Berechnen, liefert aber exaktere Ergebnisse.

## 3.2 World Geodetic System 1984

Größere regionale Abweichungen der Erdfigur vom idealen Ellipsoid glichen Geodäten früher dadurch aus, indem sie im betreffenden Gebiet ein eigenes Ellipsoid berechneten. Viele Nationen entwickelten so eigene geodätische Systeme die nur kleine Teile der Erde umfassen. In Österreich wird zum Beispiel für vermessungstechnische Anwendungen das System *MGI* (Militärgeografisches Institut) mit dem Referenzellipsoid von BESSEL<sup>7</sup> (1841) verwendet. Das Ziel des 1984 definierten System WGS84 war die nationalen Vermessungssysteme zu einem gemeinsamen Weltsystem zu vereinheitlichen.

Das Referenzellipsoid von WGS84 wird mit der Halbachse  $a = 6378137$  Meter und der Abplattung  $f = \frac{1}{298,25}$  definiert. Weiters wird das Ellipsoid von verschiedenen physikalischen Parametern beschrieben, wie zum Beispiel der *Rotationsgeschwindigkeit*  $\omega$  oder der *Geozentrischen Gravitationskonstanten*  $GM$ .

Das WGS84-Koordinatensystem ist ein dreidimensionales, kartesisches Koordinatensystem mit dem Koordinatenursprung im Massezentrum des Referenzellipsoides. Die positive  $x$ -Achse des Ellipsoids geht vom Mittelpunkt durch den Schnittpunkt von Äquator und Greenwich-Meridian (0-Meridian). Die  $y$ -Achse liegt ebenfalls in der Äquatorebene, nur 90° versetzt zur  $x$ -Achse. Die  $z$ -Achse steht senkrecht auf der  $x$ - und  $y$ -Achse und geht durch den geografischen Nordpol.

<sup>7</sup>Friedrich Wilhelm BESSEL (\*1784 in Westfalen; †1846 in Königsberg) war einer der bedeutendsten Wissenschaftler des 19. Jahrhunderts.

Anstatt der kartesischen Koordinaten  $x, y$  und  $z$  werden normalerweise die ellipsoidischen Koordinaten  $\lambda, \varphi$  und  $h$  verwendet. Hier entspricht  $\lambda$  dem Längengrad (Longitude),  $\varphi$  dem Breitengrad (Latitude) und  $h$  der Höhe, also der Länge des Lotes vom Punkt  $P$  bis zum Ellipsoid.

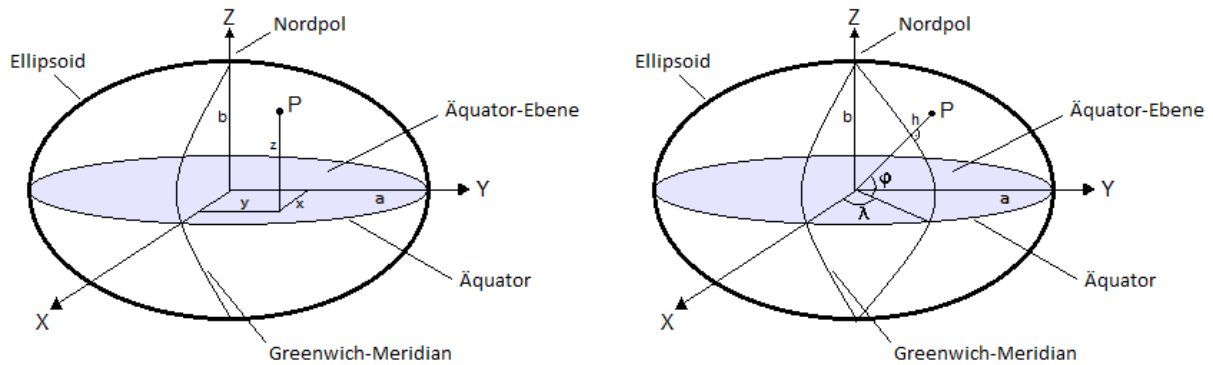


Abb. 3.5: Kartesische Koordinaten (links) und ellipsoidische Koordinaten (rechts)

Für die Umrechnung von ellipsoidischen Koordinaten zu kartesischen Koordinaten wird die Hilfsgröße  $e^2$  aus Formel 3.4 benötigt. Die kartesischen Koordinaten  $(x, y, z)$  und die weitere Hilfsgröße  $N$  ergeben sich aus folgenden Gleichungen:

$$N = \frac{a}{\sqrt{1 - [e^2 \cdot (\sin \varphi)^2]}} \quad (3.6)$$

$$x = (N + h) \cdot \cos \varphi \cdot \cos \lambda \quad (3.7)$$

$$y = (N + h) \cdot \cos \varphi \cdot \sin \lambda \quad (3.8)$$

$$z = [N \cdot (1 - e^2) + h] \cdot \sin \varphi \quad (3.9)$$

#### Beispiel 3.4.

Der Punkt  $P = (\varphi, \lambda, h) = (16,373056; 48,208333; 178)$  beschreibt die ellipsoidischen Koordinaten des Stephansdomes in Wien. Für die beiden Hilfsgrößen  $e^2$  und  $N$  werden aus den Formeln 3.4 und 3.6 folgende Werte berechnet:

$$e^2 = 0,0066943813$$

$$N = 6379834,10$$

Nun können die Winkel (umgerechnet ins Bogenmaß) und die Höhe  $h$  in die obigen Gleichungen eingesetzt werden und führen zu folgendem Ergebnis:

$$x = 4\,079\,368,95 \text{ Meter}$$

$$y = 4\,563\,861,56 \text{ Meter}$$

$$z = 1\,786\,424,24 \text{ Meter}$$

Wird also wie in Abbildung 3.5 ein kartesisches Koordinatensystem mit dem Erdmittelpunkt als Ursprung betrachtet, müssen auf der  $x$ -Achse 4 079 368 Meter, auf der  $y$ -Achse 4 563 861 Meter und auf der  $z$ -Achse 1 786 424 Meter abgezählt werden um genau beim Stephansdom zu landen.

Die umgekehrte Transformation von kartesischen Koordinaten  $(x, y, z)$  in ellipsoidische Koordinaten  $(\varphi, \lambda, h)$  ist durch direkte Rechnung nicht möglich, da die einzelnen Größen voneinander abhängen. Nur  $\lambda$  kann geradewegs durch  $\lambda = \arctan(\frac{y}{x})$  berechnet werden. Die Werte  $\varphi$  und  $h$  werden durch einen iterativen Prozess angenähert.

1. Zuerst wird der Startwert für  $\varphi = \arctan(\frac{z}{\sqrt{x^2+y^2}})$  bestimmt.
2. Nun wird mit diesem Startwert in die Formel 3.6 eingesetzt und die Hilfsgröße  $N$  berechnet.
3. Aus  $N$  und  $\varphi$  kann die Höhe  $h$  mit  $h = \frac{\sqrt{x^2+y^2}}{\cos\varphi} - N$  bestimmt werden.
4. Im letzten Schritt wird der Startwert für  $\varphi$  ersetzt:

$$\varphi = \arctan\left(\frac{\frac{z}{\sqrt{x^2+y^2}}}{1 - \frac{N \cdot e^2}{N+h}}\right)$$

Diese vier Schritte werden beliebig oft wiederholt, meist reichen jedoch fünf Iterationen aus um die gewünschte Genauigkeit für die ellipsoidischen Koordinaten zu erlangen.

### Beispiel 3.5.

Es sollen die im vorherigen Beispiel errechneten kartesischen Koordinaten  $(x, y, z)$  für den Stephansdom in ellipsoidische Koordinaten  $(\varphi, \lambda, h)$  transformiert werden.

Aus der direkten Berechnung von  $\lambda$  folgt  $\lambda = 48,208333$ .

Der oben beschriebene Prozess wurde in diesem Beispiel fünf Mal wiederholt. Die Differenz zwischen dem wahren  $\varphi$  und dem hier errechneten beträgt nur  $4,250538 \cdot 10^{-10}$ .

Startwerte	nach fünf Iterationen
$\varphi = 16,269266$	$\varphi = 16,373056$
$N = 6379813,22$	$N = 6379834,10$
$h = -3184,41$	$h = 177.99996727$

### 3.3 GPS - Grundlagen der Ortung mit Satelliten

Zur Bestimmung geografischer Positionen werden vorrangig Satellitensysteme verwendet. Auch für die Navigation im Gelände oder für die Bestimmung der exakten Zeit werden solche Systeme benutzt. Es gibt mehrere verschiedene satellitengestützte Positionssysteme, wie zum Beispiel GPS, GLONASS oder GALILEO. Sie alle werden unter dem Begriff *Global Navigation Satellite System* (GNSS) zusammengefasst.

Das wohl bekannteste Ortungssystem ist das *Global Positioning System* (GPS) oder auch *Navigation System with Time and Range* (NAVSTAR). Mit GPS kann überall auf der Welt der genaue Standort und die genaue Zeit bestimmt werden. Es wird durch das Verteidigungsministerium der Vereinigten Staaten betrieben, das 1973 die Entwicklung des Systems gestartet hat. Seit Dezember 1993 ist GPS voll funktionsfähig. Mittlerweile hat sich die Bezeichnung GPS sogar so sehr eingebürgert, dass sie umgangssprachlich für alle Satellitennavigationssysteme benutzt wird.

GPS kann von zivilen und militärischen Anwendern genutzt werden. Das Funksignal SPS (*Standard Positioning Service*) ist von der Allgemeinheit frei nutzbar und wird für Autonavigationsgeräte, Ballonflüge und dergleichen verwendet. Das militärische Signal PPS (*Precise Positioning Service*) darf nur von autorisierten Stellen genutzt werden und wird verschlüsselt ausgestrahlt.

Die ersten GPS-Satelliten wurden 1978 in die Umlaufbahn gebracht und mittlerweile umkreisen 32 aktive Satelliten auf sechs Bahnen die Erde. Diese Satelliten sind genau so verteilt, dass von jedem Punkt der Erde aus immer mindestens vier zu sehen sind.

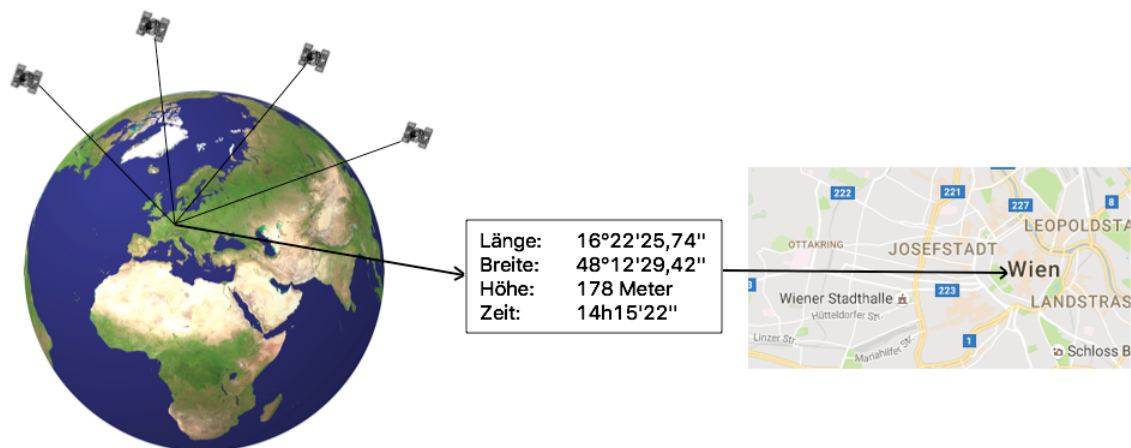


Abb. 3.6: Die grundlegende Funktion von GPS

Die Satelliten in den Umlaufbahnen senden permanent ihre genaue Position und Zeit zu dem GPS-fähigen Empfänger  $P$  auf der Erde. Da die Berechnungen in einem kartesischen, dreidimensionalen Koordinatensystem durchgeführt werden, übermittelt jeder der vier benötigten Satelliten  $S_1, S_2, S_3$  und  $S_4$  seine Position in Form des Vektors  $(x_i, y_i, z_i)^T$  mit  $i = 1, \dots, 4$ . Weiters wird von den Satelliten der Zeitpunkt der Signalausendung übertragen, der dank eingebauter Atomuhren genauestens bekannt ist. Die Differenz von den gesendeten Satelliten-Zeiten mit der GPS-Zeit auf der Erde ergibt die vier Signallaufzeiten  $\Delta t_1, \Delta t_2, \Delta t_3$  und  $\Delta t_4$ .

Um nun die Entfernungen  $R_1, R_2, R_3$  und  $R_4$  von den Satelliten zum GPS-Empfänger berechnen zu können, wird das *euklidische Skalarprodukt* und die *euklidische Norm* definiert.

**Definition 3.6. Euklidisches Skalarprodukt**

Das euklidische Skalarprodukt ist definiert durch

$$\langle x, y \rangle_2 = x^T \cdot y = x_1 \cdot y_1 + \dots + x_n \cdot y_n \quad \forall x, y \in \mathbb{R}^n.$$

**Definition 3.7. Euklidische Norm**

Die euklidische Norm oder die Länge eines Vektors ist definiert durch

$$\|x\|_2 = \sqrt{\langle x, x \rangle_2} \quad \forall x \in \mathbb{R}^n.$$

Es ergeben sich also für  $i = 1, \dots, 4$  die Gleichungen:

$$R_i = \|S_i - P\|_2 = \sqrt{(x_i - x_P)^2 + (y_i - y_P)^2 + (z_i - z_P)^2}$$

Trotz der vergleichsweise einfachen Abstandsberechnung gibt es bei der Laufzeitmessung Probleme. Für einfache GPS-Empfänger wären Atomuhren viel zu teuer (etwa 100 000 €) und viel zu schwer, deshalb werden hier einfache Quarzuhren verwendet. Da diese Uhren aber die Genauigkeit der Atomuhren nicht erreichen, wird deren Zeitangabe mit einem Uhren-Offset  $dt$  versehen.

Alle Entfernungsmessungen sind um den gleichen Anteil zu lang oder zu kurz und da sich die elektromagnetischen Wellen zwischen Satelliten und Empfänger mit Lichtge-

schwindigkeit  $c = 299\,792\,458$  m/s bewegen, beträgt dieser Anteil  $c \cdot dt$ .

Wenn  $c \cdot dt$  zu den wirklichen Entfernungen  $R_i$  addiert wird, ergeben sich die sogenannten Pseudo-Entfernungen  $PE_i$ :

$$PE_i = \|S_i - P\|_2 + c \cdot dt$$

Das zu lösende nichtlineare Gleichungssystem lautet also:

$$PE_1 = \|S_1 - P\|_2 + c \cdot dt = \sqrt{(x_1 - x_P)^2 + (y_1 - y_P)^2 + (z_1 - z_P)^2} + c \cdot dt$$

$$PE_2 = \|S_2 - P\|_2 + c \cdot dt = \sqrt{(x_2 - x_P)^2 + (y_2 - y_P)^2 + (z_2 - z_P)^2} + c \cdot dt$$

$$PE_3 = \|S_3 - P\|_2 + c \cdot dt = \sqrt{(x_3 - x_P)^2 + (y_3 - y_P)^2 + (z_3 - z_P)^2} + c \cdot dt$$

$$PE_4 = \|S_4 - P\|_2 + c \cdot dt = \sqrt{(x_4 - x_P)^2 + (y_4 - y_P)^2 + (z_4 - z_P)^2} + c \cdot dt$$

Für nichtlineare Gleichungssysteme gibt es keine allgemeingültigen Lösungsstrategien. Oft werden nur näherungsweise Lösungen mit Hilfe numerischer Verfahren bestimmt. Auf die genauen Algorithmen zur Lösung nichtlinearer Gleichungssysteme soll hier nicht näher eingegangen werden. Meistens wird zuerst ein Schätzwert für den Punkt  $P$  berechnet und dann mit den üblichen iterativen Verfahren so lange wieder in das Gleichungssystem eingesetzt, bis die gewünschte Genauigkeit erreicht wird. Je nach anfänglichem Schätzwert werden etwa drei bis fünf iterative Berechnungen benötigt um die Unbekannten  $x_P$ ,  $y_P$ ,  $z_P$  und  $dt$  mit einem Fehleranteil von unter einem Zentimeter zu erhalten.

Die Abstandsberechnungen zwischen einem Punkt auf der Erde und den GPS-Satelliten werden in dieser Diplomarbeit benötigt, um jedem Sturmrisiko eine geografische Koordinate zuzuordnen zu können. So kann in weiterer Folge jede Polizze einem in Kapitel 4.3 beschriebenen Quadrat zugewiesen werden.



# Kapitel 4

## Die Stürme Paula und Emma

### 4.1 Paula

Der Sturm Paula zog vom 26. Januar bis zum 27. Januar 2008 über Deutschland und Österreich hinweg. Er verursachte mit über sechs Millionen Erntefestmeter Sturmholz besonders in der österreichischen Forstwirtschaft einen der schwersten Waldschäden seit 1945. Der unmittelbare Schaden der ausschließlich aus dem Wertverlust des Holzes entstanden ist, betrug über 91 Millionen Euro. Schäden dieser Art sind großteils nicht versichert, trotzdem entstand für die österreichische Versicherungswirtschaft ein Schaden in der Höhe von rund 105 Millionen Euro. Obwohl die Spitzengeschwindigkeit in Niederösterreich mit 230 km/h gemessen wurde, wütete der Sturm am meisten in der Steiermark und in Kärnten. Winterstürme dieser Stärke kommen in Südostösterreich nur alle 20 bis 30 Jahre vor, meist ist die Alpensüdseite von schweren Orkanen wie Emma (2008) oder Kyrill (2007) kaum betroffen.

### 4.2 Emma

Nur einen Monat nach Paula zog der Orkan Emma vom 29. Februar bis zum 2. März 2008 über Mitteleuropa. Am Tag zuvor bildete sich das Tiefdruckgebiet vor der Nordostküste Nordamerikas und überquerte den Atlantischen Ozean.

Der Begriff Orkan wird verwendet, sobald die Windgeschwindigkeiten 118 km/h überschreiten, also Stärke 12 auf der Beaufort-Skala erreichen. Emma erreichte am Krippenstein in 2000 Metern Höhe eine Spitzengeschwindigkeit von 236 km/h. Trotz frühzeitiger Erkennung des Orkans forderte Emma europaweit mindestens 14 Tote und

verursachte einen Gesamtversicherungsschaden von rund einer Milliarde Euro. Davon entstand in Österreich ein versicherter Schaden von etwa 175 Millionen Euro.

## 4.3 Daten von Ubimet

Um einen Zusammenhang zwischen Windgeschwindigkeiten und Schadenanzahlen finden zu können werden in erster Linie Daten gesammelt und analysiert. Die hier verwendeten Windgeschwindigkeiten für die Stürme Paula und Emma stammen von der Wetterfirma *Ubimet*. Dieses Unternehmen wurde 2004 gegründet und ist der größte private Wetterdienst in Österreich<sup>8</sup>. Die verwendeten Schaden- und Bestandsdaten wurden von *UNIQA* bereitgestellt.

Um mathematisch die Schadenanzahl durch Windgeschwindigkeiten schätzen zu können, wird eine sogenannte Regressionsanalyse (siehe Kapitel 5) durchgeführt. Für die Regression kommen eine Vielzahl von Variablen in Frage, deswegen werden im folgenden Unterkapitel nur die wichtigsten zur Verfügung gestellten Variablen genauer beschrieben.

### 4.3.1 Beschreibung der Variablen

Die folgende Tabelle 4.1 enthält eine Liste der verwendeten Variablen.

lon	Geografische Länge des Gitterpunktes in Grad
lat	Geografische Breite des Gitterpunktes in Grad
id	Identifikationsnummer des Quadrates
Zip	Postleitzahl
Seehöhe	Seehöhe am Punkt (Lon, Lat)
Timestamp_Init	Zeitpunkt der Wetterprognose
Timestamp_Valid	Zeitpunkt für den die Wetterprognose gemacht wurde
wind_i	Prognostizierte Windgeschwindigkeit bei Timestamp_Valid
windgust_i	Prognostizierte Geschwindigkeit der Windböen bei Timestamp_Valid
Forecast_Horizon	Stundenanzahl zwischen Timestamp_Init und Timestamp_Valid
Forecast_Klasse	1 für Forecast_Horizon im Intervall [25, 48] 2 für Forecast_Horizon im Intervall [7, 24]

Tabelle 4.1: Variablen

<sup>8</sup>Quelle: <http://www.ubimet.com/de-AT/unternehmen/geschichte> am 01.02.2017

Die Wetterprognosen von Ubimet werden immer an den gleichen Koordinaten durchgeführt, die etwa vier Kilometer auseinander liegen. An jedem dieser Punkte ist der geografische Längen- und Breitengrad (Variablen `lon` und `lat`) bekannt. Mit den 165 angegebenen geografischen Längen und den 96 geografischen Breiten ergibt sich eine rechteckige Fläche die Österreich beinhaltet (siehe Abbildung 4.1).

Um jeden dieser geografischen Koordinatenpunkte wird ein Quadrat definiert, wobei `(lon, lat)` den Mittelpunkt des Quadrates beschreibt. Die Quadrate werden der Einfachheit halber mit einer fortlaufenden Identifikationsnummer `id` ausgestattet.

Aus den Rohdaten ergeben sich 15840 Quadrate, die eine Fläche von etwa 266 000 km<sup>2</sup> abdecken. Da nur Messpunkte in Österreich betrachtet werden sollen wurde mit dem Punkt-in-Polygon-Test von Jordan (siehe Kapitel 4.3.2) die Anzahl der Quadrate auf 4967 (siehe Abbildung 4.2) reduziert.

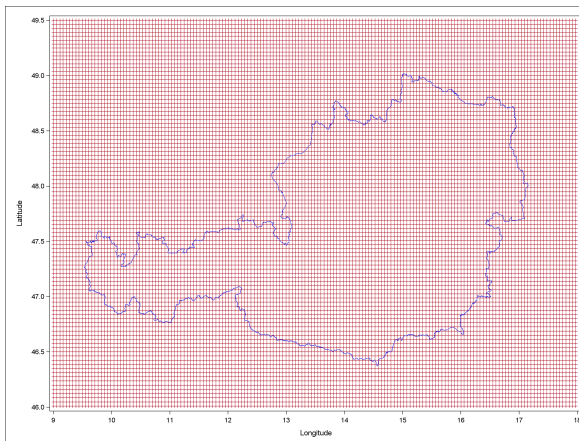


Abb. 4.1

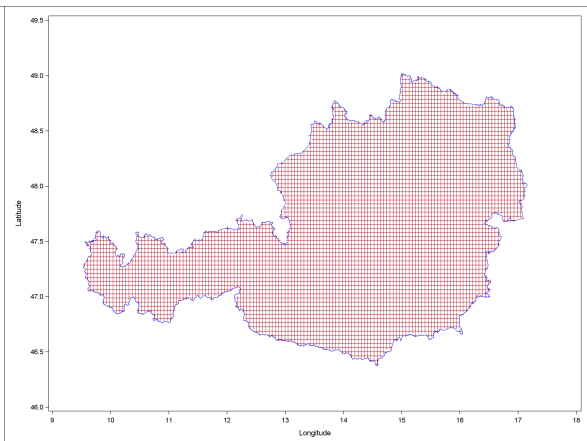


Abb. 4.2

Die Windprognose für den Sturm Emma beginnt am 29. Februar 2008 um 12:00 und wird im Zwölf-Stundentakt wiederholt bis zur letzten Prognose am 02. März 2008 um 12:00. Es ergeben sich fünf verschiedene Zeitpunkte zu denen Prognosen angefertigt werden, diese Zeitpunkte werden in den Rohdaten mit `Timestamp_Init` bezeichnet.

Zu jedem `Timestamp_Init` wird auch der Zeitpunkt angegeben für den die Prognose gemacht wurde, dieser zweite Zeitpunkt wird `Timestamp_Valid` genannt. Die verschiedenen `Timestamp_Valid` Zeiten liegen jeweils eine Stunde auseinander.

Um die unterschiedlichen Prognosen besser unterscheiden zu können wurde der Prognosehorizont `Forecast_Horizon` eingeführt, der den Abstand zwischen `Timestamp_Init` und `Timestamp_Valid` angibt.

Die Tabelle 4.2 beschreibt die verschiedenen Prognosehorizonte für den Sturm Emma. Außerdem kann ihr entnommen werden, dass für jedes Quadrat 132 verschiedene Wetterprognosen angefertigt wurden.

Timestamp_Init	Min. Prognosehorizont	Max. Prognosehorizont	Anz. Prognosen
29FEB2008:12:00	13	48	36
01MAR2008:00:00	7	48	42
01MAR2008:12:00	7	36	30
02MAR2008:00:00	7	24	18
02MAR2008:12:00	7	12	6
<b>Summe</b>	-	-	<b>132</b>

Tabelle 4.2: Anzahl der Prognosen pro Quadrat (Sturm Emma)

Für jede einzelne Prognose beschreibt die Variable `wind_i` die prognostizierte mittlere Windgeschwindigkeit in zehn Metern Höhe und wird in Metern pro Sekunde angegeben. Die prognostizierte Geschwindigkeit der Windböen trägt im Datensatz den Namen `windgust_i` und wird ebenfalls in Metern pro Sekunde angegeben.

Um die Güte der Wetterprognose zu bewerten wurde die Variable `Forecast_Klasse` hinzugefügt, sie teilt den Datensatz in zwei Teile. Prognosen die einen `Forecast_Horizon` von 25-48 Stunden haben, fallen unter die `Forecast_Klasse 1`. `Forecast_Klasse 2` umfasst alle Prognosen die 7-24 Stunden vor `Timestamp_Valid` gemacht wurden.

Für jede dieser Forecast-Klassen werden die Prognosedaten zusammengefasst, sodass es für jedes Quadrat nur eine einzige Windprognose gibt. Für die Variablen `wind_i` und `windgust_i` werden hierbei jeweils das Maximum und das arithmetische Mittel aller Prognosewerte für die jeweilige Forecast-Klasse ausgegeben. Die Abbildung 4.3 zeigt als Beispiel die ersten 15 Zeilen des Datensatzes von Emma.

	Lon	Lat	zip	Seehöhe	id	Timestamp_Init	Wind_i	Windgust_i	Timestamp_Valid	ForecastHorizon	Forecast_Klasse
1	9.551	47.286	6800	426	1	29FEB08:12:00:00	2.50	7.30	01MAR08:01:00:00	13	2
2	9.551	47.286	6800	426	1	29FEB08:12:00:00	2.50	8.40	01MAR08:02:00:00	14	2
3	9.551	47.286	6800	426	1	29FEB08:12:00:00	3.50	9.70	01MAR08:03:00:00	15	2
4	9.551	47.286	6800	426	1	29FEB08:12:00:00	5.20	10.70	01MAR08:04:00:00	16	2
5	9.551	47.286	6800	426	1	29FEB08:12:00:00	5.70	11.80	01MAR08:05:00:00	17	2
6	9.551	47.286	6800	426	1	29FEB08:12:00:00	4.90	17.80	01MAR08:06:00:00	18	2
7	9.551	47.286	6800	426	1	29FEB08:12:00:00	4.30	22.50	01MAR08:07:00:00	19	2
8	9.551	47.286	6800	426	1	29FEB08:12:00:00	2.50	22.00	01MAR08:08:00:00	20	2
9	9.551	47.286	6800	426	1	29FEB08:12:00:00	5.70	23.40	01MAR08:09:00:00	21	2
10	9.551	47.286	6800	426	1	29FEB08:12:00:00	8.30	22.10	01MAR08:10:00:00	22	2
11	9.551	47.286	6800	426	1	29FEB08:12:00:00	6.30	17.20	01MAR08:11:00:00	23	2
12	9.551	47.286	6800	426	1	29FEB08:12:00:00	9.90	25.30	01MAR08:12:00:00	24	2
13	9.551	47.286	6800	426	1	29FEB08:12:00:00	6.90	17.30	01MAR08:13:00:00	25	1
14	9.551	47.286	6800	426	1	29FEB08:12:00:00	7.40	16.80	01MAR08:14:00:00	26	1
15	9.551	47.286	6800	426	1	29FEB08:12:00:00	6.70	17.20	01MAR08:15:00:00	27	1

Abb. 4.3: Beispieldaten

### 4.3.2 Punkt-in-Polygon-Test von Jordan

Um nicht alle angegebenen Quadrate zu betrachten, sondern nur die deren Mittelpunkt in Österreich liegt wurde der *Punkt-in-Polygon-Test* nach JORDAN<sup>9</sup> durchgeführt.

Für jedes Postleitzahlgebiet Österreichs wird der Test separat durchgeführt, da so jeder Gitterpunkt zusätzlich einer Postleitzahl zugeordnet werden kann.

Um zu überprüfen ob sich ein Punkt in einem bestimmten Postleitzahlgebiet befindet wird eine Strecke von diesem Koordinatenpunkt zu einem Basispunkt  $B$  gezogen, der sich außerhalb des Polygons befindet. Meistens wird die Ursprungscoordinate  $(0,0)$  gewählt. Nun werden die Schnittpunkte der Strecke mit den Kanten des Polygons gezählt, wobei vier Fälle unterschieden werden können:

1. kein Schnittpunkt
2. gerade Anzahl von Schnittpunkten
3. ungerade Anzahl von Schnittpunkten
4. unendlich viele Schnittpunkte

Gibt es keinen Schnittpunkt oder eine gerade Anzahl von Schnittpunkten, so liegt der Koordinatenpunkt außerhalb des betrachteten Polygons. Ist die Anzahl der Schnittpunkte ungerade, liegt der Punkt innerhalb des Polygons.

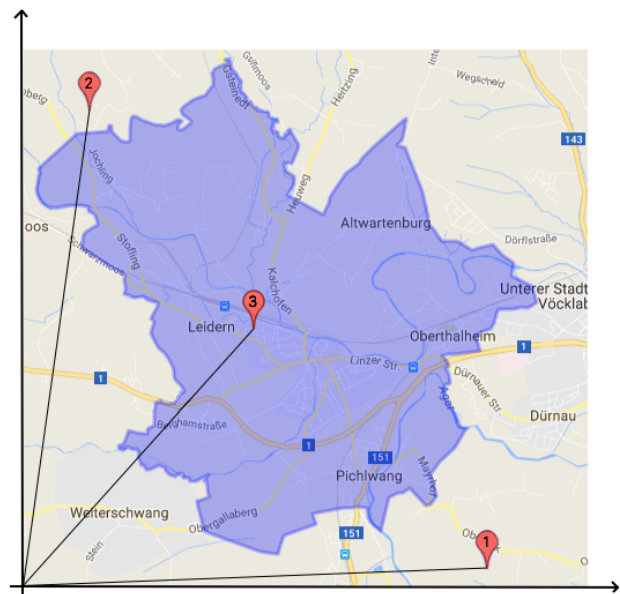


Abb. 4.4: Punkt-in-Polygon-Test

Im Fall von unendlich vielen Schnittpunkten verläuft die Gerade direkt auf einer Kante des Polygons. Der Test muss dann mit einer anderen Gerade wiederholt werden.

Um die Laufzeit des SAS-Programmes gering zu halten, erfolgte die Zuteilung der Koordinatenpunkte zu den Postleitzahlgebieten zuerst nur grob und dann erst exakt mit dem Punkt-in-Polygon-Test. Es wurde in einem ersten Schritt jedes Postleitzahlgebiet durch zwei Rechtecke angenähert, dazu wurden jeweils die maximalen und minimalen Längen- und Breitengrade des Gebietes als Eckpunkte angenommen (siehe Abbildung 4.5). Nun kann jeder Koordinatenpunkt  $P = (\text{lon}, \text{lat})$  verhältnismäßig einfach den Rechtecken zugeordnet werden und zwar indem überprüft wird ob die Bedingung  $\text{min\_long} \leq \text{lon} < \text{max\_long}$  und  $\text{min\_lat} \leq \text{lat} < \text{max\_lat}$  gilt.

<sup>9</sup>Camille JORDAN (\*1838 in Lyon; †1922 in Paris) war ein französischer Mathematiker.

Für alle Punkte  $P$  die nur genau einem Rechteck zugeordnet werden, muss der Punkt-in-Polygon-Test nicht mehr durchgeführt werden, da für diese Punkte das Postleitzahlgebiet schon eindeutig ist. Für alle Punkte denen mehr als ein Rechteck zugeteilt wurde, wird nun der exakte Test durchgeführt.

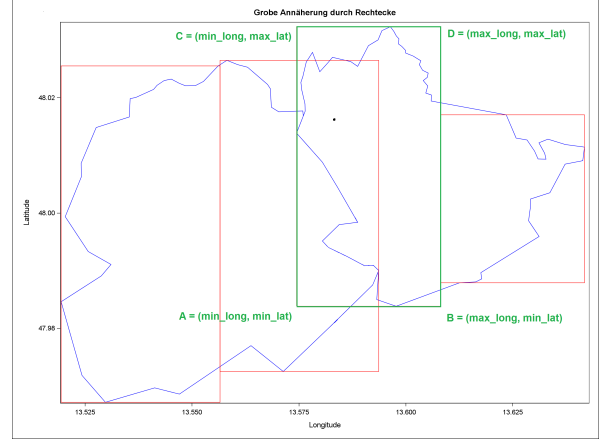


Abb. 4.5: Annäherung der PLZ-Gebiete

Für den exakten Test werden nun zwei Geradengleichungen aufgestellt. Die Gerade  $g$  verläuft durch den Basispunkt  $B = (x_B, y_B)$  und den Koordinatenpunkt  $P = (x_P, y_P)$ . Die Gerade  $h$  wird durch jeweils zwei aufeinanderfolgende Eckpunkte  $E = (x_E, y_E)$  und  $F = (x_F, y_F)$  des betrachteten Postleitzahlgebietes gezogen. Der Schnittpunkt  $S = (x_S, y_S)$  der beiden Geraden ergibt sich dabei aus:

$$x_S = \frac{(x_F - x_E)(x_P y_B - x_B y_P) - (x_P - x_B)(x_F y_E - x_E y_F)}{(y_F - y_E)(x_P - x_B) - (y_P - y_B)(x_F - x_E)}$$

$$y_S = \frac{(y_B - y_P)(x_F y_E - x_E y_F) - (y_E - y_F)(x_P y_B - x_B y_P)}{(y_F - y_E)(x_P - x_B) - (y_P - y_B)(x_F - x_E)}$$

Der Schnittpunkt  $S$  wird nur dann gezählt, wenn sich nicht nur die Geraden  $g$  und  $h$  schneiden, sondern auch die Strecken  $\overrightarrow{BP}$  und  $\overrightarrow{EF}$ . Dies wird mit den folgenden vier Bedingungen erreicht:

$$\begin{aligned} \min(x_B, x_P) &\leq x_S < \max(x_B, x_P) & \min(y_B, y_P) &\leq y_S < \max(y_B, y_P) \\ \min(x_E, x_F) &\leq x_S < \max(x_E, x_F) & \min(y_E, y_F) &\leq y_S < \max(y_E, y_F) \end{aligned}$$

Anschließend werden die Schnittpunkte pro Koordinatenpunkt  $P$  und Postleitzahlgebiet aufsummiert. So kann nun die auf der Vorderseite beschriebene Fallunterscheidung durchgeführt werden und man erhält eine eindeutige Gebietszuordnung für jeden Punkt als Endergebnis.

## 4.4 Sturmkarten

Die folgenden zwei Abbildungen zeigen die von dem Wetterdienst ZAMG gemessenen Windgeschwindigkeiten für die Stürme Paula und Emma.

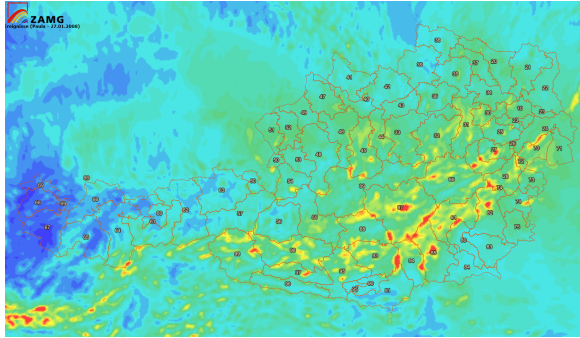


Abb. 4.6: Paula am 27.01.2008

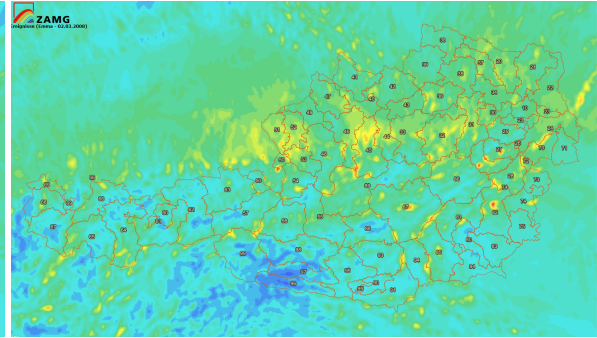
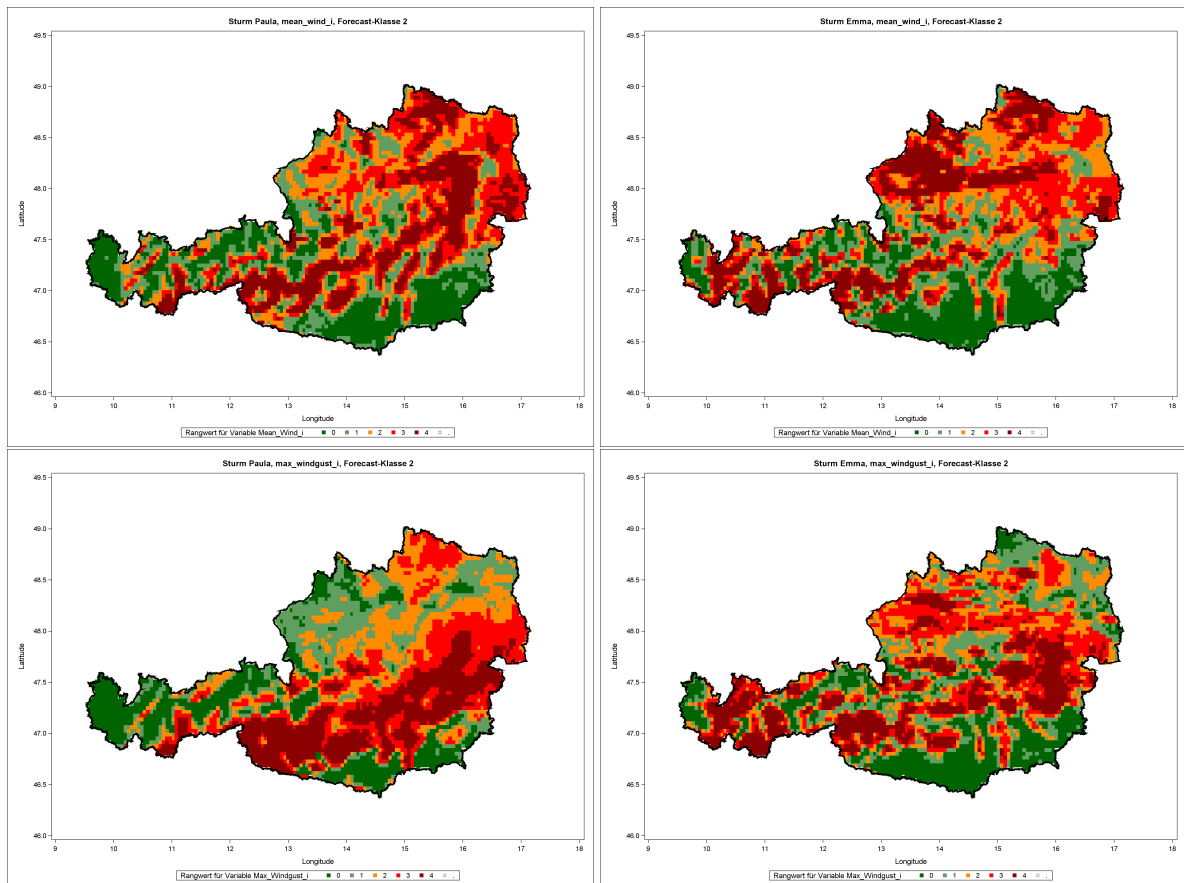


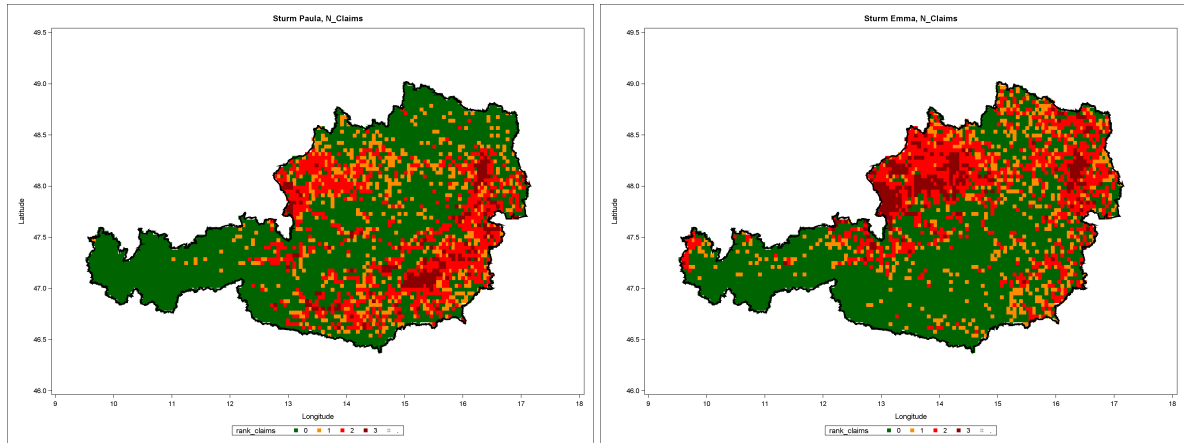
Abb. 4.7: Emma am 02.03.2008<sup>10</sup>

Im Vergleich dazu zeigen die folgenden vier Abbildungen die Windprognosewerte von Ubiomet. Auf den linken Sturmkarten werden die mittlere Windgeschwindigkeit `mean_wind_i` (oben) und die maximal prognostizierten Windspitzen `max_windgust_i` (unten) für den Sturm Paula dargestellt. Die rechten Sturmkarten zeigen die gleichen zwei Variablen für den Sturm Emma, wobei jeweils `Forecast_Klasse 2` betrachtet wird.



<sup>10</sup>Bildquelle: [www.hora.gv.at](http://www.hora.gv.at) am 05.02.2017

Die beiden unteren Abbildungen zeigen die Anzahl der Sturmschäden `n_claims` pro Quadrat. Auf den ersten Blick wirkt es, als ob `n_claims` eher mit der Variable `max_windgust_i` korreliert. Welche Variablen mit der gesuchten Zielgröße `n_claims` am höchsten korrelieren und so für die Regression am ehesten geeignet sind, wird im nächsten Kapitel behandelt.





# Kapitel 5

## Regressionsmodelle

In diesem Kapitel werden die vorher beschriebenen Daten in Anlehnung an verschiedene Regressionsmodelle analysiert. Es gibt mehrere verschiedene Modelle anhand denen Regressionen durchgeführt werden können, hier werden *Lineare Regressionsmodelle* und *Generalisierte Lineare Modelle* behandelt. Die theoretischen Grundlagen zu den linearen Modellen stammen aus den Quellen [2], [3], [4] und [7]. Für die generalisierten linearen Modelle werden die im Literaturverzeichnis angegebenen Quellen [4], [5], [9] und [14] verwendet.

Bei der Regressionsanalyse wird mit einer mathematischen Gleichung der funktionale Zusammenhang zwischen einer abhängigen Variablen und einer oder mehreren unabhängigen Variablen beschrieben. In dieser Diplomarbeit soll die aus einem Sturm resultierende Schadenanzahl durch die Windgeschwindigkeit erklärt werden. In allen Modellen wird die abhängige Zielvariable (Schadenanzahl) mit  $y$  bezeichnet und die erklärende Variable bzw. der Regressor (z.B. Windgeschwindigkeit) mit  $x$ .

### 5.1 Datenanalyse

Der erste Schritt bei der Durchführung einer Regressionsanalyse besteht darin, sich einen Überblick über die Variablen des Datensatzes zu verschaffen. Dafür werden univariate Statistiken, deskriptive Hilfsmittel (z. B. Lagemaße, Streuungsmaße) und grafische Darstellungen (z.B. Histogramme, Streudiagramme) verwendet.

Für die Regressionsanalyse wird mit den Daten des Sturmes Emma gerechnet, getestet und bewertet werden die Regressionsmodelle in Kapitel 6 anhand der Daten des Sturmes

Paula. Die verschiedenen Variablen werden in Kapitel 4.3 genau beschrieben und erklärt. Um einen ersten Überblick über die Daten des Sturmes Emma zu erhalten, werden in der folgenden Tabelle 5.1 einige deskriptive Kennzahlen angegeben, wobei sich die Prognosewerte auf Forecastklasse 1 beziehen. Die Tabelle wird pro Quadrat betrachtet, so ist zum Beispiel die größte Schadenanzahl die in einem Quadrat vorkommt 223.

	n_claims	n_pol	mean_wind	max_windgust
Arithmetisches Mittel	2,33	93,97	8,80	27,29
Varianz	63,24	207691,50	7,16	10,31
Standardabweichung	7,95	455,73	2,68	3,21
Minimum	0	0	2,20	16,10
Maximum	223	17172	19,43	39,90

Tabelle 5.1: Deskriptive Kennzahlen des Sturmes Emma

Weiters wird die Verteilung der Variablen anhand von Histogrammen grafisch dargestellt. Es ist leicht zu erkennen, dass sich in der folgenden Abbildung 5.1 bei den Variablen `n_claims` und `n_pol` über 90% der Datenmenge nur auf die ersten beiden Werte der  $x$ -Achse verteilt. Einige wenige Maximalwerte verzerren das Bild des Datensatzes, wie zum Beispiel die Quadrate mit `id = 4563` und `id = 4610`. Diese zwei Quadrate befinden sich inmitten des Stadtzentrums von Wien und weisen trotz geringer Schadenanzahlen über doppelt so viele Polizzen auf wie irgendwo sonst in Österreich. So verfälschen sie die Regressionsanalyse und werden deshalb aus dem Analyse-Datensatz ausgenommen.

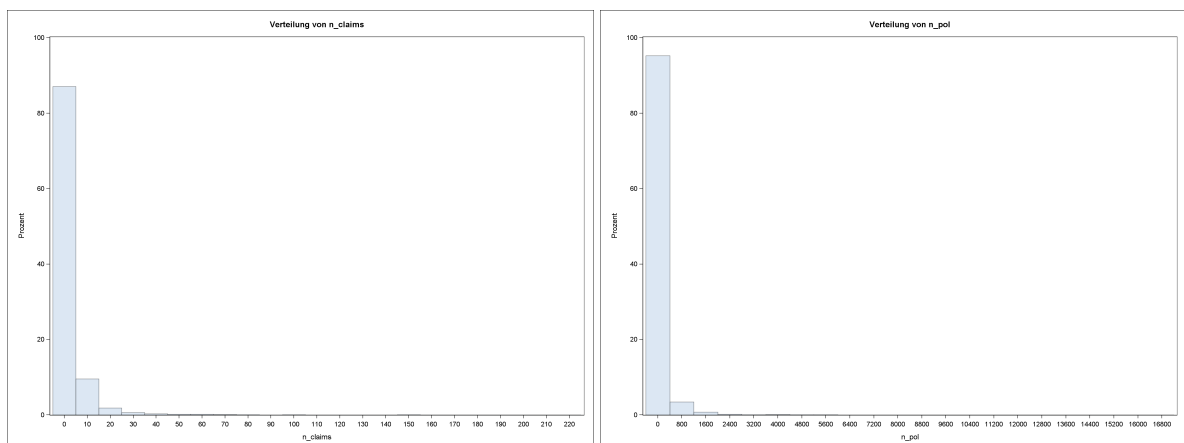


Abb. 5.1: Histogramme der Variablen `n_claims` und `n_pol`

In der folgenden Abbildung 5.2 sind die Histogramme und Kerndichteschätzer der Variablen `mean_wind_i` und `max_windgust_i` aufgezeigt.

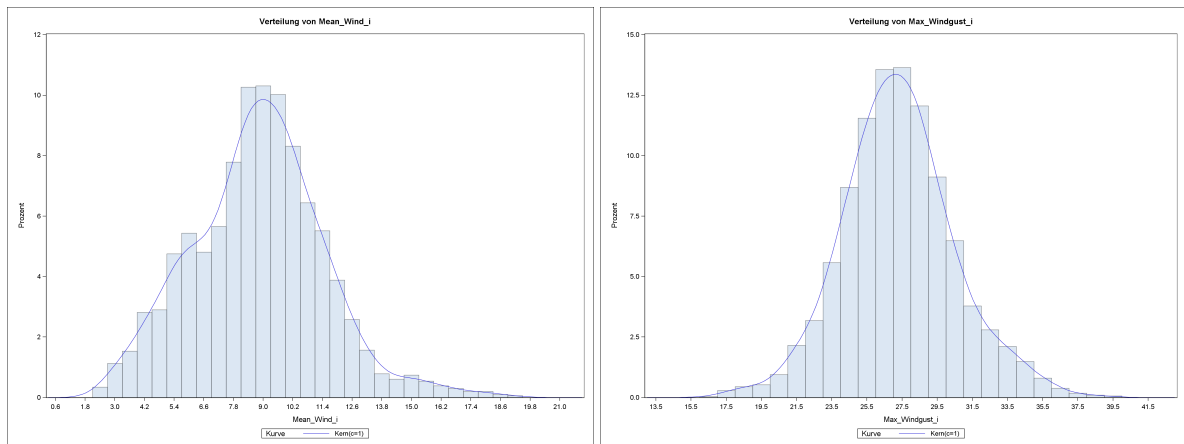


Abb. 5.2: Histogramme und Kerndichteschätzer

### 5.1.1 Zusammenhangsanalyse

In einem zweiten Schritt wird der Zusammenhang zwischen der Zielgröße und den erklärenden Variablen untersucht. Damit kann ein erster Überblick über die Art und Stärke des Zusammenhangs gewonnen werden. Hier beschränkt man sich auf zweidimensionale Zusammenhangsanalysen zwischen der Zielgröße und jeweils einer der erklärenden Variablen. Am ehesten wird eine lineare Relation zwischen `n_claims` und `n_pol` vermutet, also eine steigende Schadenanzahl je höher die Polizzenanzahl ist. Einerseits können Quadrate die keine versicherten Sachen beinhalten auch keine Schäden aufweisen und andererseits ist es eher wahrscheinlich, dass bei einem Sturmereignis an

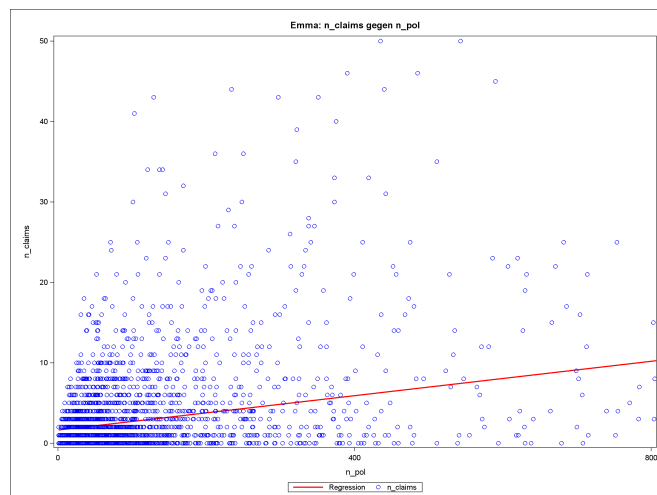


Abb. 5.3: `n_claims` gegen `n_pol`

Orten mit großer Polizzenanzahl auch dementsprechend viele Schäden entstehen. Um das Verhältnis zu verdeutlichen wurde in Abbildung 5.3 der lineare Zusammenhang durch eine rote Gerade gekennzeichnet. Die Abbildung zeigt nur einen Ausschnitt des Analyse-Datensatzes und wurde von Ausreißern bereinigt.

Bei großem Stichprobenumfang, wie hier mit fast 5000 Quadraten sind Streudiagramme meist wenig informativ. Über die Zusammenhänge zwischen der Zielvariable `n_claims` und den Windvariablen lassen sich so nur ungenaue Aussagen treffen. Da in Abbildung 5.3 der lineare Zusammenhang zwischen `n_claims` und `n_pol` schon verdeutlicht wurde, wird für die folgenden Streudiagramme mit der sogenannten *Schadenfrequenz* gearbeitet. Die Schadenfrequenz beschreibt das Verhältnis der Schadenanzahl bezogen auf die Anzahl der Polizzen pro Quadrat.

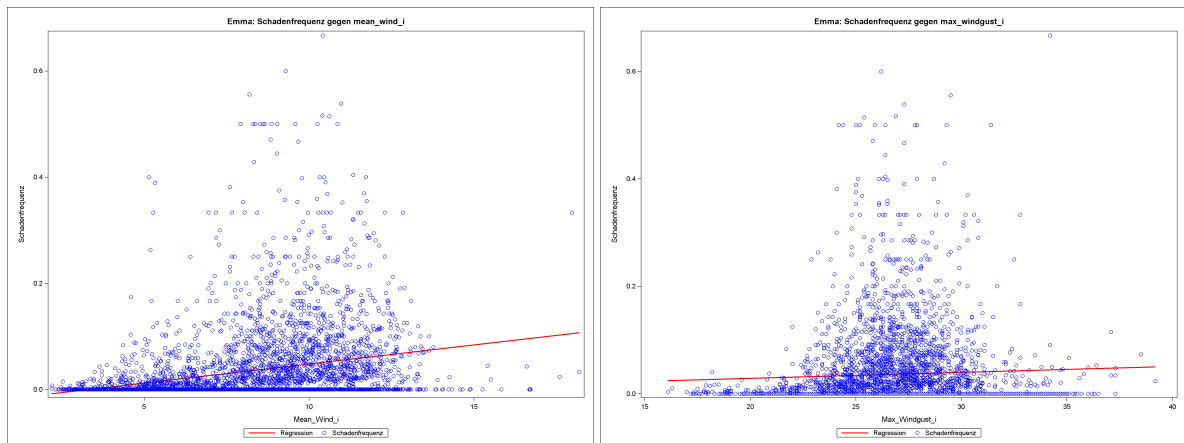


Abb. 5.4: Streudiagramme `mean_wind_i` und `max_windgust_i`

Die Relationen sind wieder mit roten Geraden gekennzeichnet. Die Variable `mean_wind_i` zeigt hier einen deutlichen linearen Zusammenhang, im Gegensatz dazu verläuft die Gerade für die Variable `max_windgust_i` eher waagrecht.

Leider gibt es nur wenige mathematische Instrumente um den Zusammenhang der verschiedenen Variablen schon vor einer Regressionsanalyse zu berechnen. Deshalb wird üblicherweise zuerst eine Regressionsrechnung durchgeführt und erst danach der Einfluss der verschiedenen erklärenden Variablen auf die Zielvariable bewertet.

Um die Güte der Anpassung an die Daten zu messen wird für jede Regression das *korrigierte Bestimmtheitsmaß*  $\bar{R}^2$  (Formel 5.1) und das *Informationskriterium nach Akaike AIC* (Formel 5.2) berechnet.

## 5.2 Lineare Regressionsmodelle

Um eine Regressionsanalyse durchzuführen wird der Einfluss der Regressoren  $x_i$  auf den Mittelwert der Zielvariable  $y$  untersucht. Es wird der bedingte Erwartungswert  $\mathbb{E}[y|x_1, \dots, x_k]$  modelliert, der als Funktion  $f$  der erklärenden Variablen  $x_1, \dots, x_k$  angegeben wird.

Die Zielvariable  $y$  kann in zwei Komponenten aufgespalten werden:

$$y = f(x_1, \dots, x_k) + \varepsilon$$

Hier beschreibt  $\varepsilon$  die zufällige, nicht von den Regressoren erklärte Abweichung vom Erwartungswert. Die Funktion  $f$  wird *systematische Komponente* genannt. Sie soll geschätzt und vom Störterm  $\varepsilon$  getrennt werden.

Am bekanntesten ist die Klasse der linearen Regressionsmodelle

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$$

in denen unterstellt wird, dass die Funktion  $f$  linear ist.

### 5.2.1 Das einfache lineare Regressionsmodell

Bei dem Modell der einfachen linearen Regression  $y = \beta_0 + \beta_1 x + \varepsilon$  wird  $y$  nur durch einen einzigen Regressor  $x$  erklärt. Im Fall der 4963 Quadrate (vier Quadrate wurden wegen Ausreißern ausgenommen) ergeben sich die Gleichungen

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, 4963.$$

Die beiden unbekannten Parameter  $\beta_0$  und  $\beta_1$  werden mit der *Methode der kleinsten Quadrate* (siehe Kapitel 5.2.2) geschätzt. Die daraus resultierende Regressionsgerade wird mit  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$  bezeichnet.

Der Schätzfehler  $y - \hat{y} = y - \hat{\beta}_0 - \hat{\beta}_1 x_1$  heißt *Residuum* und wird mit  $\hat{\varepsilon}$  gekennzeichnet.

Das lineare Regressionsmodell wird vor allem dann verwendet, wenn die Zielvariable  $y$  stetig und zumindest approximativ normalverteilt ist. Es wird angenommen, dass die Fehlervariablen  $\varepsilon_i$  unabhängig und identisch mit  $\mathbb{E}[\varepsilon_i] = 0$  und  $\mathbb{V}[\varepsilon_i] = \sigma^2$  verteilt sind. So sind dann auch die Zielvariablen  $y_i$  normalverteilt mit

$$\mathbb{E}[y_i] = \beta_0 + \beta_1 x_i \qquad \mathbb{V}[y_i] = \sigma^2.$$

### Gütekriterien

Um verschiedene Regressionsmodelle vergleichen zu können wird üblicherweise das Bestimmtheitsmaß  $R^2$  definiert. Es beschreibt den Anteil, der durch das Modell erklärten quadrierten Abweichungen der  $y$ -Werte vom Mittelwert  $\bar{y}$  und nimmt Werte zwischen 0 und 1 an. Im Extremfall  $R^2 = 1$  sind sämtliche Residuen  $\hat{\varepsilon}$  gleich Null und die Anpassung an die Daten ist perfekt.

Durch die Hinzunahme von neuen Regressoren lässt sich der Wert von  $R^2$  beliebig nahe an 1 bringen und täuscht so eine gute Anpassung vor. Darum wird in dieser Diplomarbeit das korrigierte Bestimmtheitsmaß  $\bar{R}^2$  (Formel 5.1) verwendet.  $\bar{R}^2$  „bestraft“ eine zu große Regressorenanzahl und verhindert so zu komplexe Modelle. Die Definitionen für  $n$  Beobachtungen und  $p$  Regressoren lauten:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \qquad \bar{R}^2 = 1 - \frac{n-1}{n-p} \cdot (1 - R^2) \qquad (5.1)$$

Als weiteres Gütekriterium wird das Informationskriterium nach Akaike  $AIC$  angegeben. Das im Rahmen der Maximum-Likelihood-Schätzung (siehe Kapitel 5.3.2) am häufigsten verwendete Modellwahlkriterium ist definiert durch

$$AIC = -2 \cdot l(\hat{\beta}_M, \hat{\sigma}^2) + 2|M + 1| \qquad (5.2)$$

wobei  $l(\hat{\beta}_M, \hat{\sigma}^2)$  der maximale Wert der Log-Likelihoodfunktion und  $M$  die Anzahl der Parameter ist. Bei der Wahl zwischen verschiedenen Modellen wird das Modell mit dem kleinsten AIC-Wert bevorzugt.

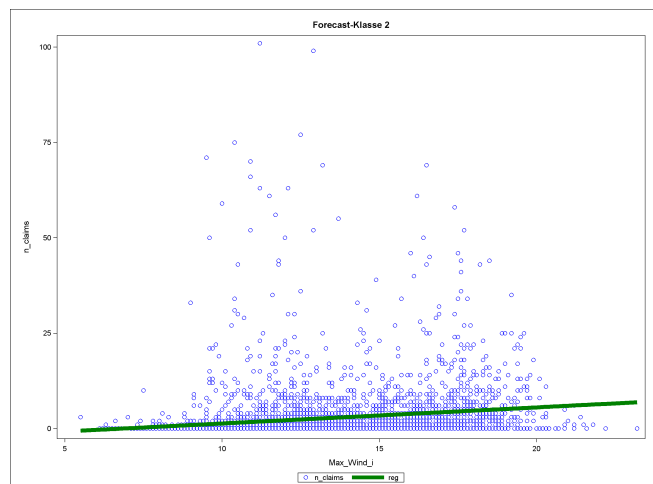
Anhand der zwei eingeführten Forecast-Klassen kann nun verfolgt werden, ob die nach und nach exakter werdenden Wetterprognosen auch zu besseren Ergebnissen in der Regressionsanalyse führen.

Die folgende Tabelle zeigt die geschätzten Parameter und die Gütemaße für das einfach lineare Regressionsmodell:

	Forecast-Klasse 1			
Regressor	$\hat{\beta}_0$	$\hat{\beta}_1$	$\bar{R}^2$	$AIC$
max_windgust_i	5,22159	-0,09113	0,001242	26532,51
mean_windgust_i	1,18157	0,07857	0,001317	26532,21
max_wind_i	-1,83606	0,36207	0,014944	26478,19
mean_wind_i	0,58079	0,26286	0,007885	26506,26
	Forecast-Klasse 2			
Regressor	$\hat{\beta}_0$	$\hat{\beta}_1$	$\bar{R}^2$	$AIC$
max_windgust_i	-1,57402	0,15578	0,004441	26519,89
mean_windgust_i	-2,70403	0,27953	0,017909	26466,34
max_wind_i	-2,86390	0,42126	0,026081	26433,48
mean_wind_i	-0,50604	0,40890	0,018601	26463,56

Tabelle 5.2: Einfache lineare Regression

Der Vergleich der zwei Forecast-Klassen zeigt, dass  $\bar{R}^2$  für Forecast-Klasse 2 in jeder Zeile einen höheren Wert angibt, folglich also besser an die Daten angepasst werden kann. Auch das AIC verbessert sich für alle Regressoren in Forecast-Klasse 2. Je kürzer also der Forecast-Horizont ist, desto besser greift die Regression.



Diese Erkenntnis passt auch zur allgemeinen Annahme, dass die Wetterprognose zwölf Stunden vor dem Sturm schon exakter ist, als noch zwei Tage zuvor. Weiters kann aus der Tabelle herausgelesen werden, dass die Regression mit Variable `max_wind_i` für beide Forecast-Klassen bessere Ergebnisse liefert.

Die Anpassung an die Daten ist trotzdem in allen Fällen sehr gering, was für die einfache lineare Regression jedoch nicht weiter überraschend ist, da hier eben nur eine erklärende Variable berücksichtigt wird.

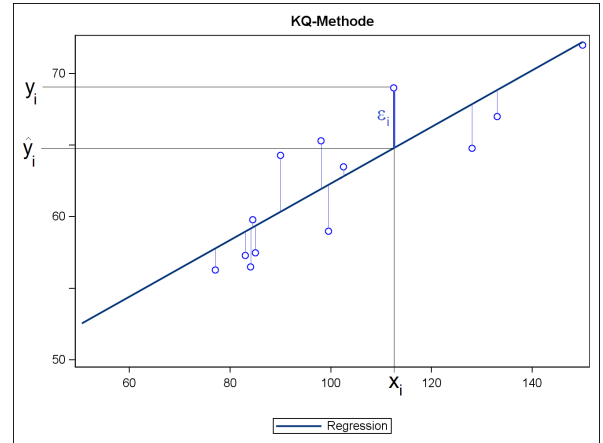
## 5.2.2 Methode der kleinsten Quadrate

Die *Methode der kleinsten Quadrate* oder *KQ-Methode* ist eines der ältesten Schätzkonzepte der Welt. Bereits 1805 wurde diese Schätzmethode von LEGENDRE<sup>11</sup> veröffentlicht.

Um eine Schätzung für die Regressionskoeffizienten  $\beta_0$  und  $\beta_1$  zu erhalten wird die Summe der quadratischen Abweichungen

$$KQ(\beta_0, \beta_1) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

minimiert.



Zur Berechnung des Minimums wird  $KQ(\beta_0, \beta_1)$  einmal nach  $\beta_0$  und einmal nach  $\beta_1$  partiell differenziert und gleich Null gesetzt.

$$\begin{aligned} \frac{\partial KQ(\beta_0, \beta_1)}{\partial \beta_0} &= \sum_{i=1}^n 2 \cdot (y_i - \beta_0 - \beta_1 x_i) \cdot (-1) \stackrel{!}{=} 0 \\ \frac{\partial KQ(\beta_0, \beta_1)}{\partial \beta_1} &= \sum_{i=1}^n 2 \cdot (y_i - \beta_0 - \beta_1 x_i) \cdot (-x_i) \stackrel{!}{=} 0 \end{aligned}$$

Durch Umformung ergeben sich die sogenannten *Normalgleichungen* (5.3 und 5.4) aus denen wiederum die Schätzer  $\hat{\beta}_0$  und  $\hat{\beta}_1$  (5.5 und 5.6) errechnet werden können.

$$\sum_{i=1}^n y_i = n\beta_0 + \beta_1 \sum_{i=1}^n x_i \quad (5.3) \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (5.5)$$

$$\sum_{i=1}^n y_i x_i = \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 \quad (5.4) \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (5.6)$$

In SAS werden die Schätzer für die lineare Regression mit der Prozedur `proc reg` berechnet. Die Prozedur ist nicht auf zwei Parameter beschränkt, es können beliebig viele  $\hat{\beta}_0, \dots, \hat{\beta}_k$  geschätzt werden.

<sup>11</sup>Adrien-Marie LEGENDRE (\*1752 in Paris; †1833 ebenda) war ein französischer Mathematiker.



### 5.2.3 Das multiple lineare Regressionsmodell

Das Modell der *multiplen linearen Regression* mit  $k$  Regressoren und  $n$  Beobachtungen lautet  $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i$  für  $i = 1, \dots, n$ . Die Variable  $x_{ij}$  beschreibt den Wert der  $j$ -ten erklärenden Variable für die  $i$ -te Beobachtung.

In der folgenden Tabelle wird die multiple Regressionsanalyse für die erklärenden Variablen  $x_{i2} = \text{n\_pol}$  und  $x_{i3} = \text{seehoehe}$  durchgeführt. Die erklärende Variable  $x_{i1}$  ist jeweils in der ersten Spalte angegeben.

Forecast-Klasse 1						
Regressor $x_{i1}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\bar{R}^2$	$AIC$
max_windgust_i	-0,71835	0,14571	0,01049	-0,00208	0,26853	25309,87
mean_windgust_i	-1,22389	0,21084	0,01062	-0,00192	0,27582	25270,49
max_wind_i	-4,02982	0,52886	0,01094	-0,00150	0,29514	25164,16
mean_wind_i	-0,88773	0,44848	0,01084	-0,00169	0,28733	25207,49
Forecast-Klasse 2						
Regressor $x_{i1}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\bar{R}^2$	$AIC$
max_windgust_i	-5,48408	0,31436	0,01054	-0,00223	0,28360	25227,99
mean_windgust_i	-4,47679	0,37978	0,01074	-0,00182	0,29769	25149,92
max_wind_i	-4,54623	0,54306	0,01092	-0,00152	0,30662	25099,55
mean_wind_i	-1,57651	0,54777	0,01082	-0,00163	0,29736	25151,75

In Abbildung 5.5 befindet sich die grafische Darstellung der Regressionsanalyse aus obiger Tabelle mit `max_wind_i` für Forecast-Klasse 2. Anstatt einer Geraden kann wegen den verschiedenen Eingangsvariablen in der Abbildung eine Punktwolke erkannt werden.

Die blauen Ringe in Abb. 5.5 stehen für die beobachteten Daten  $y_i$ , in unserem Fall die Schadenanzahl `n_claims`. Die grünen Pluszeichen bilden die mit der Regressionsgleichung geschätzten Daten  $\hat{y}_i$  für die Schadenanzahl ab.

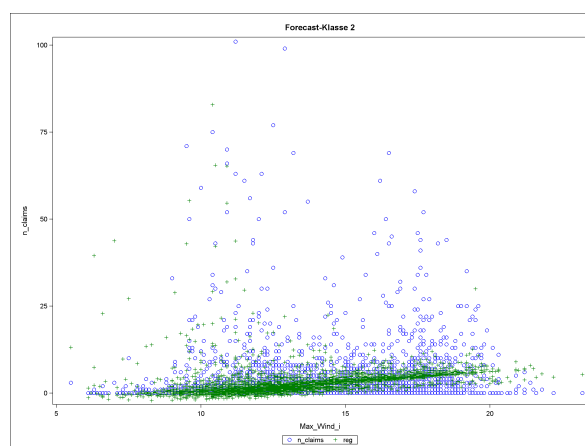


Abb. 5.5: Multiple lineare Regression

### 5.2.4 Polynomiale Regression

Entscheidend für die Anwendung eines linearen Regressionsmodells ist eine *in den Regressionskoeffizienten*  $\beta_i$  lineare Beziehung. Die erklärenden Variablen  $x_{ij}$  und auch die Zielvariable  $y_i$  dürfen dazu geeignet transformiert werden. Deshalb zählt auch die Modellierung durch Polynome zu den linearen Modellen. Für eine einzige erklärende Variable  $x_i$  wird das Schätzmodell

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots \beta_l x_i^l + \varepsilon_i$$

verwendet, wobei  $l$  den Grad des Polynomes angibt.

Es werden in der Regel selten Polynomgrade größer als drei benutzt, da sonst vor allem am Rand des Wertebereichs die Schätzungen sehr instabil werden können (siehe Abbildung 5.6). Trotzdem können die Schätzungen mittels Polynomen besser an die Daten angepasst werden. In der folgenden Tabelle 5.3 kann  $\bar{R}^2$  und AIC für die Polynomgrade 1, 3 und 5 für Forecast-Klasse 2 verglichen werden.

	$l = 1$		$l = 3$		$l = 5$	
Regressor $x_i$	$\bar{R}^2$	AIC	$\bar{R}^2$	AIC	$\bar{R}^2$	AIC
max_windgust_i	0,004441	26519,89	0,012757	26488,91	0,018210	26469,13
mean_windgust_i	0,017909	26466,34	0,025164	26439,18	0,025512	26439,77
max_wind_i	0,026081	26433,48	0,029701	26420,84	0,039120	26384,48
mean_wind_i	0,018601	26463,56	0,019361	26462,52	0,020905	26458,32

Tabelle 5.3: Vergleich der polynomialen Regression für Forecast-Klasse 2

In Abbildung 5.6 wird der Unterschied der verschiedenen Polynome für die Variable `max_wind_i` aufgezeigt.

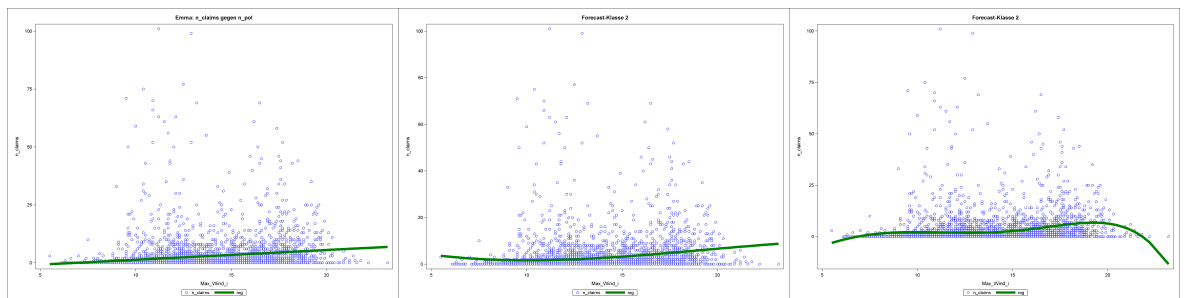


Abb. 5.6: Polynomregression für  $l = 1, 3, 5$

In Tabelle 5.3 liefert sowohl  $\bar{R}^2$  als auch AIC für ein Polynom 5. Grades das beste Ergebnis. Abbildung 5.6 zeigt aber ganz deutlich, dass dieses Polynom vor allem am rechten Rand unerwünschte Werte annimmt. Die Gütekriterien favorisieren also ein ungeeignetes Modell. Diesem Problem könnte begegnet werden, indem auch die Gütekriterien auf einem separaten Testdatensatz berechnet werden. Da aber die Güte der Modelle ohnedies in Kapitel 6 anhand eines Praxistests verglichen werden, wurde hier davon abgesehen.

Mit der polynomialen Regression eröffnen sich nun auch neue Möglichkeiten für die Regression mit mehreren erklärenden Variablen. Jeder Regressor könnte nun durch ein Polynom ersetzt werden. In der Praxis wird jedoch versucht nicht mehr als zwei Variablen durch Polynome zu ersetzen, da die Auswirkungen nicht immer genau einschätzbar sind. Um die polynomiale mit der multiplen Regression vergleichen zu können wurde in der folgenden Tabelle 5.7 das Modell (5.7) gerechnet. Hier wurde anstatt dem Regressor  $x_1$  ein Polynom vom Grad 3 eingesetzt. Die Schätzer  $\beta_i$  werden genauso wie beim multiplen Modell mit der Methode der kleinsten Quadrate (Kapitel 5.2.2) geschätzt.

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_1^2 + \beta_3 \cdot x_1^3 + \beta_4 \cdot \text{n\_pol} + \beta_5 \cdot \text{seehoehe} + \varepsilon \quad (5.7)$$

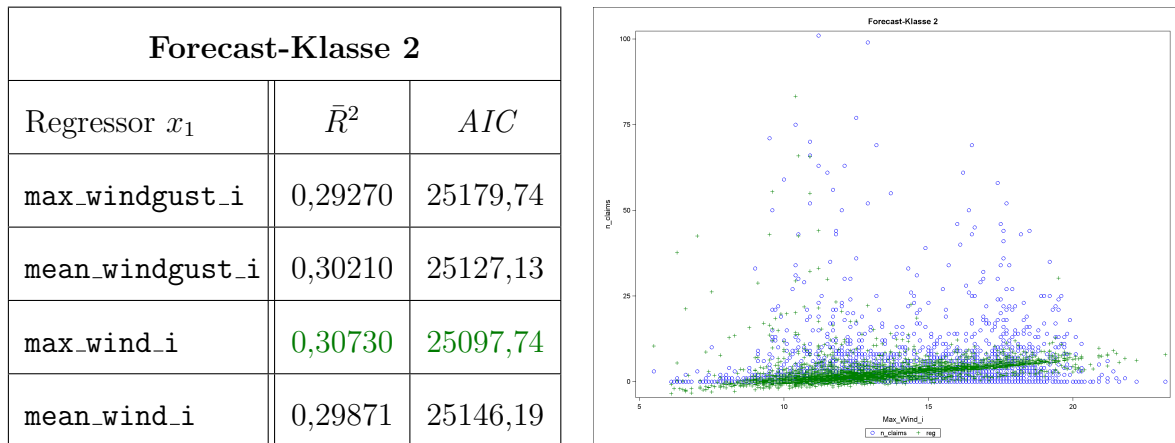


Abb. 5.7: Vergleich der polynomialen Regression für Forecast-Klasse 2

Im direkten Vergleich mit dem multiplen Modell aus Kapitel 5.2.3 erzielt das polynomiale Modell 5.7 zwar geringfügig bessere Ergebnisse, wird aber durch das Polynom um einiges komplexer zu berechnen.

### 5.3 Generalisierte lineare Modelle (GLM)

Lineare Regressionsmodelle eignen sich vor allem für Analysen bei denen die Zielvariable stetig ist und zumindest approximativ durch eine Normalverteilung modelliert werden kann. *Generalisierte lineare Modelle (GLM)* sind eine notwendige Erweiterung der linearen Modelle um auch anders verteilte Zielvariablen betrachten zu können. Da bei einer Modellierung von Schadenanzahlen die Zielvariable nicht stetig ist, sondern nur Werte im Bereich  $\{0, 1, 2, \dots\}$  annimmt, erscheint eine andere Verteilung sinnvoller. Genauer gesagt können mit einem GLM Zielvariablen modelliert werden, deren Verteilung zur Familie der Exponentialverteilungen gehört. Die wichtigsten Mitglieder dieser Verteilungsfamilie sind die Normal-, Binomial-, Poisson- und Gammaverteilungen.

Ein GLM wird durch die drei folgenden Komponenten definiert:

- dem Erwartungswert  $\mathbb{E}[y_i]$
- dem *Prädiktor*  $\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$
- der *Responsefunktion*  $h$ .

So entsteht das Grundmodell eines GLM mit:

$$\mathbb{E}[y_i] = h(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}) = h(\eta_i) \quad (5.8)$$

Die Funktion  $g = h^{-1}(\mathbb{E}[y_i])$  wird *Linkfunktion* genannt. Ein spezifisches GLM ist somit durch den Typ der Exponentialverteilung und durch die Wahl der Link- oder Responsefunktion vollständig charakterisiert.

#### 5.3.1 GLM für Zähldaten - Poisson Regression

Zur Modellierung von Zählvariablen wird am häufigsten die *Poisson-Verteilung* verwendet. Für eine Zufallsvariable  $Y$  die einer Poisson-Verteilung mit dem Parameter  $\lambda$  folgt, in Zeichen  $Y \sim Poi(\lambda)$  lautet die (diskrete) Dichtefunktion für  $y \in \mathbb{N}$

$$\mathbb{P}[Y = y] = \frac{\lambda^y \cdot \exp(-\lambda)}{y!}. \quad (5.9)$$

Eine charakteristische Eigenschaft für die Poisson-Verteilung ist die Gleichheit von Erwartungswert und Varianz  $\mathbb{E}[y] = \mathbb{V}[y] = \lambda$ .

Im bekanntesten Modell für Zähldaten wird die Poisson-Verteilung mit der Linkfunktion *log* verknüpft. Dieses Modell heißt *log-lineares Poisson-Modell*. Der Effekt der

Regressoren auf  $\mathbb{E}[y_i|x_1, \dots, x_k] = \lambda_i$  ist also nun nicht mehr linear-additiv sondern exponentiell-multiplikativ. Für  $g = \log(\lambda_i)$  bzw.  $h = \exp(\eta_i)$  gilt

$$\lambda_i = \exp(\eta_i) = \exp(\beta_0) \cdot \exp(\beta_1 x_{i1}) \cdot \dots \cdot \exp(\beta_k x_{ik}) \quad (5.10)$$

In der praktischen Berechnung eines GLM wird gerne zusätzlich eine *Offset-Variable*  $z_i$  verwendet. Sie beschreibt eine erklärende Variable mit konstantem Koeffizient 1 für jede Beobachtung. Durch die Offset-Variable wird genau genommen keine Anzahl sondern eine Rate geschätzt. Das Modell wird also in  $\log(\lambda_i) = \log(z_i) + \eta_i$  abgeändert. Bei der Berechnung der Schadenanzahl eignet sich die Variable `n_pol` als Offset-Variable. Der Erwartungswert beschreibt nun das Verhältnis der Schadenanzahl zur Polizzenanzahl pro Quadrat. Die Schätzung der Regressionskoeffizienten erfolgt bei GLM nicht mehr mit der KQ-Methode, sondern mit der Maximum-Likelihood-Methode aus Kapitel 5.3.2. In der folgenden Tabelle 5.4 wird das log-lineare Poisson-Modell mit Offset-Variable `n_pol` berechnet. Die Variable  $x_{i2}$  beschreibt den Regressor `max_seehoehe`. Es fällt auf, dass hier für die Seehöhe sehr niedrige Koeffizienten geschätzt werden. Wird das Modell ohne Seehöhe berechnet (hier nicht ausgeführt), führt das auf etwa gleich gute AIC-Werte. Das bedeutet, dass die Seehöhe und die Schadenanzahl in diesem Modell fast garnicht korreliert. Trotzdem fällt das AIC in Tabelle 5.4 besser aus als in den vergleichbaren linearen Modellen.

Forecast-Klasse 1				
Regressor $x_{i1}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	AIC
<code>max_windgust_i</code>	-6,6064	0,1204	-0,0003	22724,63
<code>mean_windgust_i</code>	-6,6797	0,1535	-0,0000	21622,35
<code>max_wind_i</code>	-6,4802	0,2256	-0,0000	20090,16
<code>mean_wind_i</code>	-5,6669	0,2479	0,0000	20054,42
Forecast-Klasse 2				
Regressor $x_{i1}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	AIC
<code>max_windgust_i</code>	-7,9405	0,1587	-0,0002	21985,89
<code>mean_windgust_i</code>	-8,3317	0,2322	0,0002	19588,24
<code>max_wind_i</code>	-6,5027	0,2131	-0,0000	19619,20
<code>mean_wind_i</code>	-5,9348	0,2786	0,0001	19540,80

Tabelle 5.4: Generalisiertes lineares Modell mit Offset-Variable

### 5.3.2 Maximum-Likelihood-Schätzung

Die *Maximum-Likelihood-Schätzung* oder *ML-Schätzung* wird bei Modelltypen mit nichtstetigen Zielvariablen verwendet. Grob erklärt wird der ML-Schätzer so bestimmt, dass die Wahrscheinlichkeit für das Auftreten der beobachteten Stichprobe  $y_1, \dots, y_n$  maximal wird und somit diese Stichprobe möglichst plausibel (*likely*) ist.

Um eine ML-Schätzung für die Regressionskoeffizienten  $\beta_0$  und  $\beta_1$  zu erhalten wird das folgende Poisson-Regressionsmodell betrachtet:

$$\begin{aligned} Y_i &\sim \text{Poi}(\lambda_i) & i = 1, \dots, n \\ \lambda_i &= \exp(\beta_0 + \beta_1 x_i) = \exp(\eta_i) \\ \eta_i &= \beta_0 + \beta_1 x_i \end{aligned}$$

Im ersten Schritt wird die *Likelihoodfunktion*  $L(\beta_0, \beta_1) = \mathbb{P}[Y_1 = y_1 | \lambda] \cdot \dots \cdot \mathbb{P}[Y_n = y_n | \lambda]$  bestimmt. Unter Vernachlässigung der Faktoren die nicht von  $\beta_0$  und  $\beta_1$  abhängen und der Dichteformel 5.9 ergibt sich

$$L(\beta_0, \beta_1) = \prod_{i=1}^n \lambda_i^{y_i} \exp(-\lambda_i).$$

Um im zweiten Schritt die *Log-Likelihoodfunktion*  $l(\beta_0, \beta_1)$  zu bestimmen wird  $L(\beta_0, \beta_1)$  logarithmiert.

$$l(\beta_0, \beta_1) = \sum_{i=1}^n (y_i \log(\lambda_i) - \lambda_i) = \sum_{i=1}^n (y_i(\beta_0 + \beta_1 x_i) - \exp(\beta_0 + \beta_1 x_i))$$

Für die weitere Vorgehensweise werden die partiellen Ableitungen der Log-Likelihoodfunktion  $l_i$  gebildet:

$$\begin{aligned} \frac{\partial l_i(\beta_0, \beta_1)}{\partial \beta_0} &= y_i - \exp(\beta_0 + \beta_1 x_i) = y_i - \lambda_i \\ \frac{\partial l_i(\beta_0, \beta_1)}{\partial \beta_1} &= y_i x_i - \exp(\beta_0 + \beta_1 x_i) x_i = x_i (y_i - \lambda_i) \end{aligned}$$

Mit diesen Ableitungen kann nun die sogenannte *Score-Funktion* bestimmt werden.

Die partiellen Ableitungen der Log-Likelihoodfunktion  $l$  werden nun in den Vektor  $\mathbf{s}(\beta_0, \beta_1)$  gefasst, der die *Score-Funktion* beschreibt.

$$\mathbf{s}(\beta_0, \beta_1) = \begin{pmatrix} \sum_{i=1}^n (y_i - \lambda_i) \\ \sum_{i=1}^n x_i (y_i - \lambda_i) \end{pmatrix} \stackrel{!}{=} 0$$

Im letzten Schritt wird die Log-Likelihood-Funktion maximiert, also die Score-Funktion Null gesetzt. So ergibt sich ein nichtlineares Gleichungssystem, das mit numerischen Verfahren gelöst werden muss. Die in SAS für GLM verwendete Prozedur `proc genmod` berechnet die ML-Schätzer mit einem iterativen Anpassungsprozess.

### 5.3.3 Poisson-Regression mit Polynomen

Wie bei den linearen Modellen können auch bei Poisson-Regressionen erklärende Variablen durch Polynome ersetzt werden. In der Tabelle 5.8 wurde wie in Kapitel 5.2.4 die Regressorvariable  $x_{i1}$  mit einem Polynom dritten Grades transformiert. Auffallend ist hier der Verlauf des Gütekriteriums AIC. Obwohl bei der linearen Regression die Modelle mit der Variable `mean_wind_i` im Vergleich zu den anderen Variablen immer eher schlechter abgeschnitten haben, erzielt hier das Modell den besten Wert bei AIC. Ob das Modell auch im praktischen Test besser abschneidet wird in Kapitel 6 behandelt.

Forecast-Klasse 2	
Regressor $x_{i1}$	AIC
max_windgust_i	21093,76
mean_windgust_i	19383,38
max_wind_i	19230,93
mean_wind_i	18990,78

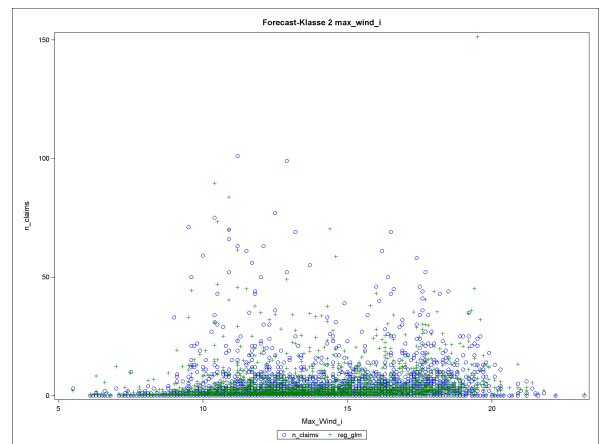


Abb. 5.8: Poisson-Regression mit Polynom ( $l = 3$ )

# Kapitel 6

## Anwendung der Regressionsanalyse

Um die verschiedenen Regressionsmodelle zu testen werden die erklärenden Variablen des Sturmes Paula in die Regressionsgleichungen eingesetzt. Berechnet wurden die Gleichungen also anhand der Daten des Sturmes Emma und nun erfolgt ein Praxistest mit den Daten des Sturmes Paula.

Alle Quadrate aufsummiert, also die berechnete Gesamtanzahl der Schäden wird mit `reg` bezeichnet. Um nicht nur die Gesamtschadenanzahl zu berücksichtigen, sondern auch die Verteilung der Schäden innerhalb von Österreich wird zusätzlich die berechnete Anzahl der Schäden in Oberösterreich `reg_oö` und der Steiermark `reg_stmk` angegeben. Oberösterreich wurde gewählt weil dort Emma die höchsten Schadenanzahlen verzeichnet. Der Sturm Paula richtete in der Steiermark die meisten Schäden an.

Die Differenz zwischen den wahren Schadenanzahlen und den errechneten Werten wird mit  $\Delta$  angegeben. Zum Vergleich sind die wirklichen Werte für Paula angegeben mit:

<code>n_claims</code>	=	8062
<code>n_claims_oö</code>	=	655
<code>n_claims_stmk</code>	=	3131

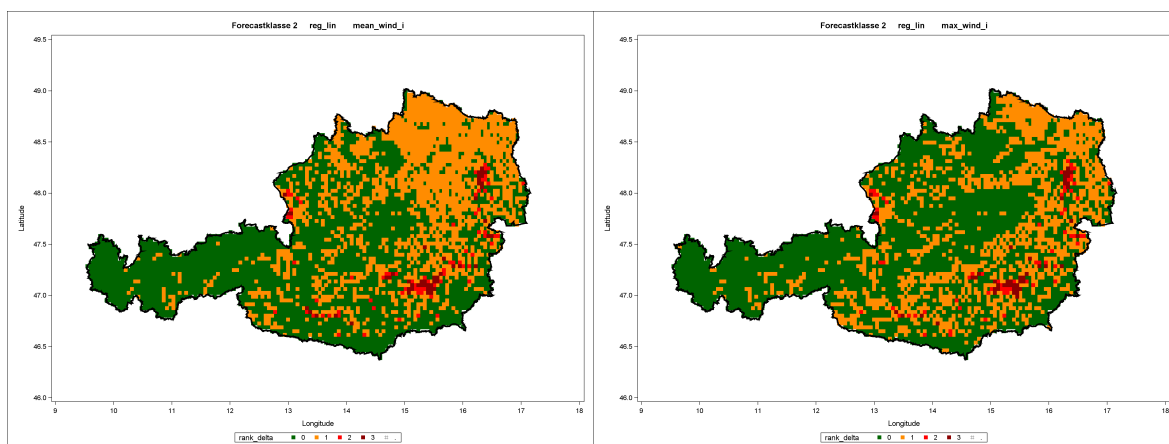
Bei den nun folgenden Vergleichen der linearen Regressionsmodelle muss beachtet werden, dass die Schadenanzahl für ein Quadrat durchaus negative Werte annehmen kann. Es ist also möglich, dass sich Schäden beim aufsummieren gegenseitig aufheben. In der folgenden Tabelle 6.1 werden die Ergebnisse der einfach linearen Regression verglichen.



Forecast-Klasse 1						
Regressor	reg	$\Delta_{Ges}$	reg_oö	$\Delta_{OÖ}$	reg_stmk	$\Delta_{Stmk}$
max_windgust_i	11488	3426	1882	1217	1946	-1185
mean_windgust_i	10321	2259	1538	873	2066	-1065
max_wind_i	8689	627	1197	532	1769	-1362
mean_wind_i	9665	1603	1506	841	1795	-1336
Forecast-Klasse 2						
Regressor	reg	$\Delta_{Ges}$	reg_oö	$\Delta_{OÖ}$	reg_stmk	$\Delta_{Stmk}$
max_windgust_i	9297	1235	1185	520	2122	-1009
mean_windgust_i	6907	-1155	1049	384	1495	-1636
max_wind_i	6958	-1104	958	293	1411	-1720
mean_wind_i	7987	-75	1279	614	1394	-1737

Tabelle 6.1: Vergleich einfache lineare Regression

Die in Kapitel 5.2.1 definierten Modellwahlkriterien  $\bar{R}^2$  und  $AIC$  ergeben bei der einfachen linearen Regression die besten Werte für das Modell mit der Variablen `max_wind_i`. Hier im Praxistest erreicht aber das Modell mit `mean_wind_i` (Forecast-Klasse 2) den mit Abstand besten Wert bei der geschätzten Gesamtschadenanzahl. Nur 75 Schäden weniger wurden mit der Regressionsanalyse berechnet. Bei genauerer Analyse erkennt man jedoch, dass die  $\Delta$ -Werte für Oberösterreich und Steiermark schlechter sind als die der anderen Modelle der gleichen Forecast-Klasse. In der unteren Abbildung sind links die  $\Delta$ -Werte für das Modell `mean_wind_i` und rechts für `max_wind_i` geplottet. Dabei bedeuten grüne Felder  $\Delta < 2$ , gelbe Felder  $\Delta \in [2, 9]$ , rote Felder  $\Delta \in [10, 19]$  und dunkelrote  $\Delta > 20$ . Die grüne Farbkodierung lässt erkennen, dass das rechte Modell die Schadenanzahlen viel besser den richtigen Quadraten zuordnet und somit auch das exaktere Modell ist.



Wenn nicht nur eine sondern mehrere erklärende Variablen im Regressionsmodell betrachtet werden sollen wird eine multiple lineare Regressionsanalyse durchgeführt. In Tabelle 6.2 sind die Ergebnisse für ein Modell mit drei Regressoren dargestellt, wobei  $x_{i2} = \text{n\_pol}$ ,  $x_{i3} = \text{seehoehe}$  und  $x_{i1}$  die in der ersten Spalte angegebene Windvariable beschreibt. Die zweite Tabelle 6.3 zeigt die Ergebnisse der Polynomregression aus Kapitel 5.2.4 mit denselben Regressoren.

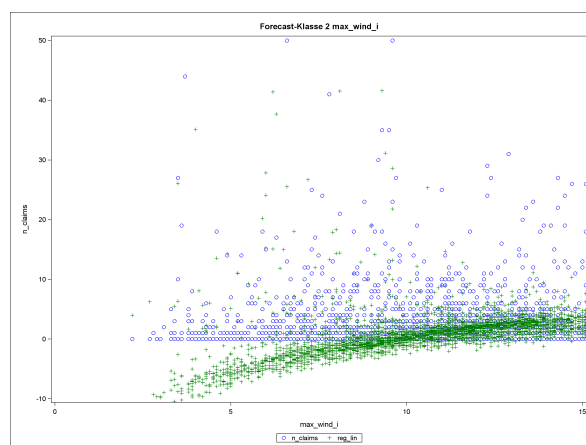
Forecast-Klasse 2						
Regressor	reg	$\Delta_{Ges}$	reg_oö	$\Delta_{O\ddot{O}}$	reg_stmk	$\Delta_{Stmk}$
max_windgust_i	8092	30	962	297	1790	-1341
mean_windgust_i	5932	-2130	1045	380	925	-2206
max_wind_i	6285	-1777	936	271	868	-2263
mean_wind_i	7456	-606	1344	679	808	-2323

Tabelle 6.2: Vergleich multiple lineare Regression

Forecast-Klasse 2						
Regressor	reg	$\Delta_{Ges}$	reg_oö	$\Delta_{O\ddot{O}}$	reg_stmk	$\Delta_{Stmk}$
max_windgust_i	3325	-4737	541	-124	1005	-2126
mean_windgust_i	6597	-1465	915	250	891	-2240
max_wind_i	4632	-3430	894	229	521	-2610
mean_wind_i	7153	-909	1352	687	714	-2417

Tabelle 6.3: Vergleich multiple Polynom-Regression

Wie bei der Einfachregression wurden auch bei der multiplen und polynomialen Regression die besten Werte für  $\bar{R}^2$  und  $AIC$  bei dem Modell mit der Variable `max_wind_i` errechnet (siehe Kapitel 5.2.3). Umso überraschender sind die vergleichsweise schlechten bzw. sehr niedrigen Ergebnisse in der Anwendung. Bei genauerer Betrachtung ergeben sich vor allem bei der polynomialen Regression sehr hohe negative Werte und diese aufsummiert verfälschen natürlich das Ergebnis der gesamten Analyse. In der Abbildung können die berechneten Schadenanzahlen (`reg`) in grün mit den wahren Schäden (`n.claims`) in blau für die Polynomregression verglichen werden.



Um dem Problem mit den negativen Schadenanzahlen aus dem Weg zu gehen, werden nun Regressionen mit generalisierten linearen Modellen betrachtet. Das hier verwendete log-lineare Poisson-Modell errechnet nur positive Werte.

In der anschließenden Tabelle 6.4 können die Schadenanzahlen für ein GLM mit Poisson-Verteilung und Offset-Variable `n_pol` verglichen werden.

Forecast-Klasse 1						
Regressor	reg	$\Delta_{Ges}$	reg_oö	$\Delta_{OÖ}$	reg_stmk	$\Delta_{Stmk}$
max_windgust_i	11419	3357	1167	502	2392	-739
mean_windgust_i	10501	2439	1206	541	1510	-1621
max_wind_i	8003	-59	1057	410	1229	-1902
mean_wind_i	8688	626	1393	728	936	-2195
Forecast-Klasse 2						
Regressor	reg	$\Delta_{Ges}$	reg_oö	$\Delta_{OÖ}$	reg_stmk	$\Delta_{Stmk}$
max_windgust_i	8855	793	757	92	2122	-1009
mean_windgust_i	6551	-1511	650	-15	770	-2361
max_wind_i	6310	-1752	919	254	982	-2149
mean_wind_i	6532	-1530	1043	378	656	-2475

Tabelle 6.4: Vergleich Poisson-Regression mit Offset-Variable `n_pol`

Bei der Poisson-Regression erreichte das Modell mit dem Regressor `mean_wind_i` die besten AIC-Werte. Leider ist aber genau dieses Modell sowohl bei  $\Delta_{OÖ}$  als auch bei  $\Delta_{Stmk}$  das ungenaueste in der jeweiligen Forecast-Klasse. Das bedeutet, dass die Regressionsgleichungen zwar gut an die Daten des Sturmes Emma angepasst werden konnten, aber nicht an die des Sturmes Paula. Um besser analysieren zu können, wo genau das Modell Probleme hat sich an die Validierungsdaten anzupassen, zeigt Abbildung 6.1 alle  $\Delta$ -Werte mittels Farbkodierung. Die dunkel-

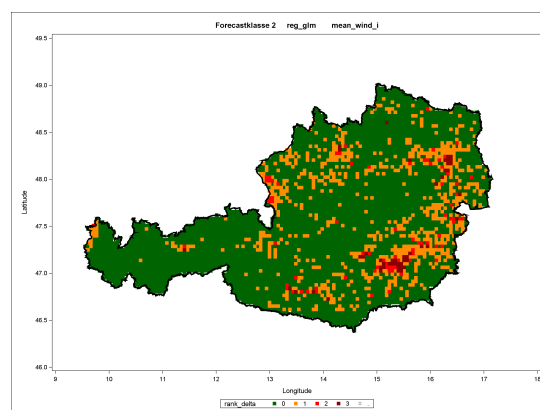


Abb. 6.1:  $\Delta$  von Poisson-Regression

roten Quadrate die  $\Delta > 20$  abbilden sind vor allem in Österreichs größeren Städten zu finden. Die rötlich eingefärbten Städte Wien, Graz, Linz und Salzburg lassen vermuten, dass vor allem Quadrate mit einer hohen Polizzenanzahl sehr fehlerbehaftet sind.

Das letzte berechnete Regressionsmodell beinhaltet ein log-lineares Poisson-Modell mit einer Polynomtransformation. Dabei wird die Windvariable, die in Tabelle 6.5 jeweils in der ersten Spalte zu finden ist, mit einem Polynom vom Grad  $l = 3$  ersetzt. Die Koeffizienten  $\hat{\beta}_i$  werden, wie bei allen generalisierten linearen Modellen, wieder mit der Maximum-Likelihood-Methode geschätzt.

Forecast-Klasse 1						
Regressor	reg	$\Delta_{Ges}$	reg_oö	$\Delta_{O\ddot{O}}$	reg_stmk	$\Delta_{Stmk}$
max_windgust_i	8665	603	951	286	967	-2164
mean_windgust_i	9988	1926	1117	452	854	-2277
max_wind_i	6811	-1251	1070	405	1103	-2028
mean_wind_i	8286	224	1467	802	893	-2238
Forecast-Klasse 2						
Regressor	reg	$\Delta_{Ges}$	reg_oö	$\Delta_{O\ddot{O}}$	reg_stmk	$\Delta_{Stmk}$
max_windgust_i	6176	-1886	387	-278	1303	-1828
mean_windgust_i	6722	-1340	549	-116	761	-2370
max_wind_i	5510	-2552	848	183	956	-2175
mean_wind_i	5997	-2065	979	314	594	-2537

Tabelle 6.5: Vergleich Poisson-Regression mit Polynom und Offset-Variable `n_pol`

Laut dem Gütekriterium AIC ist das hier betrachtete Modell mit der Windvariable `mean_wind_i` (Forecast-Klasse 2) das beste Modell, das in dieser Arbeit ausgeführt wird.

Kein anderes Modell unterschreitet den Wert  $AIC = 18990,78$ . Trotzdem überzeugen die Werte im Praxistest nicht wirklich. Um eine Erklärung zu

	zip	n_claims	reg	delta
1	8010	120	19	101
2	8572	109	4	105
3	8583	112	4	108
4	8160	132	13	119
5	8580	162	7	155

finden, warum das Modell nicht exaktere Ergebnisse liefert, werden die fünf höchsten  $\Delta$ -Werte auf Postleitzahlebene (nicht wie sonst auf Quadratebene) aus dem Datensatz herausgesucht. Aus der obigen Abbildung kann sofort gefolgert werden, dass sich alle ausgegebenen Postleitzahlen im Bundesland Steiermark befinden. Da der Sturm Paula nur mit sehr geringen Windgeschwindigkeiten in der Steiermark sehr viele Schäden anrichtete, kann kein Regressionsmodell diese Schäden wirklich erklären. Auf die Frage wie dieses Problem gelöst werden könnte bzw. weitere Vorgehensmöglichkeiten wird im Kapitel 7 *Fazit und Ausblick* eingegangen.

# Kapitel 7

## Fazit und Ausblick

Ziel der vorliegenden Diplomarbeit war es, einen Zusammenhang zwischen der Windstärke und der Schadenanzahl eines Sturmereignisses herzustellen. Daraus sollte eine Schätzung für Schadenanzahlen zukünftiger Stürme hergeleitet werden.

Zu diesem Zweck wurden Zusammenhangs- und Regressionsanalysen berechnet, die die gefragten Variablen miteinander in Verbindung setzten. In einem durchgeführten Praxistest konnten die Ergebnisse verglichen werden. Dabei ergab sich, dass vor allem mit generalisierten linearen Regressionsgleichungen vergleichsweise exakte Ergebnisse erzielt werden konnten. Zwar kann anhand dieser Ergebnisse der mathematische Zusammenhang zwischen Windgeschwindigkeit und Schadenanzahl gezeigt werden, dennoch gibt die im Praxistest berechnete Schadenanzahl nur einen Richtwert an.

Um definitive Entscheidungen anhand der vorgestellten Modelle treffen zu können, müsste eine größere historische Wetterdatenbasis betrachtet werden.

Die hier gewonnenen Erkenntnisse ließen sich durch weitere Untersuchungen ergänzen. Mit einer umfangreicheren Wetterdatenbasis könnte Österreich in verschiedene Wetterzonen eingeteilt werden. Ein solches Scoring hätte den Vorteil, dass bestimmte geografische Gegebenheiten, wie zum Beispiel die Alpen, besser in die Regressionsanalyse miteinfließen würden.

Eine interessante Weiterführungsmöglichkeit dieses Themas wäre zum Beispiel auch eine Analyse zur Betrugserkennung. In Zeiten des steigenden Versicherungsbetruges werden Instrumente zu deren Erkennung immer gefragter. Anstelle einer Poisson-Verteilung würde eine Binär-Verteilung modelliert werden und so ergäbe sich in weiterer Folge eine Schadenwahrscheinlichkeit pro Quadrat. Sollte ein Schaden in einem Quadrat gemeldet

werden das eine geringe Wahrscheinlichkeit aufweist, könnten bestimmte Maßnahmen eingeleitet werden.

Eine nicht unwesentliche Erweiterung des Themas wäre eine Prognose für die Schadenhöhe. In dieser Diplomarbeit wurde nur die Schadenanzahl prognostiziert, um zum Beispiel am Tag eines Großereignisses schon die Anzahl von Mitarbeitern in einem Call-Center aufstocken zu können. In einem nächsten Schritt könnte die Prognose weitergeführt werden um auch einen Überblick über den voraussichtlichen finanziellen Schaden zu geben.

Abschließend ist noch zu sagen, dass bedingt durch den Klimawandel zukünftig immer häufiger extreme Unwetterereignisse auf Österreich zukommen werden. Eine frühzeitige Vorbereitung auf solche klimabedingten Katastrophen darf nicht nur im Fokus der Versicherungsunternehmen stehen, sondern auch jede Privatperson sollte ihren persönlichen Beitrag zum Klimaschutz leisten. Der Beitrag eines Einzelnen gegen den Klimawandel ist vielleicht klein, aber wer Wind sät wird Sturm ernten.

# Kapitel 8

## Anhang

### 8.1 SAS-Code für ein multiples lineares Modell

```
*Definition der gewünschten Regressoren;
%let var_1 = max_wind_i;
%let var_2 = n_pol;
%let var_3 = max_seehoehe;

*Auswahl der Forecast-Klasse;
data Emma_1;
    set Emma;
    if Forecast_Klasse = 1;
run;

*Ausreißerbereinigung des Sturmes Emma (n_pol > 15 000 oder n_claims > 150);
data Emma_1_a;
    set Emma_1;
    if id not in (1116 1156 4563 4610);
run;

*Auswertung der Maxima bzw. Summen der Regressoren pro Quadrat;
proc sql;
create table Emma_1_b as select
    id, zip, zip2, zip3,
    max(seehoehe) as max_seehoehe,
    max(Max_Wind_i) as Max_Wind_i,
    sum(n_pol) as n_pol,
    sum(n_claims) as n_claims,
    "m1" as model
from Emma_1_a
group by id;
quit;
```

```

*Schätzung der Regressionskoeffizienten (KQ-Methode) von Emma;
proc reg data=Emma_1_b outest=Emma_1_b_ rsquare adjrsq aic bic cp mse;
    m1: model n_Claims= &var_1. &var_2. &var_3. \noprint;
proc print data=Emma_1_b_;
run; quit;

*Prognosewerte von Paula mit Regressionskoeffizienten verknüpfen;
proc sql;
create table Paula_1_m1 as select
    id, zip, zip2, zip3,
    a.&var_1., a.&var_2., a.&var_3.,
    a.n_claims,
    b.intercept,
    b.&var_1. as beta_1, b.&var_2. as beta_2, b.&var_3. as beta_3
from Paula_1_b as a
left join Emma_1_b_ as b
    on a.model=b._model_;
quit;

*Berechnung der Schätzwerte für die Schadenanzahl;
data Regression;
    set Paula_1_m1;
    reg = intercept + beta_1*&var_1. + beta_2*&var_2. + beta_3*&var_3.;
run;

*Wahre Schadenanzahl gegenüber Schätzwerten abbilden;
proc sgplot data=Regression;
    title "Forecast-Klasse 1";
    scatter x=&var_1. y=n_claims / markerattrs = (color=blue);
    scatter x=&var_1. y=reg / markerattrs = (color=green);
run; quit;

```



## 8.2 SAS-Code für den Punkt-in-Polygon-Test

```

*Definition des Basispunktes B;
%let b1=0;
%let b2=0;

*Funktion für die x-Koordinate des Schnittpunktes S definieren;
%macro xschnitt(x1,y1,x2,y2,x3,y3,x4,y4);
    ((&x1.*&y2.-&y1.*&x2.)*(&x3.-&x4.)-(&x1.-&x2.)*(&x3.*&y4.-&y3.*&x4.))/
    ((&x1.-&x2.)*(&y3.-&y4.)-(&y1.-&y2.)*(&x3.-&x4.))
%mend;

*Funktion für die y-Koordinate des Schnittpunktes S definieren;
%macro yschnitt(x1,y1,x2,y2,x3,y3,x4,y4);
    ((&x1.*&y2.-&y1.*&x2.)*(&y3.-&y4.)-(&y1.-&y2.)*(&x3.*&y4.-&y3.*&x4.))/
    ((&x1.-&x2.)*(&y3.-&y4.)-(&y1.-&y2.)*(&x3.-&x4.))
%mend;

*Streckenbedingungen aus Kapitel 4.3.2;
%macro bed(xa,xb,x);
    min(&xa.,&xb.)<=&x.<=max(&xa.,&xb.)
%mend;

*Schnittpunkte berechnen und pro PLZ-Gebiet aufsummieren;
proc sql;
create table jordan as select distinct
    a.long_sl, a.lat_sl, count(*) as n_schnitte, b.id as zip
from grid as a,
    zipat_borders as b
where a.id = b.id
and %bed(&b1.,long_sl,
    %xschnitt(&b1.,&b2.,long_sl,lat_sl,b.long1,b.lat1,b.long,b.lat))
and %bed(b.long1,b.long,
    %xschnitt(&b1.,&b2.,long_sl,lat_sl,b.long1,b.lat1,b.long,b.lat))
and %bed(&b2.,lat_sl,
    %yschnitt(&b1.,&b2.,long_sl,lat_sl,b.long1,b.lat1,b.long,b.lat))
and %bed(b.lat1,b.lat,
    %yschnitt(&b1.,&b2.,long_sl,lat_sl,b.long1,b.lat1,b.long,b.lat))
group by a.lon, a.lat, b.id
order by a.lon, a.lat, n_schnitte;
quit;

*Punkte mit einer ungeraden Anzahl von Schnittpunkten auswählen;
data jordan_1;
    set jordan (where=(mod(n_schnitte,2)=1));
    by long_sl lat_sl;
    if first.lat_sl;
run;

```

# Abkürzungsverzeichnis

Anz.	Anzahl
AIC	Akaike Information Criterion
ECMWF	Europäisches Zentrum für mittelfristige Wettervorhersagen
GNSS	Global Navigation Satellite System
GM	Geozentrische Gravitationskonstante
GPS	Global Positioning System
km/h	Kilometer pro Stunde
KQ-Methode	Methode der kleinsten Quadrate
m/s	Meter pro Sekunde
Mio.	Million
MGI	Militärgeografisches Institut
ML-Methode	Maximum-Likelihood-Methode
NAVSTAR	Navigation System with Time and Range
PPS	Precise Positioning Service
SAS	Statistical Analysis System
SPS	Standard Positioning Service
VVO	Versicherungsverband Österreich
WGS84	World Geodetic System 1984
ZAMG	Zentralanstalt für Meteorologie und Geodynamik

# Symbolverzeichnis

$^{\circ}$	Grad	$l$	Polynomgrad
$'$	Minute	$M(\alpha)$	Mediankrümmung
$''$	Sekunde	$r$	Erdradius
$\alpha$	unbestimmter Winkel	$P$	GPS-Empfänger
$\beta$	Regressionskoeffizient	$PE_i$	Pseudo-Entfernungen
$\hat{\beta}$	geschätzte Koeffizienten	$R^2$	Bestimmungsmaß
$\varphi$	Breitengrad	$\bar{R}^2$	korrigiertes Bestimmungsmaß
$\lambda$	Längengrad, Parameter der Poisson-Verteilung	$R_i$	Entfernung von GPS-Empfänger zu Satellit
$\Delta_{lat}$	Abstand zwischen Breitengraden	$S_i$	Satelliten in der Erdumlaufbahn
$\Delta_{long}$	Abstand zwischen Längengraden	$x_i$	Regressoren
$\Delta t_i$	Abstand zwischen Satellitenzeit und GPS-Zeit	$y$	Zielvariable
$\epsilon$	Lineare Exzentrizität	$\hat{y}$	Regressionsgerade
$\varepsilon$	Störterm		
$\hat{\epsilon}$	Residuum		
$\sigma$	unbestimmter Winkel		
$\omega$	Rotationsgeschwindigkeit		
$a$	große Halbachse		
$b$	kleine Halbachse		
$c$	Lichtgeschwindigkeit		
$dt$	Uhren-Offset		
$e$	Numerische Exzentrizität		
$\text{€}$	Euro		
$h$	Höhe		
$f$	Abplattung, systematische Komponente		

# Abbildungsverzeichnis

2.1	Luftbewegung zwischen Hochdruck- und Tiefdruckgebiet . . . . .	5
3.1	Längen- und Breitengrade . . . . .	9
3.2	Gestalt der Erde . . . . .	11
3.3	Lineare Exzentrizität einer Ellipse . . . . .	12
3.4	Winkel $\Delta\sigma$ . . . . .	14
3.5	Koordinatensysteme . . . . .	16
3.6	Die grundlegende Funktion von GPS . . . . .	18
4.3	Beispieldaten . . . . .	24
4.4	Punkt-in-Polygon-Test . . . . .	25
4.5	Annäherung der PLZ-Gebiete . . . . .	26
4.6	Paula am 27.01.2008 . . . . .	27
4.7	Emma am 02.03.2008 . . . . .	27
5.1	Histogramme der Variablen <code>n_claims</code> und <code>n_pol</code> . . . . .	30
5.2	Histogramme und Kerndichteschätzer . . . . .	31
5.3	<code>n_claims</code> gegen <code>n_pol</code> . . . . .	31
5.4	Streudiagramme <code>mean_wind_i</code> und <code>max_windgust_i</code> . . . . .	32
5.5	Multiple lineare Regression . . . . .	37
5.6	Polynomregression für $l = 1, 3, 5$ . . . . .	38
5.7	Vergleich der polynomialen Regression für Forecast-Klasse 2 . . . . .	39
6.1	Delta von Poisson-Regression . . . . .	47

# Tabellenverzeichnis

3.1	Abstände zwischen Breiten- bzw. Längengraden . . . . .	13
4.1	Variablen . . . . .	22
4.2	Anzahl der Prognosen . . . . .	24
5.1	Deskriptive Kennzahlen des Sturmes Emma . . . . .	30
5.2	Einfache lineare Regression . . . . .	35
5.3	Vergleich der polynomialen Regression für Forecast-Klasse 2 . . . . .	38
5.4	Generalisiertes lineares Modell mit Offset-Variable . . . . .	41
6.1	Vergleich einfache lineare Regression . . . . .	45
6.2	Vergleich multiple lineare Regression . . . . .	46
6.3	Vergleich multiple Polynom-Regression . . . . .	46
6.4	Vergleich Poisson-Regression mit Offset-Variable <code>n_pol</code> . . . . .	47
6.5	Vergleich Poisson-Regression mit Polynom und Offset-Variable <code>n_pol</code> .	48

# Literaturverzeichnis

- [1] Borth, Jan-Hendrik; *Positionsbestimmung per GPS (Global Positioning System)*; Seminararbeit; Universität Koblenz; Juni 2010
- [2] Dufner, Julius; Jensen, Uwe; Schumacher, Erich; *Statistik mit SAS*; Teubner Verlag; 3. überarbeitete Auflage; Juni 2004
- [3] Duller, Christine; *Einführung in die nichtparametrische Statistik mit SAS und R – Ein anwendungsorientiertes Lehr- und Arbeitsbuch*; Physica-Verlag; Juli 2008
- [4] Fahrmeir, Ludwig; Kneib, Thomas; Lang, Stefan; *Regression – Modelle, Methoden und Anwendungen*; Springer-Verlag; Februar 2007
- [5] Friedl, Herwig; *Generalisierte Lineare Modelle*; Institut für Statistik der Technischen Universität Graz; Oktober 2014
- [6] Klose, Brigitte; *Meteorologie – Eine interdisziplinäre Einführung in die Physik der Atmosphäre*; Januar 2008
- [7] Krämer, Walter; Schoffer, Olaf; Tschiersch, Lars; *Datenanalyse mit SAS<sup>®</sup> – Statistische Verfahren und ihre grafischen Aspekte*; Springer-Verlag; 2. überarbeitete und erweiterte Auflage; April 2008
- [8] Lamaszewska, Iwona; *Klimawandel in Österreich*; Hintergrundpapier von Global 2000; April 2012
- [9] McCullagh, Peter; Nelder, John A.: *Generalized Linear Models*; 2. Auflage; Chapman & Hall; April 1990
- [10] National Imagery and Mapping Agency: *World Geodetic System 1984 – Its Definition and Relationships with Local Geodetic Systems*; 3. Auflage; Januar 2000

- [11] Okerson, Barbara; *Distance Mapping in Health and Health Care: SAS<sup>®</sup> as a Tool for Health Geomatics*; Paper 299-2011; November 2014
- [12] Peißl, Markus; *Mathematische Grundlagen für GPS*; Karl-Franzens-Universität Graz; März 2016
- [13] Rascher-Friesenhausen, Richard; *Die Mathematik des GPS*; MNU Tagung Bremerhaven; November 2005
- [14] Rodríguez, Germán; *Lecture Notes on Generalized Linear Models*; Princeton; 2007; <http://data.princeton.edu/wws509/notes/>
- [15] SAS Institute Inc.: *SAS/STAT<sup>®</sup> 9.2 User's Guide, Seconde Edition*; North Carolina; September 2009
- [16] Schamp, Heinz: *Wind und Wetter – Wetter- und klimakundliche Anmerkungen zu dem Phänomen Wind*; Geowissenschaften in unserer Zeit; Weinheim 1984
- [17] Torge, Wolfgang: *Geodesy – Third completely revised and extended edition*; 3. Auflage; New York 2001
- [18] Zogg, Jean-Marie; *GPS und GNSS: Grundlagen der Ortung und Navigation mit Satelliten*; 9.Version; Mai 2011  
[www.diepresse.com](http://www.diepresse.com) am 15.02.2017  
[www.gps.gov](http://www.gps.gov) am 4.10.2016  
[www.gpskoordinaten.de](http://www.gpskoordinaten.de) am 4.10.2016  
[www.hora.gv.at](http://www.hora.gv.at) am 29.12.2016  
[www.igwindkraft.at](http://www.igwindkraft.at) am 28.11.2016  
[www.met.fu-berlin.de](http://www.met.fu-berlin.de) am 30.12.2016  
[www.mo-energy.at](http://www.mo-energy.at) am 30.12.2016  
[www.oeev1.orf.at](http://www.oeev1.orf.at) am 15.02.2017  
[www.versichert.at/sturmschadenversicherung](http://www.versichert.at/sturmschadenversicherung) am 03.01.2017  
[www.vvo.at](http://www.vvo.at) am 28.12.2016  
[www.webgeo.de](http://www.webgeo.de) am 28.12.2016  
[www.web.archive.org](http://www.web.archive.org) am 15.02.2017  
[www.windkraftanlagen-windenergie.de](http://www.windkraftanlagen-windenergie.de) am 28.12.2016  
[www.zamg.at](http://www.zamg.at) am 03.01.2017

# Eidesstattliche Erklärung zur Diplomarbeit

Ich versichere, die Diplomarbeit selbständig und lediglich unter Benutzung der angegebenen Quellen und Hilfsmittel verfasst zu haben.

Ich erkläre weiterhin, dass die vorliegende Arbeit noch nicht im Rahmen eines anderen Prüfungsverfahrens eingereicht wurde.

Wien, am 5. Mai 2017

---