

PROGNOSE DER LOKALEN
WALDBRANDGEFAHR IM TESSIN
DURCH MODELLIERUNG MIT HILFE
DER k -NEAREST NEIGHBORS (k NN)
METHODE

INAUGURALDISSERTATION

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel
von

Andreas Felber
aus Basel, Schweiz

Davos, 2005

Genehmigt von der Philosophisch-Naturwissenschaftlichen
Fakultät der Universität Basel auf Antrag der Herren

Prof. Dr. E. Parlow

Prof. Dr. D. Scherer

Dr. P. Bartelt, SLF Davos

Basel, den 11. April 2005

Prof. Dr. H.-J. Wirz, Dekan

*There are three simple rules for creating a model.
Unfortunately, nobody knows what they are.*

W. Somerset Maugham

Inhaltsverzeichnis

Abbildungsverzeichnis	ix
Tabellenverzeichnis	xiii
Symbole und Abkürzungen	xv
Zusammenfassung	xvii
Abstract	xix
1 Einleitung	1
1.1 Motivation	1
1.2 Zielsetzung	3
1.3 Gliederung der Arbeit	4
2 Grundlagen	7
2.1 Das Phänomen Waldbrand und seine ökologischen Folgen	7
2.2 Geschichte der Waldbrände auf der Alpensüdseite	10
2.3 Klima im Tessin	14
2.4 Vegetation im Tessin	18
3 Prognosemodelle	19
3.1 Einführung	19
3.2 Modelltypen	19
3.3 Schritte der Modellentwicklung	21
3.4 Prognosemodelle in der Waldbrandforschung	22
3.5 Klassifikation von Ereignissen	24

4	Die k-Nearest Neighbors (kNN) Methode	27
4.1	Einführung	27
4.2	Grundlagen	31
4.3	Die k NN-Methode	31
4.4	Entwicklung des k NN-Modells	34
5	Das kNN-Waldbrand Prognosemodell	39
5.1	Stationsauswahl und Gebietseinteilung	39
5.2	Grundwahrscheinlichkeit s (<i>base rate</i>)	42
5.3	Homogenität der Messreihen	45
5.4	Variablenwahl	49
5.4.1	Empirische Bestimmung der Variablen	50
5.4.2	Bestimmung der Variablen mittels logistischer Regression	52
5.4.3	Skalierung der Variablen	56
5.5	Variablengewichtung mittels Genetischen Algorithmen (GA)	56
5.5.1	Methode	57
5.6	Erstellung einer k NN-Prognose	60
5.6.1	Binäre Prognose (BF)	60
5.6.2	Probabilistische Prognose (PF)	61
5.6.3	Deskriptive Prognose (DF)	63
6	Verifikation von Prognosemodellen	67
6.1	Einführung	67
6.2	Trainingsdatensatz und Testdatensatz	68
6.3	Arten der Verifikation	70
6.4	Verifikation von binären Prognosen	72
6.4.1	Relative Operating Characteristic (ROC)	77
6.5	Verifikation von probabilistischen Prognosen	79
6.5.1	Zuverlässigkeitsdiagramm (<i>attributes diagram</i>)	81
6.6	Verifikation von deskriptiven Prognosen	81
7	Resultate	83
7.1	Einführung	83
7.2	Binäre Prognose (BF)	83
7.2.1	Relative Operating Characteristic (ROC)	87
7.3	Probabilistische Prognose (PF)	87

7.3.1	Zuverlässigkeitsdiagramm (attributes diagram)	88
7.4	Deskriptive Prognose (DF)	90
7.5	Fallbeispiele	91
7.6	Sensitivitätsanalyse	99
8	Diskussion	107
8.1	Kritische Betrachtung der Ergebnisse	107
8.1.1	Fazit - Verifikation	107
8.1.2	Fazit - binäre Prognose	108
8.1.3	Fazit - probabilistische Prognose	109
8.1.4	Fazit - deskriptive Prognose	110
8.1.5	Fazit - Grenzen und Möglichkeiten der k NN-Methode . . .	111
8.2	Vergleich mit anderen Modellen	113
8.3	Bedeutung der Ergebnisse in der Waldbrandforschung	114
8.4	Künftige Modellentwicklung	115
9	Schlussfolgerungen und Perspektiven	117
10	Danksagung	121
A	Logistisches Regressionsmodell	123
B	Gebietszuordnung der Gemeinden	127
	Literaturverzeichnis	135
	Curriculum Vitae	143

Abbildungsverzeichnis

1.1	Löscharbeiten am teilweise zerstörten Schutzwald von Leuk nach dem Waldbrand vom August 2003.	2
2.1	Zusammenhänge zwischen den verschiedenen Waldbrandtypen beziehungsweise Ausbreitungsarten.	8
2.2	Brandhäufigkeit auf der Alpensüdseite 16500 v. Chr. bis heute. . .	10
2.3	Verteilung der Anzahl Brände und der Extremereignisse. Bezugszeitraum 1920-2000.	11
2.4	Durchschnittliche monatliche Verteilung der Anzahl Brände und der betroffenen Fläche auf der Alpensüdseite (1981-1999).	13
2.5	Zusammenfassung der bedeutendsten Waldbrand-Einflussfaktoren in der Schweiz.	13
2.6	Klimadiagramme der MeteoSchweiz-Stationen Piotta, Locarno-Monti sowie Lugano.	15
2.7	Perioden von mindestens 60 Tagen mit einer Niederschlagssumme < 10mm im 20. Jahrhundert.	17
3.1	Schematische Darstellung der wichtigsten Schritte und Abläufe der Modellentwicklung.	21
4.1	Unterscheidung der drei Hauptfaktoren Grunddisposition, variable Disposition und auslösendes Ereignis.	29
4.2	Konzeptionelles Schema, welches den Zusammenhang zwischen Waldbränden und Lawinen in Bezug auf die Prognosemodellierung wiedergibt.	30
4.3	Systematische Darstellung der Diskriminierung von \mathbf{x}^+ (Ereignistagen) und \mathbf{x}^- (Nichtereignistagen) in die beiden Klassen Ω^+ und Ω^- mittels k NN-Methode.	34

4.4	Systematischer Ablauf des k NN-Modells zur Waldbrandprognose.	35
5.1	Standorte der verwendeten meteorologischen Mess- und Phänologiestationen.	40
5.2	Fiktives Beispiel um den Einfluss eines Trends zu verdeutlichen.	46
5.3	STL-Plot für die maximale Lufttemperatur der Station Lugano 1991-2002.	49
5.4	STL-Plot für die minimale relative Luftfeuchte der Station Lugano 1991-2002.	49
5.5	Phänologischer Zeitraum.	52
5.6	Schema eines genetischen Algorithmus.	58
5.7	Bestimmung des Schwellenwertes (<i>decision boundary</i>) mittels dem HSS.	61
5.8	Korrelation zwischen der berechneten Wahrscheinlichkeit p_y eines Ereignisses und der tatsächlichen Häufigkeit p des Auftretens eines Waldbrandes.	62
5.9	Möglicher k NN-Output in Form einer Karte.	64
5.10	Statistische Auswertung der DF Prognose.	64
6.1	Subjektive Einschätzung der “Nützlichkeit” vom NXD-Lawinenmodell.	71
6.2	Gegenüberstellung POD und FAR.	74
6.3	ROC-Diagramm für die Waldbrandprognose des k NN-Modells.	78
6.4	Zuverlässigkeitsdiagramm für die Waldbrandprognose des k NN-Modells.	81
7.1	Gegenüberstellung POD und FAR der Region Lugano (1991-2002).	84
7.2	Gegenüberstellung POD und FAR der Region Locarno (1991-2002).	85
7.3	Gegenüberstellung POD und FAR der Region Piotta (1991-2002).	85
7.4	Heidke Skill Score der k NN-Modelle für die Prognose der Regionen Lugano, Locarno und Piotta (1991-2002).	86
7.5	ROC-Diagramm für die Prognosen der drei Regionen Lugano, Locarno und Piotta (1991-2002).	87
7.6	Zuverlässigkeitsdiagramm der Region Lugano (1991-2002).	89
7.7	Zuverlässigkeitsdiagramm der Region Locarno (1991-2002).	89
7.8	Zuverlässigkeitsdiagramm der Region Piotta (1991-2002).	90

7.9	Einfache statistische Auswertung der deskriptiven Prognose (DF) vom 26.03.1994.	91
7.10	Berechnete Anzahl positiver Nachbarn η^+ und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1991-31.12.2002 (volles Modell).	92
7.11	Berechnete Anzahl positiver Nachbarn η^+ und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1991-31.12.2002 (reduziertes Modell).	93
7.12	Berechnete Anzahl positiver Nachbarn η^+ und tatsächlich aufgetretene Waldbrände der Region Locarno für den Zeitraum 01.01.1991-31.12.2002 (reduziertes Modell).	94
7.13	Berechnete Anzahl positiver Nachbarn η^+ und tatsächlich aufgetretene Waldbrände der Region Piotta für den Zeitraum 01.01.1991-31.12.2002 (reduziertes Modell).	94
7.14	Berechnete Anzahl positiver Nachbarn $\eta^+ \geq 3$ und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1991-31.12.2002 (reduziertes Modell).	95
7.15	Berechnete Anzahl positiver Nachbarn und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1997-31.12.1997 (reduziertes Modell).	96
7.16	Berechnete Anzahl positiver Nachbarn $\eta^+ \geq 3$ und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1997-31.12.1997 (reduziertes Modell).	97
7.17	Berechnete Anzahl positiver Nachbarn und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.04.1997-30.04.1997 (volles und reduziertes Modell).	98
7.18	Meteorologische Situation an der Station Lugano im April 1997.	98
7.19	Berechnete Anzahl positiver Nachbarn η^+ und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1991-31.12.2003.	100
7.20	Berechnete Anzahl positiver Nachbarn η^+ und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1997-31.12.1997.	101

7.21	Berechnete Anzahl positiver Nachbarn und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1997-31.12.1997 (reduziertes Modell).	104
7.22	Darstellung der Problematik der Wahl von k	105
A.1	Verifikation des logistischen Prognosemodells.	125

Tabellenverzeichnis

3.1	Zusammenfassung und Überblick einiger weitverbreiteter Waldbrandprognosemodelle.	23
5.1	Übersicht der in Frage kommenden ANETZ-Stationen (Meteo-Schweiz).	41
5.2	Gebietseinteilung der Untersuchungsgebiete und entsprechende meteorologische Stationen.	42
5.3	Gemessene Variablen der ANETZ-Stationen und nach einer ersten empirisch durchgeführten Variablenreduktion dem k NN-Modell zur Verfügung stehenden Grössen.	50
5.4	Übersicht über die im k NN-Modell zu berücksichtigenden Phänologiedaten und -stationen.	51
5.5	Zusammenfassung der Signifikanzsymbole.	53
5.6	Signifikanztest der Variablen des vollen Modells mit Hilfe der logistischen Regression für die Messstation Lugano (1991-2002). . .	54
5.7	Signifikanztest der Variablen des reduzierten Modells mit Hilfe der logistischen Regression für die Messstation Lugano (1991-2002). .	54
5.8	Signifikanztest der Variablen des reduzierten Modells mit Hilfe der logistischen Regression für die Messstation Locarno (1991-2002). .	55
5.9	Signifikanztest der Variablen des reduzierten Modells mit Hilfe der logistischen Regression für die Messstation Piotta (1991-2002). . .	55
5.10	Beispieloutput einer deskriptiven Prognose (DF).	63
6.1	Zusammenfassung der wichtigsten Attribute einer Prognose. . . .	70
6.2	Kontingenzmatrix kategorieller Prognosen binärer Ereignisse. . . .	72
6.3	Kontingenzmatrix kategorieller Prognosen binärer Ereignisse. <i>Joint</i> und <i>marginal</i> Wahrscheinlichkeiten.	72

7.1	Output der deskriptiven Prognose (DF) des k NN-Waldbrand Prognosemodells für den 26.03.1994.	91
7.2	Zusammenfassung der Anzahl aufgetretener Waldbrände und der Anzahl berechneter positiver Nachbarn η^+ in den drei Untersuchungsgebieten.	93
7.3	Zusammenfassung der wichtigsten Grössen zur Bestimmung der Sensitivität des Modells.	103

Symbole und Abkürzungen

<i>Symbol</i>	<i>Bedeutung</i>
a	prognostiziertes Ereignis
A_z	Fläche unter der ROC-Kurve
b	prognostiziertes Nichtereignis
β	Koeffizienten, Schätzer (<i>Estimator</i> bei LR)
c	Nichtprognostiziertes Ereignis
d	Nichtprognostiziertes Nichtereignis
D	Abstand im p -dimensionalen Raum
$\mathcal{E}\langle Y_i \rangle$	Erwartungswert der binären Grösse (liegt zwischen 0 und 1)
e_p	Prognoseunsicherheit
η_i	linearer Prädiktor bei LR
η^+	Anzahl gefundener Nachbarn
j	Index der Beobachtungen
k	Anzahl Nachbarn
L	Kategorien bei <i>Brier Score</i>
n	Umfang der Stichprobe
p	Dimensionen (Anzahl der Merkmale)
p_y	Eintrittswahrscheinlichkeit eines Waldbrandtages, bestimmt nach der probabilistischen Auswertung des kNN-Modells
p_{LR}	Eintrittswahrscheinlichkeit eines Waldbrandtages, berechnet mit LR
Ω	Merkmalsgrundraum
s	Grundwahrscheinlichkeit (<i>base rate</i>)
S	Stichprobe
x_i	meteorologische Variablen
\mathbf{x}	Tag (Vektor), gebildet aus x_i
\mathbf{y}	zu prognostizierender Tag (Vektor)
\hat{x}_i, \hat{y}_i	skalierte und gewichtete Grössen (vergänger Tage, an zu prognostizierenden Tagen)
Y_i	Zielgrösse bei der LR
x_i	erklärende Grössen bei LR (meteorolog. Variablen)

<i>Abkürzung</i>	<i>Text</i>
B	Bias
BF	Binäre Prognose
BS	<i>Brier Score</i>
CSI	<i>Critical success index</i>
CT	Classification trees (Entscheidungsbäume)
DA	Diskriminanzanalyse
DB	Schwellenwert (<i>decision boundary</i>)
DF	Deskriptive Prognose
F	<i>False alarm rate</i>
FAR	<i>False alarm ratio</i>
FBI	<i>Frequency bias</i>
SLF	Eidg. Institut für Schnee- und Lawinenforschung, Davos
FWI	Kanadischer-Feuer-Wetter Index
GA	genetic algorithms (Genetische Algorithmen)
HSS	<i>Heidke Skill Score</i>
kNN	k-Nearest Neighbors (k-nächste Nachbarn)
LR	logistische Regression
NNnet	Neuronale Netze
NXD	kNN-Lawinenprognosemodell, entwickelt am SLF
PEC	Treffergenauigkeit (<i>percent correct score</i>)
PF	Probabilistische Prognose
POD	<i>Probability of detection</i>
ROC	<i>Relative Operating Characteristics</i>
s	Grundwahrscheinlichkeit (<i>base-rate</i>)
SS	Skill Score
STL	<i>Seasonal Trend Decomposition Procedure</i>
TSS	<i>True Skill Statistic</i>
WSL	Forschungsanstalt für Wald-, Schnee- und Landschaft

Zusammenfassung

Waldbrände stellen weltweit ein sehr grosses Naturgefahrenpotential dar. Unzählige Waldbrände wüteten in den letzten Jahren in den USA, Brasilien, Indonesien, den Mittelmeerrändern Europas und auch in der Schweiz, hauptsächlich im Kanton Tessin.

Waldbrandprognosemodelle sind ein wichtiges Instrument zur Waldbrandprävention und daher für eine effiziente Waldbrandmanagementstrategie unabdingbar.

Im Rahmen dieser Arbeit wurde ein statistisches Modell zur Prognose der lokalen Waldbrandgefahr, basierend auf der k -Nearest Neighbors (k NN) Methode, entwickelt. Die k NN-Methode ist eine nichtparametrische Diskriminanzanalyse, deren Annahmen darauf basieren, dass in einem durch Variablen aufgespannten Vektorraum mittels eines Euklid'schen Abstandmasses ähnliche Datensätze gefunden werden. Voraussetzung für das Funktionieren der Prognosemethode ist, dass sich ein Variablensatz finden lässt, sodass sich Tage mit Waldbränden innerhalb des durch die Variablen aufgespannten Ereignisraumes gruppieren lassen. Weiter setzt die Methode voraus, dass die erklärenden Variablen keinem zeitlichen Trend unterliegen und über das gesamte Untersuchungsgebiet extrapoliert werden können.

Im Rahmen dieser Arbeit sind allgemein gültige "Regeln" für die Anwendung der k NN-Methode als Prognosemodell für die Naturgefahren Waldbrand und Lawinen formuliert worden. Basierend auf diesen "Regeln", ist eine einfache Weiterentwicklung der k NN-Methode auf andere Naturgefahren möglich.

Die Ergebnisse der k NN-Methode können auf drei unterschiedliche Arten dargestellt werden: Die *binäre* Prognose stellt einen dichotomen Ja-Nein-Entscheid dar. Die Verifikation dieses Prognosetyps basiert auf der Auswertung von Kon-

tingenzmatrizen. Bei der *probabilistischen* Prognose wird eine Wahrscheinlichkeit für das Eintreten eines Ereignisses berechnet. *Brier Score* und Zuverlässigkeitsdiagramme stellen dabei die Verifikationsmasse dar. Die *deskriptive* Prognose ist die dritte und umfassendste Ausgabeform des *k*NN-Modells. Sie beinhaltet im Gegensatz zu den anderen beiden Prognosen die gesamte, nicht selektierte Information, welche in der Datenbank über die nächsten Nachbarn enthalten ist.

Das *k*NN-Modell stellt mit seinen unterschiedlichen Ausgabemöglichkeiten ein weiteres Unterstützungswerkzeug im Entscheidungsprozess einer Gefahrenabschätzung der aktuellen Waldbrandsituation dar und kann von entsprechenden Behörden somit sinnvoll zur lokalen Waldbrandprognose eingesetzt werden.

Abstract

Forest fire are a major natural hazard in Europe. While writing this work (2001-2004) wildfires raged in Spain, Portugal and France. Since the 1960's forest fire activity has also increased in Switzerland, primary in the Canton of Ticino.

The development of fire danger prediction methods is considered to be one of the most important elements of an effective forest fire management strategy. In this work a statistical forecast model, based on the k -nearest neighbors method (k NN), was developed and applied to asses forest fire danger.

Statistically, the k NN method is a non-parametric pattern classification technique which arranges data in a multi-dimensional space and applies a distance metric (usually Euclidean) to define the distance between past and present data. The method assumes that similar events are likely to exhibit similar precursors and that fire and weather factors can be extrapolated over the geographic forecast area.

General rules underlying the application of a k -nearest neighbors model were formulated in this work. Based on these fundamental principles a further development of the k NN model in the field of natural hazard prediction, both spatially and thematically, is enabled.

The output of the k NN forecast can be used in different ways. A *binary* forecast is used with decision boundaries to predict event occurrence. Contingency tables for this type of forecast are evaluated. A *probabilistic* forecast can be used to predict the probability of occurrence and therefore danger levels. *Brier score* and *attributes diagrams* are presented that are used to verify this forecasting method. In a *descriptive* forecast, which records a detailed list of events and all associated individual observations, the responsibility for the danger assessment is with the user.

The absolute quality of the k NN approach, however, is difficult to quantify because the method provides examples, which often trigger decision making proces-

ses. Thus, the k NN model can be a valuable addition to general decision making tools.

1 Einleitung

1.1 Motivation

Hochwasser, Stürme, Steinschlag, Muren, Erdbeben, Waldbrände und Lawinen bedrohen die Menschheit seit ihren Anfängen und haben einen nachhaltigen Einfluss auf die Entwicklung der verschiedenen Kulturen der Erde gehabt. Treten diese Ereignisse in von Menschen stark genutzten Räumen auf, sind Personen- und Sachschäden oft direkte Folgen. Durch die frühzeitige Erkennung von Naturgefahren und die Kommunikation an die Betroffenen können die Ausmasse der Schäden bei entsprechenden Massnahmen reduziert oder ganz eliminiert werden (Brabec, 2001).

Global brennen jährlich nach Schätzungen der Ernährungs- und Landwirtschaftsorganisation der Vereinten Nationen (FAO) 11.3 Mio. ha Wald (Gesamtwaldfläche: 3.4 Mia. ha) (FAO, 1999). In jüngster Zeit haben Naturkatastrophen wie die Waldbrände vom Sommer 2003 (Kantone Wallis und Tessin), die Murgänge von 2002 (Kanton Graubünden), die Lawinen vom Winter 1999 (gesamter Alpenraum) und die Winterstürme 1999 ("Lothar") die schweizerische Öffentlichkeit stark bewegt.

Der zunehmende Siedlungsdruck und die Ausbreitung touristischer Aktivitäten in potentiellen Gefahrenräumen erfordern ein weit entwickeltes Risikomanagementkonzept. Für die Minimierung von Schäden und Evakuierung gefährdeter Räume ist die Prognose potentieller Naturgefahren ein wichtiger und mitentscheidender Bestandteil.

Die Schweiz ist hauptsächlich von Hochwasser, Lawinen und Rutschungen, in geringerem Ausmass aber auch von Waldbränden betroffen. Obwohl die Naturgefahr Waldbrand nicht sehr bekannt ist, tritt sie - lokal vorkommend (Alpensüdseite, dabei insbesondere im Kanton Tessin) - relativ häufig auf. Im Durchschnitt brennen jährlich bei 80-100 Bränden rund 740ha Wald, dies hauptsächlich in den

1. EINLEITUNG

trockenen Winter- und den heissen Sommermonaten. Die Brände können schwerwiegende ökologische und ökonomische Auswirkungen zur Folge haben, insbesondere dann, wenn der Schutz, den die Wälder vor Lawinen oder Steinschlag bieten, zerstört werden (Abb. 1.1). Die Installation von einer Hektar schützender Struktur kostet rund 1 Mio. SFr. (Allgöwer, 2001). Neben dem ökonomischen Aspekt stellen die Gebirgswälder auch aus ökologischer Sicht ein hoch sensibles Ökosystem dar, welches nach einer Zerstörung Jahrzehnte braucht, um wieder seinen ursprünglichen Zustand zu erreichen.



Abb. 1.1: Löscharbeiten am teilweise zerstörten Schutzwald von Leuk nach dem Waldbrand vom August 2003. Photo: A. Felber.

Waldbrände sind in ihrer Art und Intensität, in ihrem Verhalten und ihren Auswirkungen stark variierend, sodass eine Vielzahl von Faktoren benötigt werden, um ein Feuer zu beschreiben.

Auch wenn die Schweiz kein typisches Waldbrandland wie etwa die USA oder die Mittelmeeranrainer ist, zeigt die Vergangenheit, dass ausgedehnte Katastrophenwaldbrände vorkommen können. Die globale Klimaveränderung und die damit einhergehende mögliche Veränderung des Niederschlagregimes auf der Alpensüdseite fordern geeignete planerische Mittel und Strategien, um einem möglichen künftigen Anstieg der Waldbrandgefahr vorzubeugen.

1.2 Zielsetzung

Ausgereifte Modelle für das Verhalten von Waldbränden in Abhängigkeit von Vegetationstyp, Menge und Anordnung des zur Verfügung stehenden Brennmaterials, Topographie und meteorologischen Rahmenbedingungen liegen für eine Reihe von aussereuropäischen Ländern vor. Auch die südeuropäischen Länder haben sich in der Vergangenheit mit diversen Waldbrandprognosemodellen auseinandergesetzt, wobei diese mehrheitlich auf der Brandgutfeuchte basieren. In der Schweiz wird dieser Ansatz hauptsächlich für den Kanton Graubünden durch Arbeiten des Geographischen Institutes der Universität Zürich verfolgt (Allgöwer, 2001).

Neben solchen deterministisch-physikalischen Modellansätzen soll die vorliegende Arbeit einen statistischen Ansatz zur Modellierung der Waldbrandgefahr aufzeigen. Für die Lawinenprognose hat sich gezeigt, dass mit Hilfe eines k -Nearest Neighbors (k NN) Modells für ein begrenztes Gebiet gute Ergebnisse erzielt werden können (Heierli et al., 2003). Basierend auf demselben k NN-Ansatz soll in der vorliegenden Arbeit ein Prognosemodell für die Naturgefahr Waldbrand entwickelt werden.

Damit ein erfolgreiches Modell entwickelt werden kann, müssen zuerst die zu erfüllenden Anforderungen bekannt sein.

Zur Entwicklung eines operationellen Waldbrandprognosemodells für die Regionen auf der Alpensüdseite gelten folgende Ziele, die unter bestimmten Randbedingungen zu erfüllen sind:

1. Das Modell muss den Ansprüchen der für die Warnung von Waldbränden zuständigen Behörden genügen und von diesen für ihre zu erstellenden Prognosen angewendet werden können.
2. Die räumliche Auflösung des Modells soll im Skalenbereich *lokal* bis *regional* liegen.
3. Die zeitliche Auflösung des Modells soll grundsätzlich bei einem Tag liegen. Es sollen aber sowohl Prognosen für den aktuellen Tag, als auch 3-Tagesprognosen berechnet werden können.
4. Der Output des Modells soll eine Analyse nach unterschiedlichen Einflusskriterien (z.B. Höhenlage, Exposition, Waldtyp, Ursache, etc.) ermöglichen.

1. EINLEITUNG

5. Da das Modell sowohl auf Messdaten als auch auf meteorologische Prognosedaten der MeteoSchweiz basiert, muss eine Schnittstelle für einen “einfachen” Datenimport implementiert werden.
6. Eine Modellverifikation muss einfach möglich sein. Entsprechende Analyse-Tools müssen entwickelt und in das Modell integriert werden.
7. Die Struktur des Modells soll modular so aufgebaut werden, dass eine Erweiterung (räumlich oder thematisch) mit angemessenem Aufwand realisierbar ist.

Können die geforderten Bedingungen erfüllt werden, kann das Modell für neue Fragestellungen anderer Themenbereiche oder Regionen adaptiert werden.

Neben der Datenaufbereitung und der Prüfung ihrer Qualität soll in der vorliegenden Arbeit die Problematik der Diskriminierung binärer Ereignisse angesprochen werden. Die k NN-Methode stellt eine Möglichkeit zur Lösung dieses Problems dar.

Die Resultate sollen hauptsächlich in Bezug auf eine praxisorientierte Anwendung des Modells überprüft werden. Zwar gibt es in der Waldbrandforschung zahlreiche theoretische Modellansätze, jedoch fehlen praxisbezogene, in der Realität sinnvoll einsetzbare Modelle.

1.3 Gliederung der Arbeit

Gegliedert ist die vorliegende Arbeit wie folgt: Im **Kapitel 2 Grundlagen** werden jene Punkte behandelt, welche die Grundlage dieser Arbeit bilden. Einen Überblick über bereits existierende Waldbrandprognosemodelle und eine Einführung in die Problematik der Diskriminierung von erklärenden Grössen zur Beschreibung binärer Ereignisse liefert das **Kapitel 3 Prognosemodelle**. **Kapitel 4, die k -Nearest Neighbors Methode**, liefert einen Überblick über die k NN-Methode und eine Modellbeschreibung. Dabei werden ebenfalls Grundlagen erarbeitet, welche für die Modellentwicklung gebraucht werden. Im **Kapitel 5, das k NN-Waldbrand Prognosemodell**, wird schrittweise die Entwicklung des k NN-Waldbrand Prognosemodells erklärt. Eine Einführung in die Verifikation von Prognosemodellen und die Anwendung bezüglich der k NN-Resultate folgt im **Kapitel 6, Verifikation von Prognosemodellen**. Im **Kapitel 7, Resultate**

werden die Ergebnisse der Arbeit aufgeführt und in **Kapitel 8, Diskussion** anschliessend diskutiert. Abschliessend folgt in **Kapitel 9, Schlussfolgerungen** eine Bewertung der Diskussion. Im Anhang werden Grundlagen des logistischen Regressionsmodells besprochen und es folgt eine Übersicht über die Gebietszuordnung der Gemeinden.

2 Grundlagen

2.1 Das Phänomen Waldbrand und seine ökologischen Folgen

Unter dem Begriff “Waldbrand” werden verschiedene Arten und Ausprägungen eines Feuers zusammengefasst. Flur- oder Wiesenbrände fallen ebenfalls in die Kategorie der Waldbrände, welchen grundsätzlich besser der englische Begriff *wildfire* zugeordnet werden sollte. Da es für diesen Begriff keine sinngemäß gute Übersetzung gibt, wird im Folgenden trotzdem vom allgemeinen Überbegriff “Waldbrand” gesprochen.

Je nach Entstehung und Ausbreitung können verschiedene Waldbrandtypen unterschieden werden. Waldbrände werden in Abhängigkeit von Entstehungsursache und Ausbreitung kategorisiert (Abb. 2.1).

Die häufigste Waldbrandart ist das Lauffeuer. Es breitet sich in der bodennahen Schicht durch Wind rasch aus. Die Streu-, Kraut- und Strauchschichten sind davon am stärksten betroffen.

Das Kronenfeuer bildet eine weitere Gruppe. Ausgehend von Bodenfeuern kann sich ein Brand, sofern zwischen dem Boden und der Baumkrone brennbares Material vorhanden ist, in die Baumkronen fortpflanzen. Bei starken Winden kann sich ein Kronenfeuer aufgrund der durch Hitze entstehenden konvektiven Strömungen rasch über grosse Distanzen ausbreiten (Flugfeuer).

Als Erdfeuer bezeichnet man unterirdische Glutherde. Aufgrund des nur spärlich vorhandenen Sauerstoffs breitet sich das Feuer langsam aus. Da Erdfeuer an der Oberfläche kaum sichtbar sind, gestaltet sich ihre Bekämpfung sehr schwierig.

Als letzte Gruppe sind die Stockfeuer zu erwähnen. Sie glimmen in toten und hohlen Baumstämmen. Über diese gelangen die Feuer in den Boden bis zu den Wurzeln, wo sie neue Erdfeuer entfachen können.

2. GRUNDLAGEN

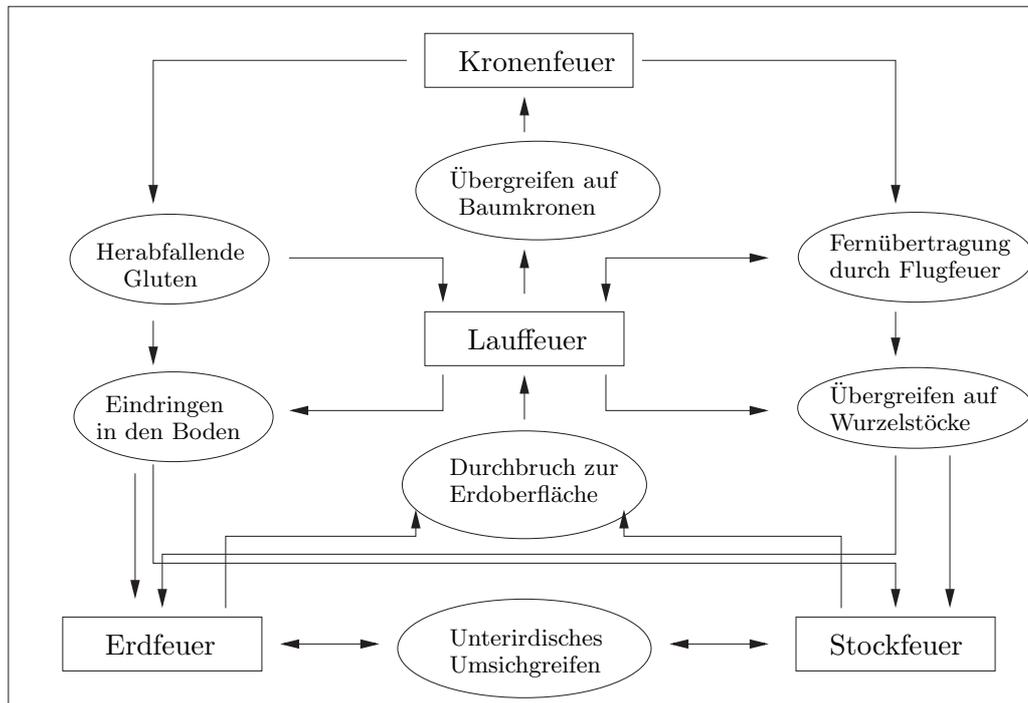


Abb. 2.1: Zusammenhänge zwischen den verschiedenen Waldbrandtypen beziehungsweise Ausbreitungsarten. Nach Jeklin & Schöb (1993). (Leicht verändert).

Wird ein Brand durch eine natürliche Ursache (z.B. Blitz) ausgelöst, ist ein Waldbrand ein natürlicher Vorgang. Sowohl in diesem Fall als auch bei einem durch Menschen verursachten Feuer müssen die Brandfolgen nicht unbedingt negativ sein. In von Bränden regelmässig betroffenen Gebieten hat sich eine entsprechend feuertolerante Fauna und Flora entwickelt. Regelmässige Brände erneuern auch den Waldbestand. Brandflächen weisen in den Jahren nach einem Brand eine im Vergleich zum benachbarten Wald höhere Artenvielfalt auf (Conedera et al., 1996).

Aussagen bezüglich den ökologischen Folgen eines Waldbrandes beruhen zu einem grossen Teil auf einem Experiment, welches im März 1998 in Copra (TI) durchgeführt wurde. Im Rahmen dieses Versuches wurde ein rund 0.5ha grosses Waldstück kontrolliert in Brand gesteckt. Neben der Sottostazione Sud delle Alpi (SdA) beteiligten sich rund 15 weitere Forschergruppen (Conedera et al., 1998).

Während eines Waldbrandes entsteht in der Regel starke oberirdische Thermik. Diese wirkt sich schädlich auf pflanzliches Gewebe aus. Nur Baumarten wie Ei-

chen, alte Birken und Kastanien können aufgrund der dicken Borke teilweise überleben. Unterirdische Pflanzenteile hingegen werden vom Feuer kaum beeinträchtigt. Temperaturmessungen ergaben innerhalb der Streuschicht Werte von 400-650°C, in 2.5cm Tiefe nur noch Werte um 35°C (Marxer, 2003). Wieder austreibende Arten wie die Kastanie sind somit gegenüber jenen Arten, welche sich nicht vegetativ vermehren im Vorteil und können Brandflächen rasch besiedeln. Entgegen den Erwartungen hat sich die Artenzusammensetzung der wirbellosen Fauna nach einem Lauffeuer kaum verändert (Moretti et al., 1998). Je nach Lebens- und Verhaltensweisen der einzelnen Tierarten reduziert zwar das Feuer deren Individuenzahl stark, löscht in der Regel jedoch keine aus.

Waldbrände haben oft aber auch negative Folgen. Durch die grosse Hitzeeinwirkung kann das Pflanzen- und Tierleben zumindest teilweise zerstört werden. So können u.a. auch Mikroorganismen im Boden geschädigt werden, was eine Humifizierung des abgestorbenen Materials verunmöglicht. Bodenstrukturveränderungen und erhöhte Erosionsgefahr sind die Folgen. Der Waldboden wird nicht mehr von Vegetation bedeckt und damit anfälliger für Witterungseinflüsse. Die Folgen sind Bodenabschwemmung, erhöhtes Risiko von Winderosion und erhöhte Evaporation (Winkler, 1971).

Viele Wälder erfüllen eine Schutzfunktion gegen Naturgefahren (Lawinen, Murgänge, etc.) oder dienen als Erholungsraum für die Bevölkerung. Mit dem Verlust des Waldes geht auch diese Funktion verloren. Die von verheerenden Wald- und Buschbränden betroffenen Regionen sind oft auch bewohnt, wie Beispiele aus verschiedenen Ländern immer wieder zeigen.

Ganz allgemein lässt sich sagen, dass die Auswirkungen von Waldbränden auf das Ökosystem davon abhängen, zu welcher Jahreszeit, wie intensiv, wie lange und wie oft es brennt. Die Kombination dieser Einflussfaktoren bezeichnet man als Feuerregime (Winkler, 1971). In einer feueranfälligen Region wie der Alpensüdseite, muss man das Feuerregime als Standortfaktor bezeichnen, der die Entwicklung der Waldökosysteme massgeblich beeinflusst.

2.2 Geschichte der Waldbrände auf der Alpensüdseite

Pollenanalytische Untersuchungen zeigen, dass die Alpensüdseite der Schweiz von Natur aus eine waldbrandgefährdete Region ist. Seit dem Neolithikum vor 6500 Jahren lösten menschliche Aktivitäten regelmäßig Brände aus (vgl. Abb. 2.2). Während der Bronze- und Eisenzeit kam es am häufigsten zu Waldbränden, weil der Mensch das Feuer für die Brandrodung einsetzte. Mit der Aufforstung der Kastanie während der Römerzeit und der späteren intensiven Landnutzung im Mittelalter nahm die Häufigkeit von Waldbränden deutlich ab (Tinner et al., 1999).

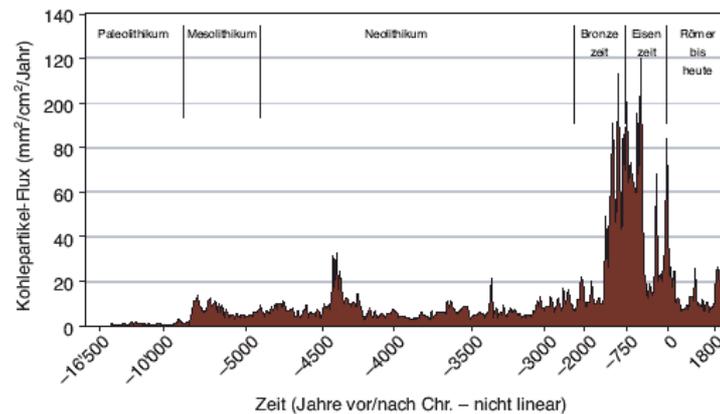


Abb. 2.2: Der Mensch hat die Häufigkeit von Waldbränden auch früher beeinflusst. Ablagerungen von Kohlepartikeln in der Nacheiszeit, nachgewiesen in den Sedimenten des Origliesees, sind proportional zur Brandhäufigkeit (Tinner et al., 1998). Die Periode von 5000-7200 v. Chr. zeigt das Niveau natürlicher Waldbrände unter ähnlichen klimatologischen Bedingungen wie heute. Die extreme Häufigkeit von Waldbränden in der Bronze- und Eisenzeit ist auf Brandrodungen durch den Menschen zurückzuführen. Die Brandrodungen fanden ein Ende, als die Römer um das Jahr 0 begannen die Edelkastanie aufzuforsten (Tinner et al., 1999).

Im 20. Jahrhundert nahm die jährliche Anzahl der Brände auf der Alpensüdseite seit den 60er Jahren von durchschnittlich 30 auf 80 zu (Abb. 2.3 (a)). Dies weil die land- und forstwirtschaftliche Nutzungsintensität abnahm. Ehemals offene Flächen verwaldeten und brennbares Material häufte sich an. Der generelle Anstieg der Feuergefahr führte an feuergefährdeten Standorten zu hohen Feuerfrequenzen, welche die Waldentwicklung stark beeinflussten (Conedera et al.,

1997). Seit den 80er Jahren sinkt die jährliche Anzahl der Waldbrände wieder und auch die Grössen der betroffenen Brandflächen nehmen seit den 60er Jahren ab. Eine Ausnahme bildete das ausserordentliche Jahr 1973. Die Extremereignisse, bei denen sich die Waldbrände auf Flächen von über 100ha ausweiteten, waren zwischen 1941 und 1980 am häufigsten (Abb. 2.3 (b)).

Zwischen 1981 und 2000 sank die jährliche Anzahl der Brände auf das Niveau von 1921-1940 zurück (Conedera, 2003). Leider fehlt in der Statistik der trockene und von einigen Waldbränden betroffene Sommer 2003.

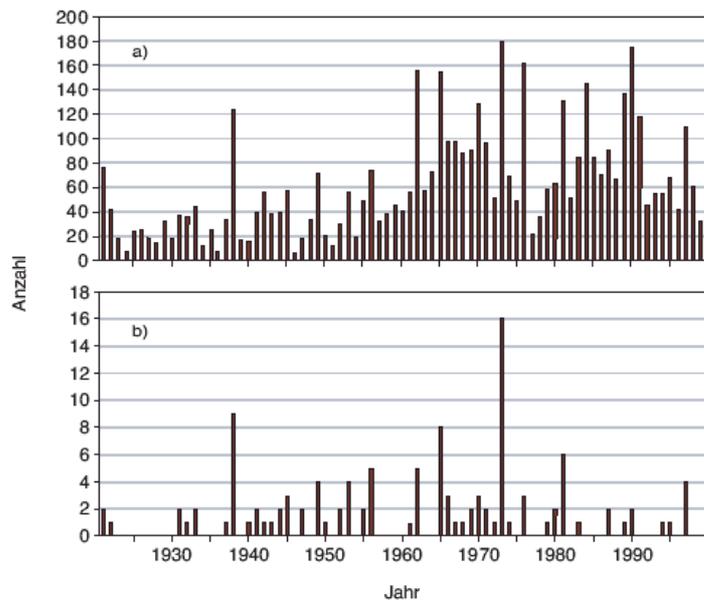


Abb. 2.3: Verteilung (a) der Zahl aller Brände pro Jahr und (b) der Zahl der Extremereignisse, die eine Fläche > 100ha erfassten. Bezugszeitraum 1920-2000 (Conedera, 2003).

Die Entwicklung der Häufigkeit der Waldbrände spiegelt den menschlichen Einfluss wider. Auf der Alpensüdseite verursachte das Vordringen der Waldgebiete und die Anhäufung brennbarer Materialien auf dem Boden seit den 60er Jahren eine grosse Anzahl von Bränden. Die durchschnittlich betroffene Fläche und die Anzahl der Extremereignisse nehmen jedoch vor allem dank der guten Organi-

2. GRUNDLAGEN

sation der Brandverhütung ab. Im Tessin sind entsprechende Massnahmen seit 1980 in Kraft. Seit 1987 gilt die Verordnung zum absoluten Verbot von Feuern im Freien (partielle Lockerung seit 1996). Diese Massnahmen führten zu einer Abnahme der jährlichen Anzahl Brände und somit zu einer Veränderung des Feuerregimes, was bei der Anwendung des entwickelten statistischen Prognosemodells mitberücksichtigt werden muss. Die Voraussetzung von homogenen Verhältnissen (meteorologisch, phytologisch und bezüglich der Verteilung der Brände), welche als Brandbedingung für das k NN-Modell gelten, sind zwischen 1980 und 1990 nicht gegeben. Dies hat zur Folge, dass die Gültigkeit des Modells auf die Periode nach 1991 beschränkt wurde.

Im Gegensatz zum Mittelmeerraum, wo der Höhepunkt der Waldbrandsaison gewöhnlich im Hochsommer erreicht wird, ereignen sich die meisten Brände am Alpensüdhang in den trockenen Wintermonaten, insbesondere zwischen Januar und April. Während der heissen Sommermonate ist ein sekundärer Kulminationspunkt feststellbar, wobei sowohl die Anzahl als vor allem auch die betroffenen Flächen deutlich geringer sind, wie Abbildung 2.4 zeigt. Es handelt sich dabei meist um Oberflächenfeuer (Lauffeuer), welche sich zum Teil sprunghaft ausbreiten, sodass in den Brandgebieten immer wieder nicht verbrannte Inseln verbleiben (Conedera et al., 1997). Diese Tatsache muss bei der Interpretation der ökologischen Folgen von Waldbränden mitberücksichtigt werden.

Durch Landschaftspflege, Gesetzgebung sowie Massnahmen zur Brandverhütung verändert der Mensch die Voraussetzung für Brände. Durch Massnahmen der Brandbekämpfung kontrolliert er die Brände. Trotzdem ist er durch Fahrlässigkeit und Brandstiftung für die meisten Brände selbst verantwortlich. Die Abbildung 2.5 gibt neben einem Überblick der bedeutendsten Waldbrand-Einflussfaktoren auch die prozentuale Verteilung der Ursachen an.

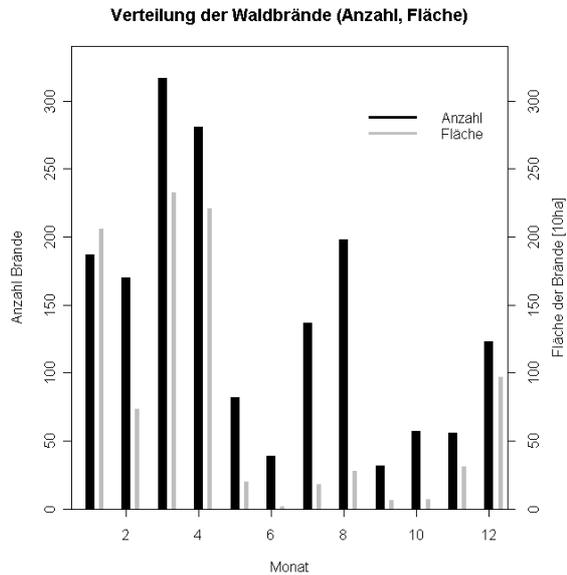


Abb. 2.4: Durchschnittliche monatliche Verteilung der Anzahl Brände und der betroffenen Fläche auf der Alpensüdseite (1981-1999).

			Häufigkeit/ Gewichtung	Zeithorizont
Voraussetzungen	natürlich	Niederschläge	++	ein Tag bis Jahre
		Windhäufigkeit	+	ein bis mehrere Tage
		Windgeschwindigkeit	++	ein Tag
		relative Luftfeuchtigkeit	++	ein Tag
		Sonneneinstrahlung	+	ein Tag
		Feuchtigkeitsdefizit	+	ein Tag bis Jahre
	Mensch	sozioökonomische Entwicklung	+	Monate bis Jahre
		Landschaftspflege	+	Monate bis Jahre
		gesetzlicher Rahmen	++	Monate bis Jahre
		Umwelterziehung	+	Monate bis Jahre
direkt auslösend	natürlich	Blitzschlag	8.6%	
		Mensch	Fahrlässigkeit	26.1%
	Brandstiftung	15.2%		
	Eisenbahn	3.6%		
	Militär	1.7%		
	el. Leitung (Kurzschluss)	1.7%		
	grenzübergreifend	1.7%		
	übrige	2.4%		
Ursache unklar bzw. unbekannt		38.9%		

++ = häufig / massgebend; + = wiederkehrend / einflussreich
Die Prozentangaben beziehen sich auf die Periode 1981–2000.

Abb. 2.5: Zusammenfassung der bedeutendsten Waldbrand-Einflussfaktoren in der Schweiz (Conedera et al., 1996).

2.3 Klima im Tessin

Die klimatischen Eigenschaften des Tessins sind neben der Vegetation die wichtigste Einflussgrösse zur Entstehung von Waldbränden. Die Alpensüdseite wird in Richtung Norden und Westen durch die Alpen begrenzt. Der Einfluss dieses morphologischen Hindernisses auf die atmosphärischen Hauptströmungsrichtung und somit des Wetters kann sehr gross sein. Feuchte und warme west-südwestlichen Strömungen werden oft durch den Alpenkamm in Richtung Norden abgelenkt. Nördliche, alpenüberquerende Strömungen werden zu katabatischen Winden (Nordföhn). Der Föhn bewirkt, dass auf der Alpensüdseite keine Niederschläge fallen. Die aus den nördlichen Richtungen stammenden Luftmassen werden durch den Alpenhauptkamm zum Aufstieg gezwungen. Wird das Kondensationsniveau der aufsteigenden Luftmassen unterschritten, kommt es zur Kondensation respektive zu Niederschlägen. Die Temperaturabnahme auf der Nordseite folgt bis zum Kondensationsniveau dem trockenadiabatischen Temperaturgradienten (ca. $1^\circ K * 10^{-2}m$), danach aufgrund der freiwerdenden Verdunstungswärme dem feuchtadiabatischen Temperaturgradienten (ca. $0.6^\circ K * 10^{-2}m$). Auf der Leeseite des Alpenhauptkammes sinkt die Luft ab und erwärmt sich mit dem trockenadiabatischen Temperaturgradienten. Dies hat zur Folge, dass auf der Südseite die relative Luftfeuchte geringer und die Temperatur auf gleicher Meereshöhe höher ist als auf der Nordseite der Alpen. Die Kanalisation des Luftstroms in der Leventina führt zu einer Zunahme der Windgeschwindigkeit und einer Reduktion der relativen Luftfeuchte. Beides sind entscheidende Faktoren für die Entstehung respektive Ausbreitung von Waldbränden. Der Effekt der Austrocknung des Brandgutes ist vornehmlich auf die Reduktion der relativen Luftfeuchte, und nicht auf die Temperaturzunahme zurückzuführen. Beim Eintritt von Nordföhnsituationen kann die relative Luftfeuchte auf der Alpensüdseite innerhalb kurzer Zeit von 80% auf Werte um 20% fallen. Die Häufigkeit der Nordföhnsituationen erreicht mit Nordweststaulagen im Spätwinter und Frühjahr ihren Höhepunkt. Gleichzeitig erreicht auch die Verteilung der Waldbrände während diesen Perioden jeweils den höchsten Wert.

Die große morphologische Variabilität des Untersuchungsgebietes erschwert eine eindeutige Klassifizierung des Klimas. Es lassen sich mehrere lokale Klimatypen unterscheiden, welche die durchschnittlichen Werte der Wetterphänomene zusammenfassen. Die große Anzahl von Mikroklimatypen bilden ein Mosaik in

den unterschiedlichen Teilen des Gebietes. Die verschiedenen Täler mit ihren unterschiedlichen Orientierungen und Höhen, die “Küstengebiete” um die Seen und der Übergang zur Poebene im südlichsten Teil sind einige Beispiele, die die klimatische Heterogenität der Alpensüdseite charakterisieren.

Die Lufttemperatur, welche auf der Alpensüdseite im gesamtschweizerischen Vergleich die höchsten Werte aufweist (Locarno-Monti: 11.5°C; Lugano: 11.6°C mittlere Jahrestemperatur für den Zeitraum 1961-1990), hat auf die Waldbrandgefahr keinen signifikanten Einfluss. Abbildung 2.6 zeigt Klimadiagramme der Stationen Piotta, Locarno-Monti und Lugano.

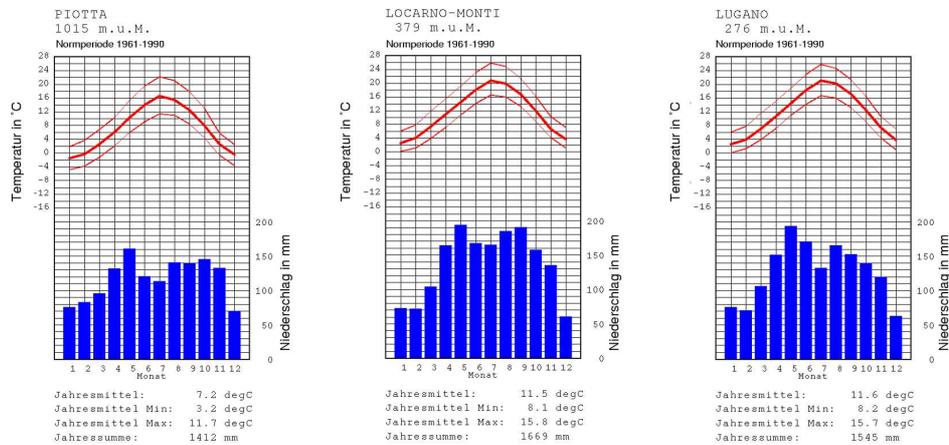


Abb. 2.6: Klimadiagramme der MeteoSchweiz-Stationen Piotta, Locarno-Monti sowie Lugano.

Die Mehrheit der durch Feuer betroffenen Gebiete weisen südliche Expositionen auf, was darauf hindeutet, dass auch die Globalstrahlung einen wichtigen Faktor in Bezug auf die Waldbrandgefahr darstellt. Die höchsten Globalstrahlungswerte werden wiederum bei Nordstaulagen gemessen, bei welchen auf der Alpensüdseite - bedingt durch den Föhn - niedrige Luftfeuchtigkeitswerte und ein geringer Aerosolanteil herrschen.

Das insubrische Klima (*clima insubrico*) der Alpensüdseite ist durch milde und trockene Winter sowie feucht-warme Sommer charakterisiert. Besonders niederschlagsreich sind die Frühjahrs- und Herbstmonate. In den gewöhnlich heißen Sommermonaten kann es zu heftigen konvektiven Gewittern mit hoher Niederschlagsintensität kommen. Nicht selten werden Werte bis 70mmh^{-1} registriert (Alexakis, 2000). Derart starke Niederschlagsereignisse können an steilen Berg-

2. GRUNDLAGEN

flanken zu einer grossen Gefahr von Murgängen führen. Niederschlagsarm sind hingegen die Wintermonate von Dezember bis Februar (Abb. 2.6), wobei eine eigentliche Trockenzeit nicht vorkommt (Alexakis, 2000).

Der Niederschlag stellt einen weiteren, in Bezug auf die Waldbrandgefahr bedeutenden Faktor dar, insbesondere wenn man nicht die Tageswerte, sondern über einen längeren Zeitraum summierte Werte betrachtet. Es scheint paradox, dass die Alpensüdseite, welche eine der niederschlagsreichsten Gebiete der Schweiz darstellt, zugleich die grösste jährliche Anzahl an Waldbränden aufweist. Im Durchschnitt fallen in Locarno-Magadino (1961-1990) 1770mm Niederschlag pro Jahr. Bezüglich des Auftretens von Waldbränden ist jedoch entscheidend, dass die Summe der niederschlagsfreien Tage auf der Alpensüdseite jene nördlich der Alpen deutlich übertrifft. Das Niederschlagsregime weist im Tessin eine wesentlich stärkere jährliche Variation auf. Seit den 80er Jahren lässt sich ein Zunahme geringerer Niederschlagsaktivität zwischen November und März feststellen (Rebetez, 1999).

Möglicher Einfluss einer globalen Klimaänderung

Betrachtet man nicht die jährlichen Niederschlagssummen, sondern jene der trockenen Monate, so wird klar ersichtlich, dass sich daraus aufgrund sehr grosser jährlicher Schwankungen keine allgemeine Tendenz ableiten lässt. Zieht man jedoch nur die extremsten Perioden dieser Serie in Betracht (insbesondere jene mit den längsten Trockenzeiten), so werden sehr interessante Klimatendenzen deutlich. Berücksichtigt man nur Jahre mit Trockenperioden von mindestens zwei Monaten, während welchen nicht mehr als 10mm Niederschlag fielen, so erfüllen im 20. Jahrhundert 15 Jahre dieses Kriterium. Abbildung 2.7 stellt diese allgemeine Tendenz zu immer längeren Trockenperioden dar. Rebetez (1999) zeigte zudem, dass Trockenperioden vornehmlich zu Zeiten starker Temperaturzunahme auftreten. Die intensivsten winterlichen Trockenperioden weisen also eine hohe zeitliche Korrelation zu starken Temperaturzunahmen während dieser Jahre auf. Aus den Analysen der Trockenperioden im Tessin geht jedoch kein direkter Zusammenhang mit einer Zunahme von Waldbränden im entsprechenden Zeitraum hervor, wie ein Vergleich der Abbildungen 2.7 und 2.3 zeigt.

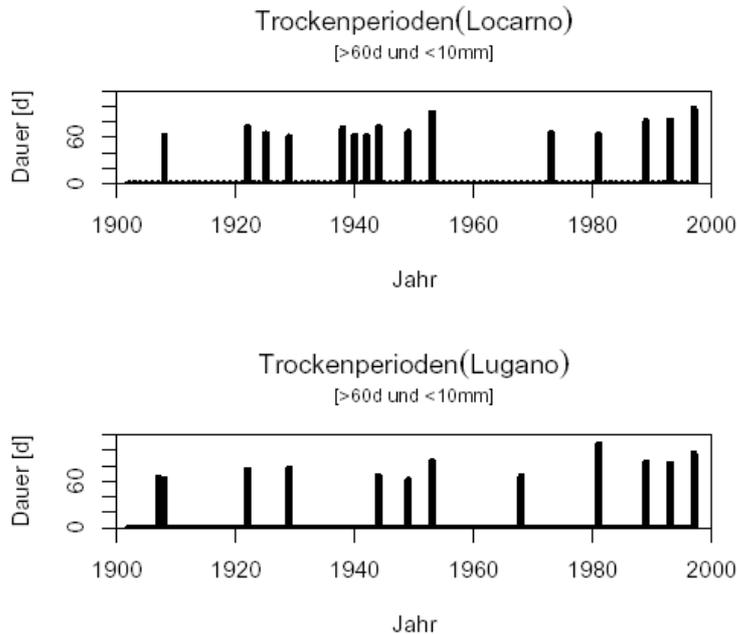


Abb. 2.7: Perioden von mindestens 60 Tagen mit einer Niederschlagssumme $< 10\text{mm}$ im 20. Jahrhundert.

Die Gefahr extremer Waldbrandereignisse wird auch in Zukunft vornehmlich dann eintreten, wenn ausserordentlich förderliche Witterungsbedingungen auftreten. Trotz der abnehmenden Tendenz von Waldbränden (Abb. 2.3 (a)) haben sich an der Alpensüdseite beispielsweise im Frühjahr 1997 nach einer längeren Trockenperiode und während Tagen mit starkem Föhn einige grössere Waldbrände ereignet (Conedera et al., 1998).

Würde es zukünftig infolge der globalen Klimaänderung im Sommer zu niedrigeren Niederschlagsmengen und höheren Temperaturen kommen, könnte dies infolge verminderter Abflüsse regelmässig zu Niederwasser in Vorflutern sowie zu niedrigerer Bodenfeuchte führen. Eine mögliche Klimaänderung könnte somit die Wahrscheinlichkeit erhöhen, dass mehrere förderliche Voraussetzungen für Brände zusammentreffen. Einerseits wird die zukünftige Entwicklung von Extremereignissen bei Waldbränden weiterhin mit menschlichem Handeln wie Landschaftspflege, Brandverhütung und Brandbekämpfung verknüpft sein. Andererseits spielt die Häufigkeit von extremen Wetterlagen wie längeren Trockenperioden in Verbindung mit starken Winden eine grosse Rolle. In diesem Sinne muss

2. GRUNDLAGEN

eine Erforschung der Auftrittstendenzen von Waldbränden stets beide Faktoren berücksichtigen (Conedera, 2003). Nach Rebetez (1999) ist dagegen allgemein damit zu rechnen, dass im 21. Jahrhundert aufgrund der zu erwartenden globalen Erwärmung ungeachtet aller soziökonomischen Veränderungen die winterliche Waldbrandgefahr auf der Alpensüdseite jedenfalls weiter zunehmen wird.

2.4 Vegetation im Tessin

Sämtliche Vegetationsschichten (Kraut-, Strauch-, Baumschicht) können durch Feuer mehr oder weniger stark betroffen und zerstört werden. Eine hohe Individuenzahl an Laubbäumen auf der gesamten Alpensüdseite stellt viel potentielles Brandgut dar, was die Ausbruchswahrscheinlichkeit und die rasche Ausbreitung von Bränden erheblich erhöht.

Die Mehrzahl der Feuer entsteht in einer trockenen Krautschicht (*Molinia littoralis*, *Calamagrostis*, *Festuca* und *Pteridium aquilinum*) (Ceschi, 1977). In der Strauchschicht begünstigen vor allem folgende Arten eine weitere Feuerausbreitung: *Calluna vulgaris*, *Sarotamnus scoparius* und *Juniperus* (Ceschi, 1977). Bäume müssen in die Gruppe Laub- und Nadelhölzer (von Bränden relativ selten betroffen) unterteilt werden. Innerhalb der Gruppe der vorherrschenden Laubbäume können verschiedene "Risikoklassen" unterschieden werden. Da die Humifizierung der Edelkastanie (*Castanea sativa*) im Vergleich zu anderen Laubbäumen am meisten Zeit beansprucht, stellt sie aufgrund der mächtigen bodenbedeckenden Mull- und Moderschicht für Brände die grösste Gefährdung dar (Ellenberg, 1996). Neben der Art spielt das Alter der Bäume bezüglich der Waldbrandgefahr eine Rolle. Junge Pflanzen sind aufgrund ihrer geringmächtigeren Rinde weniger geschützt und somit anfälliger.

Die am häufigsten von Bränden betroffene Edelkastanie (*Castanea sativa*) wächst hauptsächlich in wintermilden Lagen der kollinen und submontanen Stufen und bevorzugt mässig trockene bis feuchte, nährstoffarme Böden (mässig frische, lockere, auch podsolige Braunerde; meidet kalkhaltige Böden; (Ellenberg, 1996)).

Die grösste Auftrittshäufigkeit weisen innerhalb der vielverbreiteten Kastanienbestände die Oberflächenfeuer (Lauffeuer) auf, bei welchen hauptsächlich die Humusschicht respektive Streu- und Unterholz zerstört werden.

3 Prognosemodelle

3.1 Einführung

Um Aussagen bezüglich der Gefährdung eines Gebietes durch eine Naturgefahr zu treffen, müssen zahlreiche komplexe Zusammenhänge bekannt sein. Für die Einschätzung der Gefahr müssen daher in grossem Umfang viele Daten und Informationen aus verschiedenen Messnetzen und Regionen miteinander verknüpft werden.

Ziel der Modellierung von solcherart komplexen Problemen ist, die zum Teil bedrohlichen Konsequenzen von teilweise unverstandenen Prozessen vorauszusagen, ohne von einer mehr oder weniger subjektiven Einschätzung durch Experten abhängig zu sein. Solche Expertisen sind zwar oft von guter Qualität, aber wegen der nicht gegebenen Nachvollzieh- und Reproduzierbarkeit ein für die wissenschaftliche Zwecke ungeeignetes Mittel.

Ziel der vorliegenden Arbeit ist, für die Naturgefahr Waldbrand ein Prognosemodell zu entwickeln, welches sowohl den Anforderungen der Wissenschaft als auch der Praxis entspricht.

3.2 Modelltypen

Die reale Umwelt, oder auch jedes wahrnehmbare System in ihr, ist zu komplex, um sie in ihrer Gesamtheit zu verstehen oder zu definieren. Bei der Behandlung eines einzelnen Problems ist man daher darauf angewiesen, das System gegenüber der Realität vereinfacht darzustellen. Eine vereinfachte und daher leichter definierbare und handhabbare Version der Realität ist ein Modell (Hebel, 2003).

Trotz einer Vielzahl von System- respektive Modellbegriffen gibt es keine einheitliche Systematik. Im Folgenden wird eine Unterteilung von Modellen nach der Art der Prozessbeschreibung durchgeführt. Danach werden folgende Modelltypen unterschieden:

3. PROGNOSEMODELLE

- (a) empirische Modelle:** Dies sind einfache Ursachen-Wirkungsmodelle, bei welchen keine physikalische Beschreibung der Prozesse stattfindet. Erkenntnisse über das Systemverhalten können dabei nicht gewonnen werden und dementsprechend ist keine unmittelbare Übertragbarkeit auf andere Standorte und Bedingungen möglich. Ihr Einsatz haben diese Modelle in den Bereichen, wo eine exakte physikalische Beschreibung der Prozesse nicht bekannt ist. Als Beispiele könnten diverse, zur Waldbrandprognose eingesetzte Modelle mit berechneten Indizes, basierend auf meteorologischen Bedingungen, erwähnt werden.
- (b) deterministische Modelle:** Sie beschreiben die Änderung der Zustandsgrößen eindeutig aus dem jeweiligen Systemzustand und den Auswirkungen von aussen (Bossel, 1992). Eine optimale Lösung, bei der alle Prozesse und Randbedingungen des Systems korrekt wiedergegeben werden, existiert in der Realität äusserst selten. Um eine Näherungslösung zu erreichen, werden analytische oder numerische Lösungsverfahren eingesetzt.
- (c) stochastisch - statistische Modelle:** Bei diesem Modelltyp wird ein statistischer Zusammenhang zwischen einer Ziel- und erklärenden Grössen hergestellt. Im Unterschied zu empirischen- oder deterministischen Modellen verlangen statistische Modelle in Abhängigkeit der Komplexität der zu beschreibenden Prozesse eine mehr oder weniger umfangreiche Datenbank. Als Beispiele können hier die k NN-Methode oder Regressionsmodelle genannt werden.

Anhand der Anforderungen an das zu entwickelnde Waldbrandprognosemodell unter Berücksichtigung diverser Einschränkungen (vgl. Kap. 1.2), muss entschieden werden, welcher Modelltyp implementiert werden soll. Zu den Einschränkungen sind beispielsweise Faktoren wie Datengrundlage oder Kenntnisse der ablaufenden physikalischen Prozesse zu zählen.

Modelle vom Typ (a) und (c) sind erst dann zu verwenden, wenn kein oder nur ein ungenügendes physikalisches Verständnis der grundlegenden Prozesse vorhanden ist. Kennt man die Physik der ablaufenden Prozesse, ausreichend genau, liefert Typ (b) die besten Modellresultate.

Im vorliegenden Fall wurde, basierend auf der Komplexität der zu einem Waldbrand führenden Prozesse ein stochastisches Modellverfahren (c) zur Prognose

der Waldbrandgefahr ausgewählt.

3.3 Schritte der Modellentwicklung

Die eigentlichen Schritte der Modellentwicklung verlaufen unabhängig vom gewählten “Modelltyp” nach demselben Schema (Abb. 3.1).

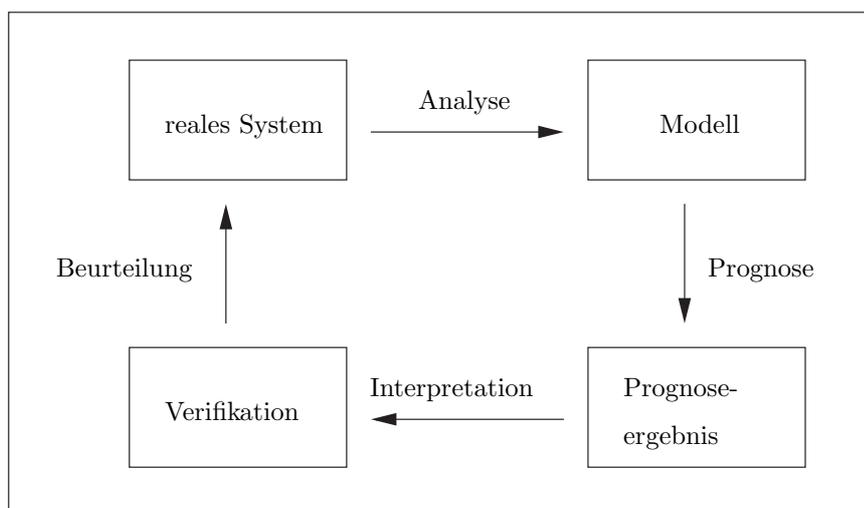


Abb. 3.1: Schematische Darstellung der wichtigsten Schritte und Abläufe der Modellentwicklung.

Die Eigenschaften eines realen Systems müssen vor einer Modellentwicklung analysiert werden. Diese Analysen können beispielsweise anhand unterschiedlicher statistischer Auswertungen der grundlegenden Prozesse vorgenommen werden. Nach der Entwicklung des Modells, müssen die Modellberechnungen (Prognosen) durchgeführt werden. Die damit berechneten Ergebnisse müssen in der Folge interpretiert und verifiziert werden. Anschliessend kann beurteilt werden, ob die Modellqualität den geforderten Ansprüchen genügt oder ob Verbesserungen notwendig sind.

Bei der Entwicklung eines Modells ist es wichtig, dass man sich ungefähr an diesen skizzierten Ablauf hält, um allfällige Schwächen bereits im Entscheidungsprozess zu erkennen. Jeder Einzelschritt muss zudem auf dessen Plausibilität überprüft werden, damit sichergestellt ist, dass das entwickelte Modell die eingangs defi-

3. PROGNOSEMODELLE

nierten Anforderungen erfüllt.

Allgemein müssen alle Prognosemodelle immer unter dem Gesichtspunkt der Rahmen- und Einsatzbedingungen erarbeitet werden.

3.4 Prognosemodelle in der Waldbrandforschung

In der Waldbrandforschung stellen Modelle zur Prognose der Waldbrandgefahr nur einen kleinen Teilaspekt im Bereich der Feuerprävention dar. Themen wie Feuererkennung, Feuerausbreitung, Risikoabschätzungen oder Feuerökologie haben hingegen eine weit grössere Bedeutung. Verlässliche Prognosemodelle können für das Risikomanagement von Waldbränden dennoch sehr hilfreich sein. In den meisten Fällen lässt sich ein prognostizierter Brandausbruch zwar nicht verhindern, doch können durch frühzeitige Gefahrenerkennung präventive Massnahmen getroffen werden. So können beispielsweise die Feuerwehr in potentiell gefährdeten Gebieten frühzeitig verstärkt und in Alarmbereitschaft gesetzt oder touristische Gebiete wie Nationalparks für die Öffentlichkeit geschlossen werden. Brände, welche vorsätzlich, beispielsweise im Rahmen von Bodenpreisspekulationen, gelegt werden, können von Prognosemodellen hingegen kaum berücksichtigt werden. Ein Modell ist prinzipiell nur in der Lage eine Aussage über eine potentielle Gefährdung zu machen. Wo ein Brandstifter möglicherweise einen Brand legt, kann nicht vorhergesehen werden.

Gegenwärtig gibt es einige wenige Modelle, welche zur Prognose von Waldbränden im Einsatz sind. Die meisten basieren auf dem Kanadischen Feuer-Wetter-Index System (FWI¹) oder ähnlichen Ansätzen.

Aufgrund der Komplexität der für die Entstehung eines Waldbrandes massgeblichen Prozesse basieren die meisten der erwähnten Prognosemodelle auf empirischen Ansätzen. Dabei stellen meistens meteorologischen Variablen die Grundlage dar. Aufbauend auf dem FWI wurden in verschiedenen südeuropäischen Ländern diverse Indizes mit unterschiedlichem Erfolg erarbeitet (Viegas et al., 1999a). In der Schweiz wird ein indexbasierter Ansatz (Incendi) im Kanton Graubünden

¹FWI: Das Kanadische Feuer-Wetter-Index System berücksichtigt fünf unterschiedliche, waldbrandbeeinflussende Grössen um den FWI zu berechnen: *Fine Fuel Moisture Code*, *Duff Moisture Code*, *Drought Code*, *Initial Spread Index* und den *Buildup Index* (Van Wagner, 1987). Unter Berücksichtigung dieser Grössen wird der FWI Index berechnet, anhand dessen eine 4-stufige Gefährdung bestimmt wird.

durch das Geographische Institut der Universität Zürich angewandt. Die berechnete Gefahrenstufe ist täglich im Internet abrufbar.

Tabelle 3.1 gibt einen Überblick über die am weitesten verbreiteten Waldbrandprognosemodelle. Eine Vollständigkeit dieser Liste kann nicht gewährleistet werden, da eine Vielzahl der entwickelten Modelle lokal und zum Teil nur kurzfristig eingesetzt werden. Einige Modelle werden mit Ausbreitungsmodellen integriert und decken dadurch weitere Fragestellungen im Bereich des Risikomanagements ab.

Tab. 3.1: Zusammenfassung und Überblick einiger weitverbreiteter Waldbrandprognosemodelle.

Name	Entwickler	Einsatzgebiet	Methode
FWI	Van Wagner (1987)	CDN	
NFDRS	Deeming et al. (1977)	USA	Wetter, Brandgut, Brandgutfeuchte
Risque numérique	Sol (1989)	F	
IREPI	Bovio et al. (1994)	I	
DFR	Garcia Diez et al. (1999)	Galizien, E	
IP	Goncalvez & Lourenco (1990)	P	
BEHAVE	Burgan & Rothermel (1984)	USA, Europa	
INMG	diverse	P, Europa	
ICONA	diverse	E, Europa	
Incendi	Schöning et al. (1998)	Graubünden	
EFFIS	Joint research center, ISPRA	Europa	Zusammenfassung der verschiedenen europäischen Indices

Neben den genannten indexbasierten empirisch-(deterministischen) Modellen, gibt es rein statistische Modellansätze, wie die von Mandallaz & Ye (1996); Langhart (1999) oder Felber & Bartelt (2002) entwickelten Prognosemodelle. Diese fallen

3. PROGNOSEMODELLE

häufig, aber nicht ausschließlich, in die Gruppe der statistisch - stochastischen Modelle (vgl. Kap. 3.2).

3.5 Klassifikation von Ereignissen

Das Ziel jedes Prognosemodells liegt darin, ein bekanntes Ereignis frühzeitig zu prognostizieren (vgl. Kap. 3.1). Durch Einschränkung mit einer ausschliesslichen Berücksichtigung von binären Ereignissen, lässt sich die Problematik vereinfacht modellieren. Das grundlegende Problem dabei ist, einen Satz von beschreibenden und messbaren Variablen zu finden, welcher Zeitpunkte (Tage) mit und ohne Ereignisse in dem durch die Variablen aufgespannten Ereignisraum unterscheidet. Diese Fragestellung wird als klassisches Problem der Diskriminierung von binären Ereignissen bezeichnet.

Das Grundproblem der Diskriminierung binärer Ereignisse lässt sich unabhängig vom gewählten Modelltyp darstellen. Man versucht im Ereignisraum Ω die Tage mit und die jene ohne Ereignisse voneinander zu trennen. Dazu sollte ein statistischer Zusammenhang zwischen den an den Prozessen mitbeteiligten Grössen (Variablen) und dem Ereignisseintritt hergestellt werden können.

Mit Hilfe der multivariaten Datenanalyse, welche eine Systematisierung von Merkmalsmustern ermöglicht, können die genannten Klassifikationsprobleme gelöst werden. Besonders hervorzuheben ist die Diskriminanzanalyse (DA). Ziel der Diskriminanzanalyseverfahren ist, einen geeigneten Variablensatz zu finden, sodass sich in dem von diesen aufgespannten Ereignisraum Ω die positiven von den negativen Ereignissen durch eine global anwendbare Funktion unterscheiden lassen. Dies ist eine praktische Problemstellung, wie sie in verschiedenen Bereichen der Medizin, Biologie oder der Meteorologie häufig auftritt.

Diskriminanzanalyseverfahren lassen sich in zwei Hauptgruppen einteilen:

- **Parametrische Verfahren zur Klassifikation:** Diese zeichnen sich dadurch aus, dass sie Annahmen über die globale parametrische Form der klassenspezifischen Dichten, der Form der Trennfläche oder der Form des Dichtequotienten treffen. Die lineare- beziehungsweise die quadratische Dis-

kriminanzanalyse kann nur dann zur Lösung von Klassifikationsproblemen eingesetzt werden, wenn die Variablen normalverteilt sind und gleiche beziehungsweise unterschiedliche Kovarianzmatrizen aufweisen (Nothnagel, 1999).

- **Nichtparametrische Verfahren zur Klassifikation:** Diese Verfahren verzichten auf globale Annahmen über Dichte oder Form der Trennflächen. Sie werden auch als verteilungsfreie Methoden bezeichnet. Die k -Nearest Neighbors (k NN) Methode wird ebenso wie beispielsweise Neuronale Netze (NNnet) oder Entscheidungsbäume *Classification trees* (CT) dieser Gruppe zugeordnet.

Häufig werden die beiden genannten Verfahren unter dem Begriff *Data Mining* (Datenmustererkennungsverfahren) zusammengefasst.

Die Wirkung der Diskriminierung und somit die Qualität des zu entwickelnden Modells hängt sehr stark von folgenden Umständen ab:

- Es müssen alle Variablen, welche an der Entstehung des Ereignisses mitbeteiligt und somit zur Beschreibung des Prozesses erforderlich sind, im Modell enthalten sein.
- Um die Klassifizierung zu optimieren, müssen die entsprechenden Variablen in zweckmässiger Weise gewichtet werden.

Sind diese beiden Bedingungen in ausreichendem Masse erfüllt, so kann mit Hilfe einer der zuvor angeführten statistischen Methoden ein zweckmässiges Prognosemodell entwickelt werden.

4 Die k -Nearest Neighbors (k NN) Methode

4.1 Einführung

Obwohl man wegen der Komplexität des zu modellierenden Prozesses im vorliegenden Fall einen deterministischen Ansatz ausschliessen muss, stehen eine Reihe statistischer und empirischer Modelle zur Auswahl (vgl. Kapitel 3.5), welche zur Lösung des Diskriminierungsproblems in Betracht gezogen werden können. Die folgenden zwei Modelle wurden zur Prognose der Naturgefahr Waldbrand auf der Alpensüdseite im Rahmen der vorliegenden Arbeit entwickelt:

- k NN-Modell: Nichtparametrische Lösung des Klassifikationsproblems, wurde aufgrund des erfolgreich im Einsatz stehenden NXD²-Lawinenprognosemodells entwickelt.
- LR-Modell³: Parametrisches Lösungsverfahren des Klassifikationsproblems, wurde als Teil des k NN-Modells entwickelt.

Da bei jedem Modelltyp sehr spezifische Anforderungen und Randbedingungen erfüllt werden müssen, wäre das Waldbrand-Prognoseproblem grundsätzlich durch verschiedene Verfahren zu lösen gewesen. Die Modellwahl wurde schliesslich aufgrund des bereits erfolgreich im Einsatz stehenden k NN-Lawinenprognosemodells getroffen.

Basierend auf demselben Ansatz kam demnach die Idee auf, auch für die Waldbrandprognose ein k -Nearest Neighbors Modell zu entwickeln.

In den 80iger Jahren führte Buser (1983) bzw. Buser et al. (1986) die k NN-Methode im Bereich der Lawinenprognose ein. Diverse Weiterentwicklungen von

² k NN-Modell: NXD-Lawinen.

³LR: logistisches Regressionsmodell zur Prognose der Waldbrandgefahr, entwickelt von Felber & Bartelt (2002).

4. DIE k -NEAREST NEIGHBORS (k NN) METHODE

Kristensen & Larsson (1994); Meister & Brabec (1999); Gassner et al. (2000); Brabec (2001) und Purves et al. (2002) führten zu einer weiten Verbreitung des k NN-Ansatzes vor allem bei lokalen Lawinenwarndiensten in der Schweiz und im Ausland. Zahlreiche Verifikationen bestätigten die hohe Qualität des Prognosemodells, und folglich wurde eine thematische Weiterentwicklung des Modells im Rahmen der vorliegenden Arbeit lanciert.

Die potentielle Gefahr von Ereignissen wie Waldbränden und Lawinen wird primär durch meteorologische Prozesse gesteuert (vgl. Kap. 2.1). Im Falle der Waldbrände ist es das Brandgut, welches die potentielle Gefährdung beschreibt, bei den Lawinen die Schneedeckenstabilität. Der Zustand letzterer lässt sich dabei durch meteorologische Abläufe erklären. Dazu müssen auch mittel- und längerfristige meteorologische Entwicklungen berücksichtigt werden. Dies geschieht, indem weiter zurückliegende Werte, wie zum Beispiel die 30-Tagesniederschlagssumme, im Modell mitberücksichtigt werden.

Sowohl bei Waldbränden als auch bei Lawinen kann man das Zustandekommen eines Ereignisses in drei Bereiche aufteilen. Dabei wird zwischen der Grunddisposition (i), der variablen Disposition (ii) und dem auslösenden Ereignis (iii) unterschieden (Abb. 4.1).

4. DIE k -NEAREST NEIGHBORS (k NN) METHODE

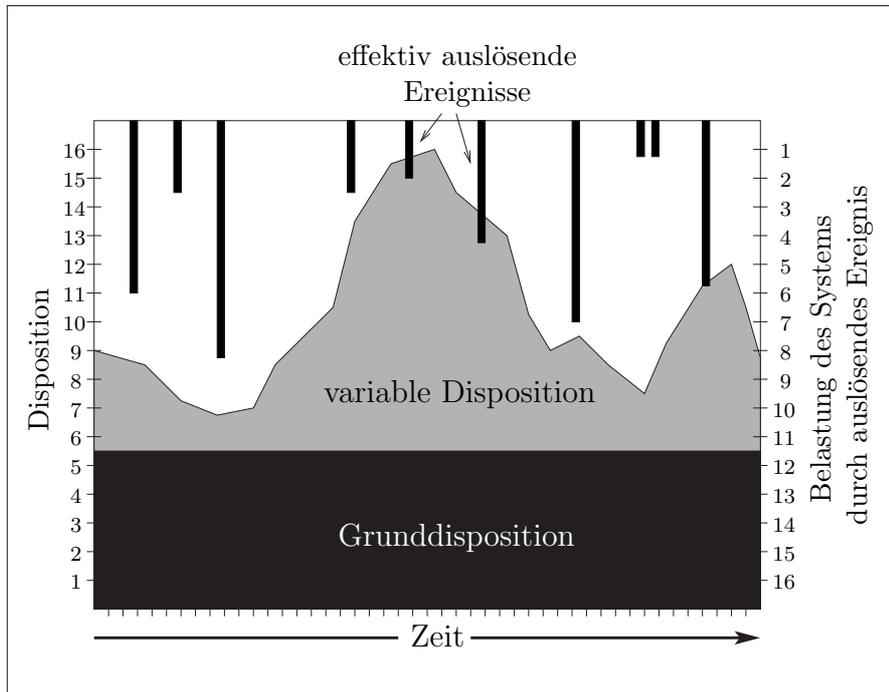


Abb. 4.1: Unterscheidung der drei Hauptfaktoren Grunddisposition (i), variable Disposition (ii) und auslösendes Ereignis (iii), welche zu einem Ereignis führen können. (Verändert nach Mani (1995)).

Die Bewertung der auslösenden Ereignisse und der variablen Disposition erfordert in der Regel einen grösseren methodischen Aufwand als die Ermittlung der Grunddisposition.

Die Grunddisposition zu gefährlichen Prozessen ist die prinzipielle, über längere Zeit gleichbleibende Anlage oder Bereitschaft gefährlicher Prozesse, abzulaufen. Sie wird durch über längere Zeit konstant bleibende Parameter wie Relief oder Klima bestimmt.

Die variable Disposition ist die bei gegebener Grunddisposition zeitlich veränderliche, in einem bestimmten Umfang schwankende oder sich entwickelnde Disposition für gefährliche Prozesse. Hierzu könnten beispielsweise die Vegetation, die Jahreszeit, Trockenheit oder längere Föhnperioden zusammengefasst werden.

Die aktuelle Gesamtdisposition (Potential) wird aus Grund- und variabler Disposition zusammengesetzt. Prognosemodelle können nur die aktuelle Disposition modellieren, das auslösende Ereignis selbst kann hingegen nicht modelliert werden.

4. DIE k -NEAREST NEIGHBORS (k NN) METHODE

Das auslösende Ereignis (z.B. Blitzschlag, etc. bei Waldbränden respektive zusätzlicher Neuschnee, etc. bei Lawinen), setzt bei einer gegebenen aktuellen Disposition den gefährlichen Prozess in Gang. Die Auslösung der gefährlichen Prozesse erfolgt dann, wenn der Schwellenwert eines oder mehrerer kritischer Faktoren erreicht bzw. überschritten wird. Je ausgeprägter jeweils die aktuelle Disposition zu einem gefährlichen Prozess ist, desto wahrscheinlicher ist es, dass der Schwellenwert von kritischen auslösenden Faktoren überschritten wird (BUWAL, 1998).

Abbildung 4.2 zeigt einen schematisch vereinfachten Überblick über die Gemeinsamkeiten der beiden Naturgefahren Waldbrand und Lawinen in Bezug auf eine Prognosemodellierung. Dabei kommt die Unterscheidung in Grunddisposition (Brandgut, Wetter, Gelände respektive Schneedecke, Wetter, Gelände) und variable Disposition (Luftfeuchte, Temperatur, Niederschlag respektive Neuschnee, Stabilität, Wind) deutlich zum Ausdruck, welche den modellierbaren Teil der Prozesse ausmachen.

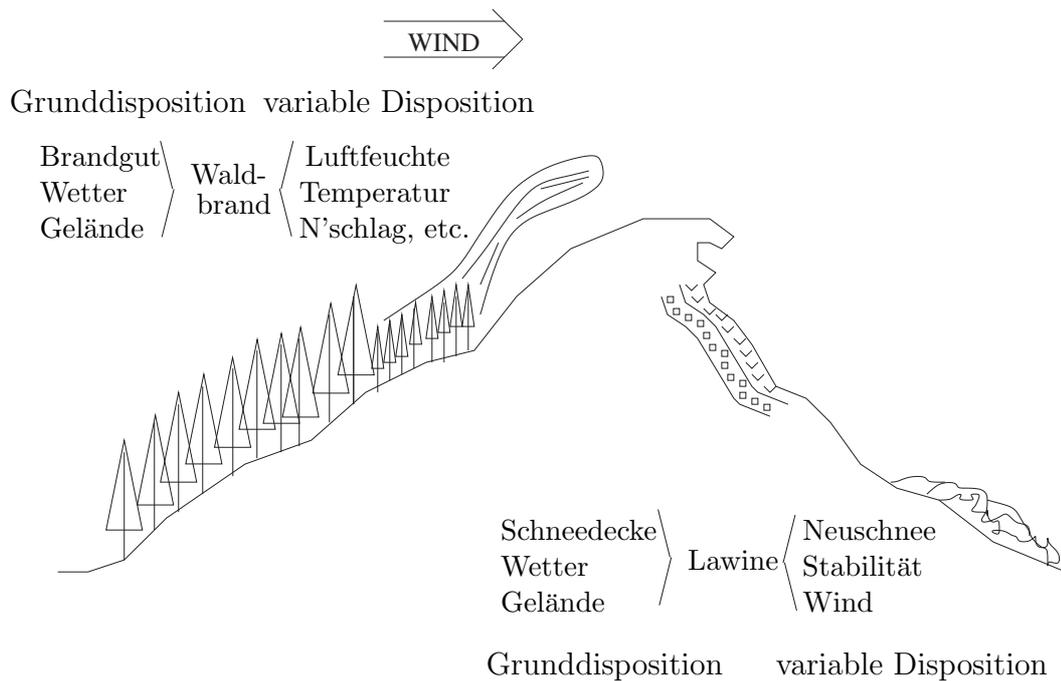


Abb. 4.2: Konzeptionelles Schema, welches den Zusammenhang zwischen Waldbränden und Lawinen in Bezug auf die Prognosemodellierung wiedergibt. (Verändert nach Jenkins & Hebertson (2000)).

4.2 Grundlagen

“Do as your neighbors do”
(Adriaans & Zantige, 1996a).

Fix & Hodges (1951, 1952) führten in den 50iger Jahren mit ihren Publikationen die k -Nearest Neighbors (k NN) Methode als nichtparametrische Diskriminanzanalyse ein. Darauf aufbauend führten Cover & Hart (1967) die k -Nearest Neighbors Regeln erstmals formal an. In den folgenden Jahren folgten zahlreiche Untersuchungen zur k NN-Methode. Ein umfassender und vollständiger Überblick über die k NN-Ansätze kann in Dasarathy (1991) gefunden werden. Die meisten Arbeiten über Data Mining (Adriaans & Zantige, 1996b; Weiss & Indurkha, 1998) und Machine Learning (Mitchie et al., 1994; Mitchell, 1997; Ripley, 1996) besprechen die k NN-Ansätze im Rahmen von distanz- oder fallbasierten Methoden.

Der k -Nearest Neighbors Ansatz basiert auf der Annahme, dass heute eine ähnlich grosse Waldbrandgefahr herrscht, wie an einem vergangenen Tag mit einer der heutigen vergleichbaren meteorologischen Situation. Grundgedanke eines k NN-Modells ist demnach, dass man eine umfangreiche Datenbank zusammenstellt, in welcher nach Tagen gesucht wird, die in Bezug auf ihre meteorologischen Variablen dem Prognosetag ähnlich sind. Man prüft dann, ob an den identifizierten Tagen Waldbrände auftraten. Weisen die referenzierten Tage (Nächste Nachbarn) das Vorkommen von Bränden auf, dann ist auch am Prognosetag höchstwahrscheinlich mit Waldbränden zu rechnen.

Voraussetzung für ein Funktionieren dieser Prognosemethode ist, dass sich ein Variablensatz findet, sodass sich Tage mit Waldbränden innerhalb des durch die Variablen aufgespannten Ereignisraumes gruppieren lassen.

4.3 Die k NN-Methode

Die meteorologische Situation an einem bestimmten Tag wird durch $p \in \mathbb{N}$ meteorologische Variablen beschrieben, sodass jedem Tag ein Vektor $\mathbf{x}=(x_1, \dots, x_p)^\dagger$ zugeordnet werden kann. Der Raum, welcher dadurch aufgespannt wird, heisst Ereignisraum Ω . Aus praktischen Gründen kann eine Datenbank nur $n \in \mathbb{N}$ Ein-

4. DIE k -NEAREST NEIGHBORS (k NN) METHODE

träge enthalten. Jedem gemessenen Punkt \mathbf{x} wird ein Attribut $+$ bzw. $-$ beigeordnet, je nach dem, ob sich an diesem Tag Waldbrände ereignet haben oder nicht. Einträge mit positivem Attribut werden der Klasse Ω^+ , solche mit negativem der Klasse Ω^- zugeordnet. Die Prognose beruht in der Folge darauf, einen zu prognostizierenden Tag \mathbf{y} entweder der Klasse Ω^+ oder der Klasse Ω^- zuzuordnen.

Die k NN-Methode findet für einen Prognosetag $\mathbf{y} = (y_1, \dots, y_p)^\dagger$ mittels dem Mass

$$D(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^p (\hat{x}_i - \hat{y}_i)^2} \quad (4.1)$$

die nächsten $k \in \mathbb{N}$ Nachbarn $\mathbf{x}_1, \dots, \mathbf{x}_k$, wobei \hat{x}_i und \hat{y}_i die skalierten und gewichteten Variablen darstellen. Danach wird festgestellt, wie gross die Anzahl η^+ der gefundenen Nachbarn ist, welche ein positives Attribut tragen.

Es gibt drei verschiedene Möglichkeiten, wie man aus den gefundenen k Nachbarn eine Prognose für \mathbf{y} erstellt:

- Die **binäre Prognose** stellt einen einfachen, dichotomen Ja-Nein-Entscheid dar. Dazu muss ein Schwellenwert DB (*decision boundary*) definiert werden. Ist $\eta^+ \geq DB$ wird der zu prognostizierende Tag in die Kategorie Ereignistag (Ω^+) eingeteilt, andernfalls wird er der Kategorie Nichtereignistag (Ω^-) zugeordnet.
- Bei der **probabilistischen Prognose** wird eine Wahrscheinlichkeit

$$p_y = \frac{\eta^+}{k} \quad (4.2)$$

für das Eintreten eines Ereignisses an dem zu prognostizierenden Tag \mathbf{y} berechnet. Diese Prognose liefert mehr Informationen über die Lage von \mathbf{y} , was jedoch dazu führt, dass diese Informationen für die Prognose zunächst noch interpretiert werden müssen.

- Bei der **deskriptiven Prognose** wird im Gegensatz zu den anderen Prognosen die gesamte Information, welche in der Datenbank über die nächsten k Nachbarn von \mathbf{y} enthalten ist, ausgegeben. Es bleibt dem Benutzer des Modells überlassen, aus den Daten die entsprechenden Interpretationen vorzunehmen.

Anstelle der k NN-Methode, bei der k zu \mathbf{y} nächsten Nachbarn aus der Datenbank gesucht werden, können statt einer genau definierten Anzahl von Nachbarn

4. DIE k -NEAREST NEIGHBORS (k NN) METHODE

diejenigen Tage \mathbf{x}_j gesucht werden, für die gilt:

$$D(\mathbf{x}_j, \mathbf{y}) \leq R, \quad \text{mit} \quad (j > 0). \quad (4.3)$$

Diese Methode liefert durch die Anzahl gefundener Nachbarn einen Hinweis auf die Dichte der bestehenden Messpunkte in Ω an der Stelle \mathbf{y} und somit indirekt über die Qualität der Prognose.

Abbildung 4.3 stellt anhand eines Beispiels graphisch die Vorgehensweise der k NN-Methode zur Lösung des vorliegenden Klassifikationsproblems dar. Zur einfachen und übersichtlichen Darstellung wird in diesem Beispiel nur in einem 2-dimensionalen Ereignisraum Ω nach den nächsten Nachbarn gesucht. Zum Prognosezeitpunkt \mathbf{y} , der bestimmt durch seine meteorologischen Variablen irgendwo im Ereignisraum Ω zu liegen kommt, werden anhand der Gleichung 4.1 die k nächsten Nachbarn bestimmt. Diese werden anschliessend auf ihr jeweiliges Attribut überprüft. Je nach gewünschter Ausgabeform müssen die Anzahl gefundener positiver Nachbarn η^+ beurteilt werden; Dies in Bezug zu einem gewählten Schwellenwert DB für die binäre Prognose, bzw. in Bezug zur Anzahl k für die probabilistische Prognose. Bei der deskriptiven Prognose werden alle im Kreis um den Prognosezeitpunkt \mathbf{y} liegenden Tage \mathbf{x} in Form einer Liste ausgegeben.

4. DIE k -NEAREST NEIGHBORS (k NN) METHODE

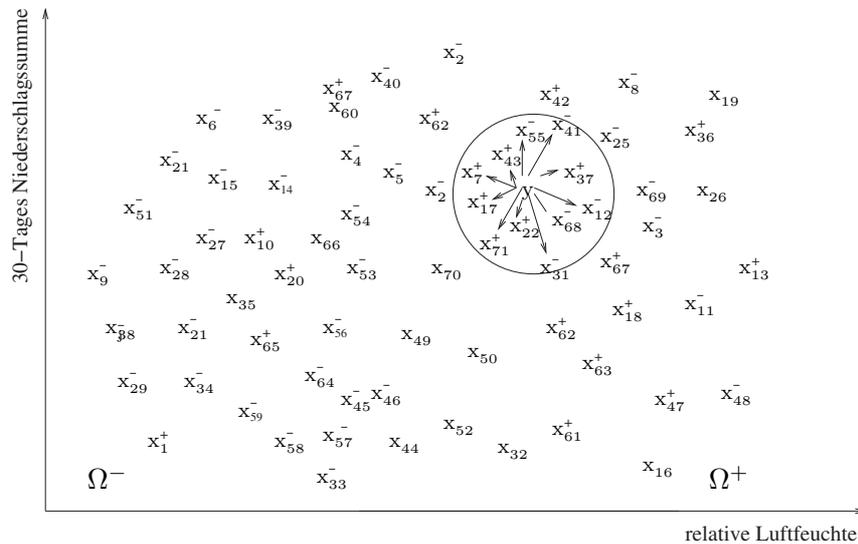


Abb. 4.3: Systematische Darstellung der Diskriminierung von \mathbf{x}^+ (Ereignistagen) und \mathbf{x}^- (Nichtereignistagen) in die beiden Klassen Ω^+ und Ω^- mittels k NN-Methode. Um den zu prognostizierenden Tag \mathbf{y} wird ein Kreis (bei $p = 2$) gelegt, in welchem die k nächsten Nachbarn liegen.

4.4 Entwicklung des k NN-Modells

Die Entwicklung eines komplexen Prognosemodells erfordert eine klare Strukturierung der einzelnen, zu bearbeitenden Abläufe. Die Einhaltung dieser chronologisch aufeinanderfolgenden Arbeitsschritte ist wichtig, nur so können allfällige Probleme frühzeitig erkannt werden. Abbildung 4.4 zeigt die einzelnen Abläufe des in Rahmen dieser Arbeit entwickelten Waldbrand-Prognosemodells.

4. DIE k -NEAREST NEIGHBORS (k NN) METHODE

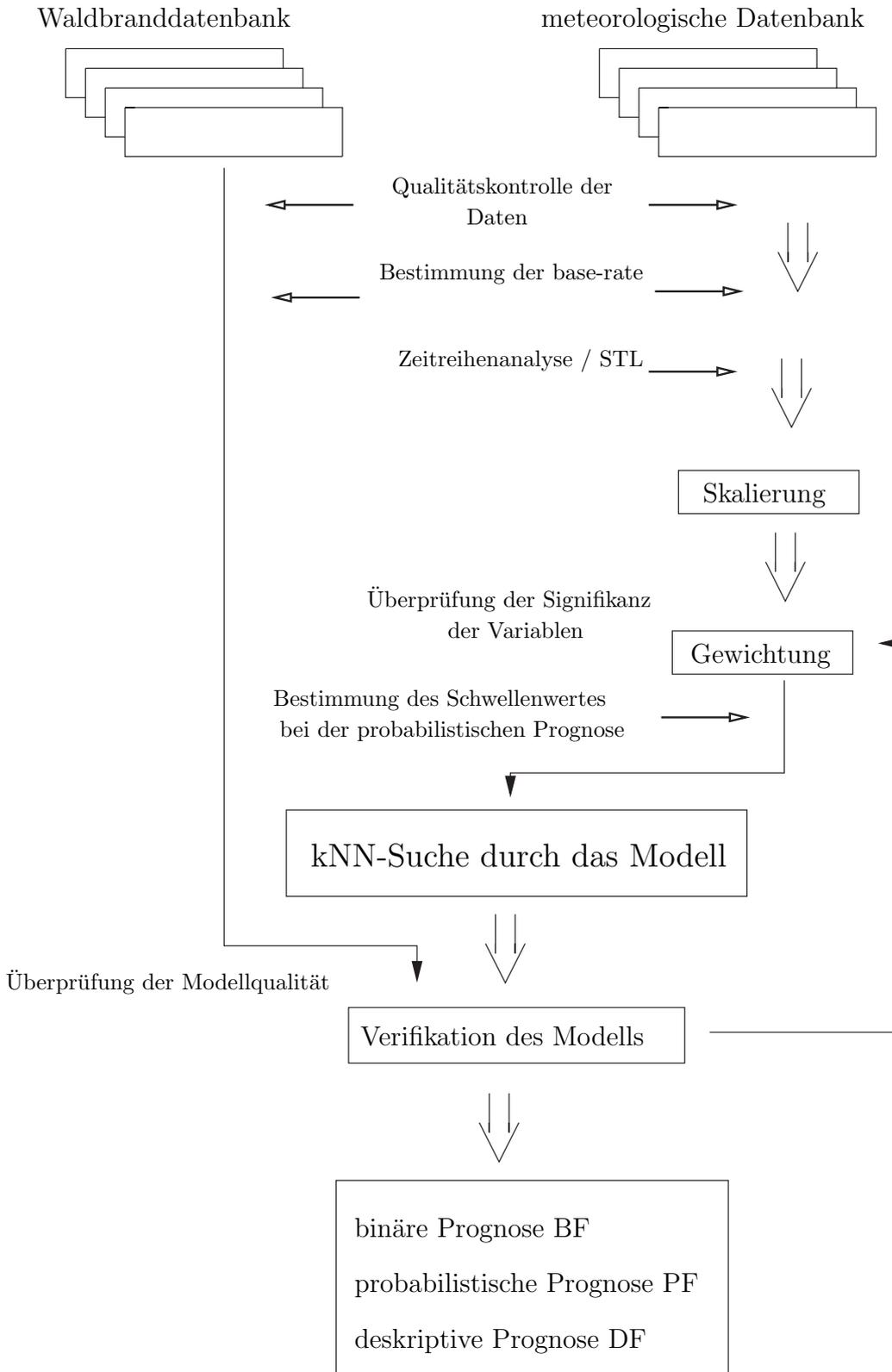


Abb. 4.4: Systematischer Ablauf des k NN-Modells zur Waldbrandprognose. (Überblick).

4. DIE k -NEAREST NEIGHBORS (k NN) METHODE

- (a) Nach der Aufbereitung der meteorologischen- und der Waldbranddatenbank müssen diese einer Qualitätskontrolle unterzogen werden. Dabei müssen die Werte auf ihre Plausibilität überprüft werden. Das k NN-Modell kann beispielsweise nicht mit fehlenden Variablen umgehen. Wenn also in der Datenbank meteorologische Messwerte fehlen, muss durch einen Vergleich mit anderen Messstationen überprüft werden, ob durch eine Interpolation die fehlenden Werte ergänzt werden können. Anderenfalls muss der entsprechende Beobachtungstag aus der Datenbank entfernt werden.
- (b) Die Grundwahrscheinlichkeit (*base-rate*) $s = \eta_{\text{tot}}^+ / n$ muss berechnet werden, wobei η_{tot}^+ die Anzahl Tage mit mindestens einem Ereignis und n die Gesamtanzahl Tage in der Datenbank darstellt. Ist s klein, wird die Klasse Ω^+ nur sehr schwer mittels einer geschickten Variablenwahl innerhalb von Ω gruppierbar. Gelingt diese Gruppierung auch bei einem sehr kleinem s gut, ist in der Regel genug Wissen über den Ereignisprozess vorhanden, um ein deterministisches Modell zu entwickeln.
- (c) Mittels einer Zeitreihenanalyse der Variablen muss untersucht werden, ob ein zeitlicher Trend in den Messdaten vorhanden ist. Obwohl dieser durch statistische Verfahren korrigiert werden könnte, kann damit belegt werden, dass sich beispielsweise die Vegetation verändert hat. Eine derartige Heterogenität in der Grunddisposition könnte das Modell nicht berücksichtigen. Es muss also die Homogenität der Grunddisposition überprüft werden. Ist diese nicht gewährleistet, funktioniert die k NN-Methode nicht.
- (d) Meteorologische Variablen, welche den Ereignisraum Ω für das k NN-Modell aufspannen, müssen definiert werden. Der Variablensatz muss so gewählt werden, dass sich die Klassifizierung der Ereignisse mittels der k NN-Methode möglichst optimal durchführen lässt. Dabei muss beachtet werden, dass nur solche Variablen mitberücksichtigt werden, die einen signifikanten Beitrag zur Klassifikation leisten, weil irrelevante Dimensionen von Ω die Qualität der k NN-Prognosen erheblich vermindern.
- (e) Die meteorologischen Variablen müssen skaliert werden. Die Skalierung ist nötig, damit bei der Nachbarsuche Einheiten und Zahlenwerte der Variablen kompatibel sind.

4. DIE k -NEAREST NEIGHBORS (k NN) METHODE

- (f) Um die Gruppierung der Klasse Ω^+ innerhalb von Ω zu optimieren, müssen die Variablen entsprechend ihrem Einfluss auf das Ereignis gewichtet werden. Diese Gewichtung beeinflusst das in Gleichung (4.1) definierte Mass $D(\mathbf{x}, \mathbf{y})$.
- (g) Mit den nun zur Verfügung stehenden skalierten und gewichteten Variablen wird die Nachbarsuche durchgeführt. Nun können mit dem k NN-Modell binäre, probabilistische sowie deskriptive Prognosen erstellt werden.
- (h) Die Qualität der k NN-Prognosen muss mit Hilfe von Verifikationsmethoden überprüft werden. Jeder einzelne der zuvor angeführten Schritte muss in der Folge anhand der Verifikationsresultate optimiert werden, bis die Ergebnisse zufriedenstellend sind und das k NN-Modell den in Kapitel 1.2 definierten Ansprüchen genügt.
- (i) Erfüllt die Qualität die gestellten Ansprüche, können die Ergebnisse interpretiert und zur Beurteilung der Waldbrandgefahr eingesetzt werden.

5 Das k NN-Waldbrand Prognosemodell

Der Kern des Prognosemodells wird durch die Nearest Neighbors (k NN) Methode selbst gebildet. Die in Abbildung 4.4 erwähnten Bedingungen und Abläufe bilden dabei die eigentlichen Grundlagen des k NN-Waldbrand Prognosemodells. In der Folge sollen diese, für die Modellbildung wichtigen Schritte, genauer betrachtet und diskutiert werden.

5.1 Stationsauswahl und Gebietseinteilung

Die Auswahl des Untersuchungsgebietes schränkte sich nach Berücksichtigung der Anzahl Waldbrände und deren Archivierung auf die Alpensüdseite (Kanton Tessin) ein. Da dieses Untersuchungsgebiet aber zu gross wäre, musste eine Aufteilung stattfinden. Basierend auf der Arbeit von Conedera et al. (1996) wurde die Alpensüdseite anhand folgender Kriterien in drei Untersuchungsgebiete aufgeteilt⁴:

- dominanter Vegetationstyp
- räumliches Niederschlagsmuster
- Charakteristika der relativen Luftfeuchte
- politische Gemeindegrenzen
- vorhandene Messstationen
- vollständige Zeitreihe der gemessenen meteorologischen Variablen

⁴Eine entsprechende Aufteilung erfordert, dass drei k NN-Modelle entwickelt werden. Die Kernstruktur bleibt dabei dieselbe, möglicherweise sind aber andere Variablen an dem Prozess einer Waldbrandentstehung beteiligt.

5. DAS k NN-WALDBRAND PROGNOSEMODELL

Eine Zuordnung der jeweiligen Gemeinden zu den entsprechenden meteorologischen Messstationen ist im Anhang B zu finden. Abbildung 5.1 gibt einen Überblick über die Standorte der für das k NN-Modell verwendeten meteorologischen Mess- und Phänologiestationen.

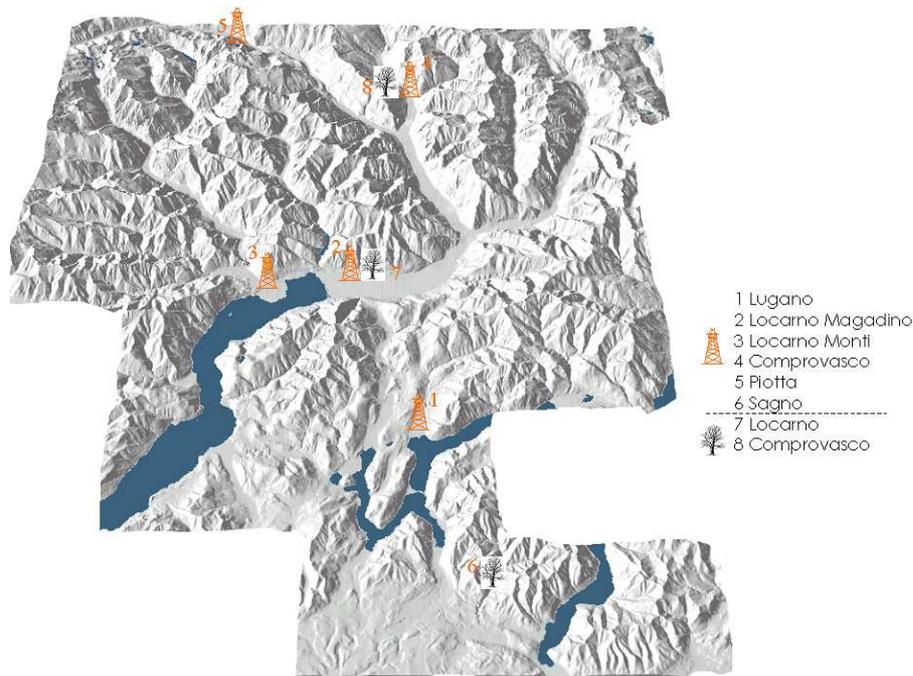


Abb. 5.1: Auf dieser Reliefkarte (hillshade) des Kantons Tessins sind die Standorte der für das k NN-Modell als Datengrundlage verwendeten meteorologischen Mess- und Phänologiestationen eingezeichnet. Quelle: Atlas der Schweiz.

Mit Hilfe räumlicher Autokorrelationsanalysen wurde zunächst versucht, eine computergestützte Stationsauswahl durchzuführen (Schneebeli & Laternser, 2004). Dabei wurden alle zur Verfügung stehenden Messstationen auf der Alpensüdseite berücksichtigt und deren Variablen jeweils einer Autokorrelationsanalyse unterzogen. Es sollte nur dann eine neue Station für das k NN-Modell berücksichtigt werden, wenn die Korrelation mehrerer Variablen zwischen zwei oder mehr Stationen niedrig ist. Im Zuge der Modelloptimierung wurde allerdings aufgrund unbefriedigender Ergebnisse der Autokorrelationsanalyse auf eine Gebietseinteilung mit Hilfe dieser computergestützten Methode verzichtet.

Im gesamten Untersuchungsgebiet standen nur fünf meteorologische Messstationen zur Verfügung, bei welchen alle nötigen Variablen gemessen werden. In

5. DAS k NN-WALDBRAND PROGNOSEMODELL

Tabelle 5.1 sind diese automatischen Beobachtungsstationen (ANETZ) der MeteoSchweiz aufgeführt. Anhand zuvor definierten Kriterien, ergibt sich nur für die nördlichste Region die Möglichkeit einer weiteren Unterteilung in Teiluntersuchungsgebiete. Diese wurde aufgrund der niedrigen Grundwahrscheinlichkeit s in dieser Region nicht ausgeführt. Es standen für diese Region die beiden Stationen Piotta oder Comprovasco als “Basisstation” für das k NN-Modell zur Auswahl. Modellberechnungen und anschliessend durchgeführte Verifikationen zeigten, dass mit der Station Piotta bessere Prognoseergebnisse erzielt werden konnten. Mit dem gleichen Vorgehen konnte für die Region mittleres Tessin Locarno Monti als Basisstation bestimmt werden.

Tab. 5.1: Übersicht der in Frage kommenden ANETZ-Stationen (MeteoSchweiz) für die Datengrundlage der k NN-Prognose.

Stationsname	Abk.	Lage-Koordinaten	Höhe (NN)	Daten seit
Piotta	PIO	694930/152500	1007	01.81
Lugano	LUG	717880/95870	273	01.81
Locarno Magadino	MAG	711160/113540	197	01.81
Locarno Monti	OTL	704160/114350	366	01.81
Comprovasco	COM	714998/146440	575	01.89

Gebirgsräume wie die Alpensüdseite sind bezüglich ihrer Klimatologie, heterogene Gebilde (Kap. 2.3). Es war daher schwierig eine repräsentative Messstelle für ein Gebiet zu finden, wie dies das k NN-Modell fordert. Je komplexer die Topographie, desto kleinräumiger sollten die Untersuchungsgebiete definiert werden. Dabei ist allerdings zu beachten, dass wie im vorliegenden Fall, ein dichtes Messnetz in diesen Gebieten häufig fehlt und eine kleinräumige Aufteilung daher nicht immer möglich war.

Für die k NN-Waldbrand Prognosemodelle der drei Untersuchungsgebiete, wurden als Grundlage die meteorologischen Daten der Stationen Lugano (südliches Tessin), Locarno Monti (mittleres Tessin) und Piotta (nördliches Tessin) verwendet.

5. DAS k NN-WALDBRAND PROGNOSEMODELL

Rohdaten bildeten 10-Minuten Werte, welche zu Stundenwerten gemittelt respektive summiert wurden. In einem weiteren Schritt wurden daraus Tagesmittel- respektive Tagessummenwerte gebildet.

Im Untersuchungszeitraum 01.01.1991 bis 31.12.2002 konnten 4344 Tage zur Modellentwicklung mitberücksichtigt werden. Die Waldbranddatenbank, welche durch die WSL (Eidg. Forschungsanstalt für Wald-, Schnee- und Landschaft) erhoben wird und jährlich aktualisiert als ORACLETM-Datenbank vorliegt, beinhaltet 38 Tage ohne gültigen Eintrag, sodass im Modell auf diese Tage verzichtet werden musste. Eine Zusammenfassung der Gebietseinteilung mit den entsprechenden repräsentativen meteorologischen Stationen und der Anzahl Waldbrände ist in Tabelle 5.2 zu finden.

Tab. 5.2: Gebietseinteilung der Untersuchungsgebiete und entsprechende meteorologische Stationen.

Gebiet	meteorol. Station	Anzahl Tage mit Bränden
südl. Tessin	Lugano	236
mittl. Tessin	Locarno Monti	227
nördl. Tessin	Piotta	141

Bei der Gebietsaufteilung, sind im Hinblick auf die Anwendung des k NN-Modells somit zwei massgebende Kriterien zu beachten. (a) die eben erwähnte Repräsentativität der einzelnen Messstation für das dazugehörige Untersuchungsgebiet und (b) die Grundwahrscheinlichkeit (*base-rate*) s der Waldbrände im gewählten Teilgebiet.

5.2 Grundwahrscheinlichkeit s (base rate)

Die Grundwahrscheinlichkeit s gibt an, wie häufig Ereignisse im Durchschnitt auftreten. Sie spielt im Zusammenhang mit der Komplexität jener Prozesse, welche zu einem Waldbrand führen, für die Qualität des Prognosemodells eine entscheidende Rolle. Viele Qualitätsmasse wie FAR , HSS oder CSI sind u.a. eine Funktion von s .

5. DAS k NN-WALDBRAND PROGNOSEMODELL

Zu kleine s -Werte führen bei komplexen Prozessen zu einer deutlichen Qualitätseinbusse. Dies konnte anhand von 24 Auswertungen von operationell in Betrieb stehenden NXD-Lawinenprognosemodellen gezeigt werden (*unpublizierte Arbeit: J. Heierli*).

Datengrundlage zum vorliegenden Beispiel bildeten 24 Auswertungen von k NN-Lawinenprognosemodellen aus verschiedenen Ländern (CH, A, GB (*R. Purves*)). Untersucht wurde jeweils der Zusammenhang zwischen verschiedenen Qualitätsmassen und der Grundwahrscheinlichkeit s . Anhand der Daten wurde ersichtlich, dass s -Werte unterhalb einer Schwelle von zirka 5%-7% zu schlechteren Modellergebnissen führten. Der Qualitätsverlust wirkt sich bei tieferer Grundwahrscheinlichkeit deutlich stärker aus, als bei höheren Werten für s . Die Ähnlichkeit der lawinen- und waldbrandauslösenden Prozesse, lässt denselben Schluss auch für die Naturgefahr Waldbrand zu.

Sind die zum Ereignis führenden Prozesse physikalisch gut beschreibbar, kann die Modellqualität trotz einer tiefen Grundwahrscheinlichkeit s gut sein. Anstelle eines statistischen Prognosemodells sollte aber in solchen Fällen vorzugsweise ein deterministisch-physikalisches Modell zur Modellierung des Prozesses verwendet werden.

Weiter ist zu berücksichtigen, dass die Grundwahrscheinlichkeit s sehr stark durch die Gebietsgrösse bestimmt wird. Je grösser das Untersuchungsgebiet ist, desto grösser wird die *base-rate*, weil bei gleicher Anzahl Tage eventuell die Anzahl der Brände steigt.

Untersuchungsgebiete dürfen somit nicht zu gross (Repräsentativität) aber auch nicht zu klein (*base-rate*) sein. Weiter sollte das Modell rasch eine Übersicht bezüglich der aktuellen Waldbrandsituation eines Gebietes liefern können.

Sollte für eine bestimmte Region respektive für eine neue thematische Fragestellung, wie beispielsweise Murgänge, ein neues k NN-Modell entwickelt werden, kann die Grundwahrscheinlichkeit s bereits als erster Indikator bezüglich einer zu erwartenden Modellqualität berücksichtigt werden, dies bevor die Ergebnisse der Modellverifikation vorliegen.

Für die Naturgefahr Waldbrand haben sich für den 12-jährigen Untersuchungszeitraum 1991-2002 auf der Alpensüdseite folgende Grundwahrscheinlichkeitswer-

5. DAS k NN-WALDBRAND PROGNOSEMODELL

te s ergeben:

- südliche Region mit der meteorologischen Station Lugano: $s = 5.4\%$
- mittlere Region mit der meteorologischen Station Locarno: $s = 5.2\%$
- nördliche Region mit der meteorologischen Station Piotta: $s = 3.2\%$

Base-rate Effekt

Die zuvor angesprochene Abhängigkeit der Modellqualität von der Grundwahrscheinlichkeit s soll anhand eines Beispiels aufgezeigt werden. Das Beispiel basiert auf den Resultaten des k NN-Modells der Region Lugano für den Zeitraum 1991-2002, wobei bei einer binären Prognose (BF) ein Schwellenwert $DB=3$ angenommen wurde.

Die Treffergenauigkeit (*percent correct*) PEC , welche ein Mass der korrekt prognostizierten Ereignisse resp. Nichtereignisse ist, beträgt beim k NN-Waldbrand Prognosemodell 87% (Vergleich: Treffergenauigkeit des MetOffice in GB bezüglich Niederschlag ca. 70-80% (Matthews, 1996a,b)). Dies bedeutet, dass mit Hilfe des Modells an 87 von 100 prognostizierten Tagen, eine korrekte Prognose erstellt werden konnte. a stellt dabei die korrekt prognostizierten Ereignistage, d die korrekt prognostizierten Nichtereignistage bei einer Gesamtanzahl untersuchter Tage n , dar.

$$PEC = \frac{a + d}{n} = \hat{p}(\hat{x} = 1) \quad (5.1)$$

Bedeutet dies nun im genannten Beispiel, dass im Falle eine Waldbrandprognose das Ereignis auch in 87% der Fälle auftritt?

Tatsächlich tritt ein Brand vielmehr nur in 30% der Fälle ein. Die Ursache dafür liegt in der niedrigen Grundwahrscheinlichkeit s des Ereignisses Waldbrand.

Von 100 zu prognostizierenden Tagen, kommt es durchschnittlich nur in 6 Fällen zu einem Brand, an 94 Tagen wird kein Ereignis eintreten. Die Treffergenauigkeit bezieht sich nun aber nur auf diese 6 Fälle, d.h. 5 von 6 werden im Schnitt richtig prognostiziert. Weiterhin weisen die übrigen 94 Nichtwaldbrandtage eine Fehlprognose von 12% auf. Es wird also insgesamt in 5 von 17 Fällen eine richtige Prognose gemacht, was bedeutet, dass nur in 29% der Fälle, in welchen

ein Feuer prognostiziert wird, tatsächlich auch eines auftritt. Dies entspricht ungefähr der Grössenordnung wie sie auch bei Niederschlagsprognose von MetOffice in England anzutreffen ist (Matthews, 1996b).

Anhand der Gleichung 5.2 kann bestimmt werden, wie gross die Grundwahrscheinlichkeit s mindestens sein muss, damit über 50% richtig prognostiziert wird.

$$s(\text{min}) = 1 - PEC \quad (5.2)$$

Im vorliegenden Beispiel müsste s somit 13% sein.

5.3 Homogenität der Messreihen

Nach der Gebietseinteilung und Stationsauswahl mussten die Messdaten einer Qualitätskontrolle unterzogen werden. Dabei mussten die Daten in einem ersten Schritt auf allfällige Lücken (fehlende Messwerte), in einem zweiten Schritt auf ihre Plausibilität hin überprüft werden.

Für das k NN-Modell ist wichtig, dass die meteorologischen Messgrössen keinem Trend unterliegen. Die Erfüllung dieser Bedingung wurde mit Hilfe einer Zeitreihenanalyse untersucht. Mögliche Trends beispielsweise in der Lufttemperaturmessreihe oder im Niederschlagsregime können negative Auswirkungen auf die Nachbarnsuche haben oder diese gar verunmöglichen. Eine gleichmässige Verteilung der einzelnen Werte wäre bei Vorliegen eines Trends nicht mehr gegeben, weil vermehrt Niederschlagsereignisse auftreten, welche es in einer vorangegangenen Periode mit durchschnittlich weniger Niederschlag nicht gegeben hatte. Dies hätte zur Folge, dass solche Ereignisse keine oder nur wenige Nachbarn haben und somit keine Aussagen möglich sind. Abbildung 5.2 zeigt an einem fiktiven Beispiel, wie sich ein Trend in den Messdaten auf die Nachbarnsuche auswirken würde. In den Diagrammen sind die beiden Messgrössen relative Luftfeuchte und 30-Tages Niederschlagssumme für die Perioden 1 und 2 aufgetragen. Beim Vergleich der beiden Graphiken kann ein Trend zu einer niedrigeren Luftfeuchte bei gleicher 30-Tages-Niederschlagssumme bzw. erhöhte Niederschlagsbereitschaft festgestellt werden. Die einzelnen Punkte stellen die Tage im vorgegeben Zeitraum dar. Angenommen man sucht in der Box I des rechten Diagramms nach den nächsten Nachbarn, findet man nur deren drei; sucht man demgegenüber in der Box I des

5. DAS k NN-WALDBRAND PROGNOSEMODELL

ersten Diagramms, ist eine viel grössere Anzahl Fälle zu finden. Gut ersichtlich wird dieser Umstand auch in der Box X.

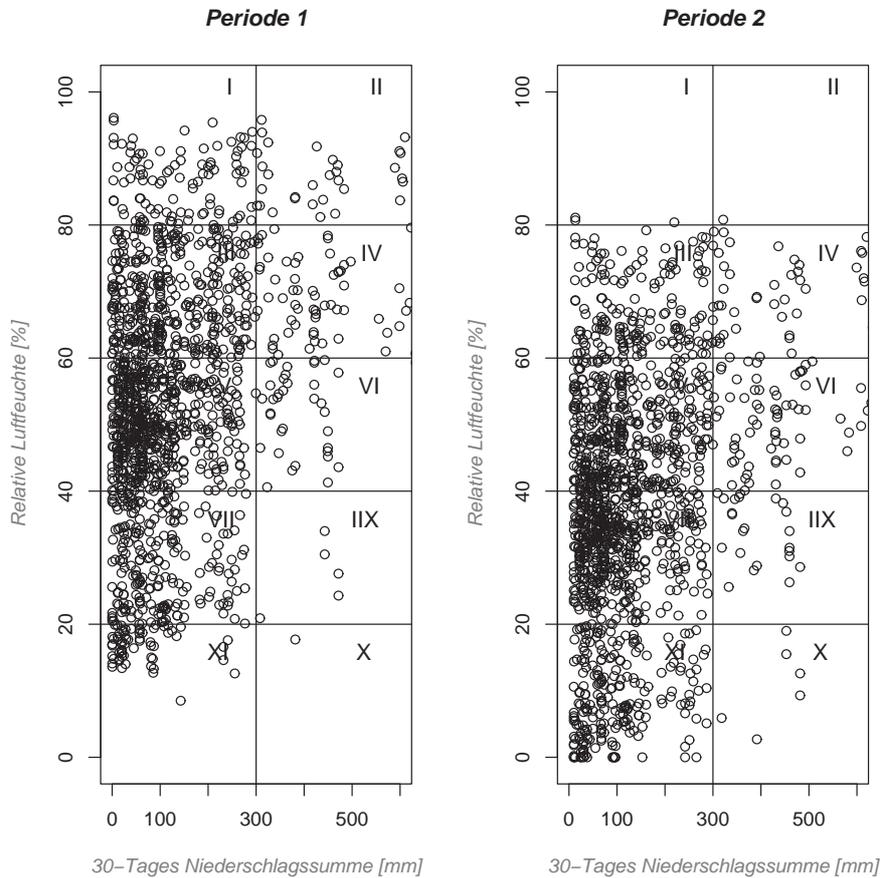


Abb. 5.2: Fiktives Beispiel um den Einfluss eines Trends zu verdeutlichen. Unterliegen die meteorologischen Variablen einem Trend, so verändert sich die Datendichte, dies vor allem in den “extremen” Bereichen.

Mit Hilfe dieser Abbildung kann auch die Problematik von seltenen Ereignissen erklärt werden. Wird die Nachbarsuche über die Anzahl der nächsten Nachbarn (z.B. $k=10$) bestimmt, werden immer zehn ähnliche Fälle gefunden. Man muss sich allerdings die Frage stellen, wie ähnlich diese Nachbarn effektiv sind. Anstelle die Nachbarsuche mit einem fixen k durchzuführen, könnte man daher auch über eine im voraus festgelegte Distanz D suchen. Die Anzahl der Nachbarn wäre variabel und würde ein Mass für die Dichte der Punkte \mathbf{x} um den Prognosetag \mathbf{y} darstellen.

Vor der eigentlichen Analyse von Zeitreihen ist es wichtig, zunächst eine graphische Voranalyse durchzuführen. Ausreisser und allenfalls fehlende Messwerte können so rasch erkannt werden. Fehlende Messwerte sollten in Abhängigkeit des zu untersuchenden Prozesses physikalisch plausibel geschätzt werden. Messwerte, welche sich offensichtlich ausserhalb physikalisch realistischer Wertebereiche befinden, sind als Ausreisser zu betrachten und durch eine unvoreingenommene Schätzung, welche die Dynamik des untersuchten Prozesses berücksichtigt, zu ersetzen.

Eventuelle Nichtstationaritäten sind in Trend (i), unkonstante Varianzen und deterministisch periodische Schwankungen (ii), sowie Monats- oder Jahreszeiteffekte (iii), aufgeteilt. Bei Hürzeler (2003) sind Vorschläge zum Testen der Stationarität zu finden. Da die verschiedenen Tests in der Regel eine kleine Macht haben, wurde für die vorliegende Arbeit eine visuelle Kontrolle, die *seasonal trend decomposition analysis* (STL), angewendet.

Weisen einzelne, für das k NN-Modell zu verwendende Messdaten einen zu starken Trend auf, muss der Einsatz des k NN-Modells in Frage gestellt werden.

Das grösste Problem sind in diesem Zusammenhang Verschiebungen in der Grunddisposition, welche z.B. durch eine Klimaänderung (Trend) verursacht werden. Für ein Waldbrandprognosemodell würde beispielsweise eine Veränderung der Vegetation einen solchen unzulässigen Trend darstellen. Waren zuvor an einem bestimmten Standort beispielsweise Buchenwälder anzutreffen, und sind dort später Kastanienwälder zu finden, hat sich die Grunddisposition derart verändert, dass keine Prognose mehr anhand der Gesamtheit von Messdaten gemacht werden können. Das Problem bei der Behandlung einer Klimaveränderung liegt somit in der Veränderung der Grunddisposition respektive des Feuerregimes und nicht in der Datendichte. Daher sollte auf jeden Fall mit dem zur Verfügung stehenden Datensatz eine Zeitreihenanalyse (STL) durchgeführt werden.

STL: A Seasonal-Trend Decomposition Procedure

Die STL zerlegt eine Zeitreihe nichtparametrisch in eine Trend- und eine Saisonkomponente sowie in einen irregulären Term. In mehreren Schritten wird jeweils unter Verwendung eines nichtparametrischen Glättungsalgorithmus "*loess*" abwechslungsweise die Trend- respektive die Saisonkomponente geglättet. Der grosse Vorteil dieser nichtparametrischen Methode gegenüber einer Modellierung ist,

5. DAS k NN-WALDBRAND PROGNOSEMODELL

dass in einfacher Weise Änderungen der Monatseffekte und allfällig kompliziertere Formen des Trends graphisch sichtbar gemacht werden können (Cleveland et al., 1990). Im vorliegenden Fall war von Interesse, ob in den letzten 12 Jahren allenfalls ein Trend in den verschiedenen meteorologischen Parametern zu finden war.

Der Umgang mit Ausreißern in Zeitreihen ist nicht trivial, da ansonsten die äquidistante Zeitstruktur zerbricht. Allfällige Ausreißer können durch plausibel erscheinende Werte ersetzt werden. Die Abbildungen 5.3 und 5.4 dienen dazu, festzustellen, ob ein allfälliger Trend in den einzelnen Parametern vorlag oder nicht. Die Abbildungen zeigen, dass in den beiden Parametern “Lufttemperatur” und “relative Luftfeuchte” kein optisch erkennbarer Trend feststellbar war. Derartige Trendanalysen wurden mit den gleichen Ergebnissen für alle Parameter aller Stationen durchgeführt. Für die weiteren Untersuchungen kann somit ein Trend und folglich eine mögliche Unsicherheit im Modell ausgeschlossen werden. Bei der Betrachtung eines längeren Zeitraumes (1980-2003) hingegen konnte u.a. bei der Lufttemperatur mit der Trendanalyse eine leichte Zunahme festgestellt werden.

Im Gegensatz zu Expertensystemen oder deterministischen Modellen, stellt diese Forderung einer Homogenität der Daten bei k NN-Modellen einen methodischen Nachteil dar.

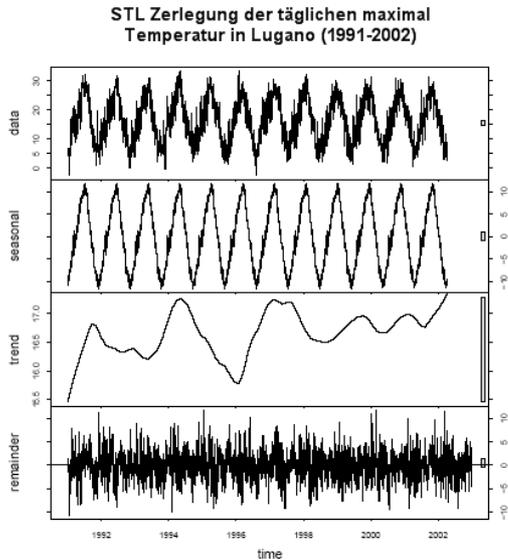


Abb. 5.3: STL-Plot für die maximale Lufttemperatur der Station Lugano 1991-2002. Ganz schwach können Tendenzen eines Trends gesehen werden. Diese spielen allerdings in Bezug auf das k NN-Modell bei diesen Größenordnungen keine Rolle.

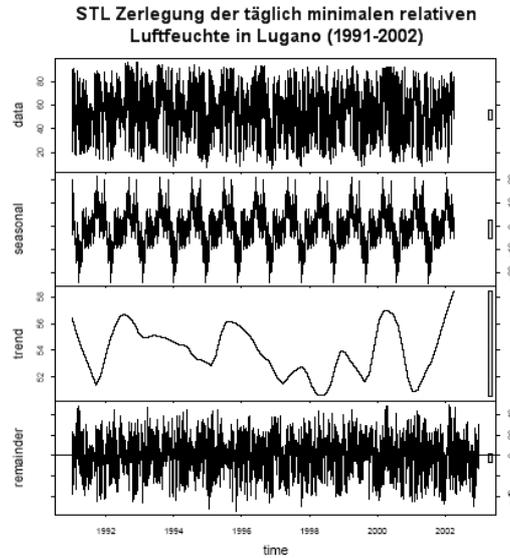


Abb. 5.4: STL-Plot für die minimale relative Luftfeuchte der Station Lugano 1991-2002.

5.4 Variablenwahl

Nichtsignifikante Variablen erhöhen die Dimension des Ereignisraumes Ω und mindern damit die Qualität des Modells (Kap. 4.4). Mit der zunehmenden Anzahl an Dimensionen steigt die Fehlerrate der Klassifikation. Dieser Umstand ist unter dem Begriff *Peaking*-Phänomen oder “Fluch der Dimensionalität” bekannt (Bellmann, 1961).

Eine Erhöhung der Anzahl Dimensionen p durch die Berücksichtigung einer zusätzlichen Variablen führt in lokaler Umgebung um einen betrachteten Punkt \mathbf{x} zu einer Abnahme der Datendichte. Da sich mehr Tage ohne Brände in der Datenbank befinden als solche mit, führt diese Verdünnung zu einer Varianzerhöhung. Eine Ausführliche Diskussion zu dieser Thematik ist in Friedman (1993) zu finden. Um diesen Qualitätsverlust im k NN-Modell zu verhindern, wird vorgängig eine Variablenwahl durchgeführt, indem nur effektiv zur Lösung der Klassifikation beitragende Variablen berücksichtigt werden.

5.4.1 Empirische Bestimmung der Variablen

In einem ersten Schritt wurde eine einfache “Vorselektion” der Variablen durchgeführt. Dazu wurden empirisch alle Variablen, die mit dem zu untersuchenden Phänomen keinen bekannten Zusammenhang haben, eliminiert. Die Anzahl der von einer ANETZ-Station zur Verfügung stehenden Variablen reduzierte sich somit um etwa 70%. Tabelle 5.3 stellt die für das k NN-Modell zur Verfügung stehenden gemessenen Variablen nach dieser ersten empirischen Selektion dar.

Tab. 5.3: Gemessene Variablen der ANETZ-Stationen und nach einer ersten empirisch durchgeführten Variablenreduktion dem k NN-Modell zur Verfügung stehenden Grössen.

Parameter	Einheit	Standard
Lufttemperatur	[°C]	Maximum
relative Luftfeuchte	[%]	Minimum
Luftdruck	[mb]	Mittel
Niederschlag	[mm]	Summe
Globalstrahlung	[Wm ⁻²]	Summe
Windspitze	[ms ⁻¹]	Maximum
Sonnenscheindauer	[min]	Summe
Verdunstung	[mm]	Summe
Fernblitze		Summe

Berücksichtigung der Phänologie

Da sich sowohl die Anzahl der Feuer, als auch deren Ausmass im Sommer- und Winterhalbjahr aufgrund des unterschiedlichen Feuerregimes deutlich unterscheiden, musste dieser Umstand im k NN-Modell mitberücksichtigt werden. Die Aufteilung des Jahres mittels der Variable Datum ist dazu nicht sinnvoll, denn Brände treten völlig unabhängig von dieser Grösse auf. Vielmehr sollte eine Variable gefunden werden, welche die Eigenschaften der Jahreszeiten in Bezug auf Waldbrände widerspiegelt. In einem nächsten Schritt wurde zu diesem Zweck die Phänologie mitberücksichtigt.

Dem Zeitraum zwischen Blattentfaltung und Blattfall respektive umgekehrt wurde ein zweiwertiger Zahlenwert zugeordnet. Zu jedem Blattentfaltungstermin

5. DAS k NN-WALDBRAND PROGNOSEMODELL

wurden zusätzlich 10 Tage hinzugezählt, da sich die Situation des Brandgutes ändert, sobald ein mehr oder weniger vollständiges Blätterdach vorhanden ist. Einerseits nimmt die Menge an potentiell Brandgut (z.B. Laub) ab, andererseits ändern sich die Strahlungsbedingungen und somit die Austrocknung des Brandgutes unterhalb des Blätterdaches massgeblich. Die Datengrundlage zur Berücksichtigung der Phänologie lieferten Daten des phänologischen Beobachtungsnetzes von MeteoSchweiz in Sagno, Locarno und Comprovasco (Tab. 5.4). Als Zeigerpflanze diente die auf der Alpensüdseite weit verbreitete und von Bränden stark betroffene Edelkastanie. Fehlende Edelkastanienwerte wurden durch Extrapolationen mit Hilfe von Phänologiedaten der Rosskastanie berechnet. Abbildung 5.5 zeigt, dass im Beobachtungszeitraum kein signifikanter zeitlicher Trend zu einem früheren respektive späteren Blattaustrieb bzw. -fall feststellbar ist. Die Schwankungen von Jahr zu Jahr können dennoch erheblich sein.

Tab. 5.4: Übersicht über die im k NN-Modell zu berücksichtigenden Phänologiedaten und -stationen.

Stationsname	Höhe (NN)	Pflanzentyp	Daten seit
Locarno	200	Edelkastanie	01.89
Sagno	670	Edelkastanie	01.87
Comprovasco	500	Edelkastanie	01.89

5. DAS k NN-WALDBRAND PROGNOSEMODELL

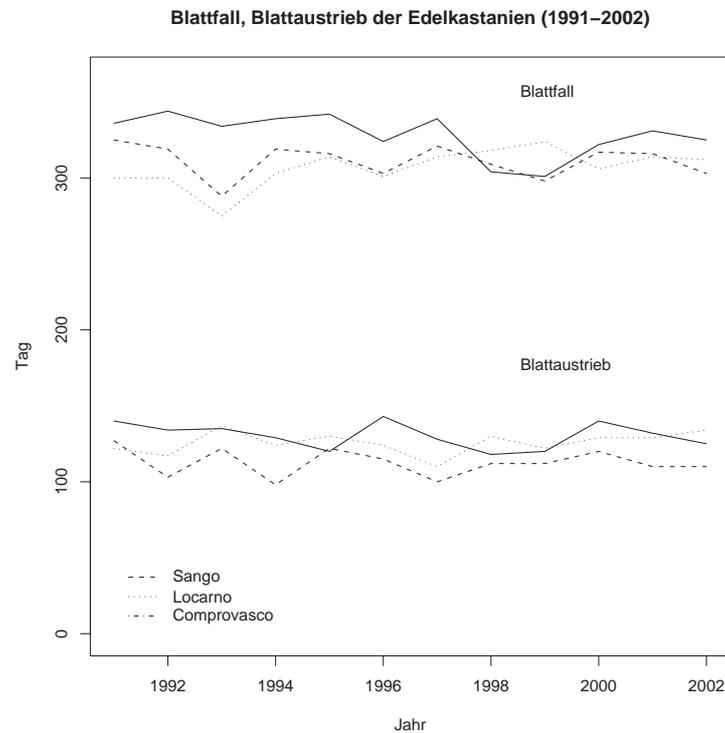


Abb. 5.5: Datum des jährlichen Blattaustriebs respektive Blattfalls bei den im k NN-Modell berücksichtigten Phänologiestationen. Es wird ersichtlich, dass im Untersuchungszeitraum kein Trend feststellbar ist, die jährlichen Schwankungen allerdings gross sein können.

5.4.2 Bestimmung der Variablen mittels logistischer Regression

Nach der in Kapitel 5.4.1 beschriebenen ersten Reduktion der Variablen, wurde die Signifikanz und somit der Einfluss der übrig gebliebenen Variablen auf das Ereignis Waldbrand mit Hilfe der logistischen Regression bestimmt. Dadurch sollte die Anzahl der zur Verfügung stehenden Variablen (Tab. 5.3) für das k NN-Modell des jeweiligen Untersuchungsgebietes wenn möglich ein weiteres mal reduziert werden.

Mit dem logistischen Regressionsmodell (LR) konnte der unmittelbare Einfluss von erklärenden Variablen auf eine binäre Zielgrösse wie Waldbrand bestimmt werden. Für die einzelnen Variablen kann ein standardisiertes Mass (P -Wert) dafür berechnet werden, wie gut die Variablen mit dem Modell der Nullhypothese übereinstimmen. Ist der P -Wert klein, weichen die Daten signifikant von der

5. DAS k NN-WALDBRAND PROGNOSEMODELL

Nullhypothese ab und der Einfluss dieser erklärenden Variable ist somit statistisch gesichert (Stahel, 2002). Zwecks einfacheren Vergleiches der P -Werte werden diese in Abhängigkeit ihrer Signifikanz mit einem * versehen (Tab. 5.5).

Tab. 5.5: Zusammenfassung der Signifikanzsymbole (Schönwiese, 1991).

P-Wert	Signifikanz	*
$p > 0.05$	nicht signifikant	
$0.01 < p \leq 0.05$	signifikant	*
$0.001 < p \leq 0.01$	sehr signifikant	**
$p < 0.001$	höchst signifikant	***

Die Herleitung der Methode und die auszuführenden Schritte bei der Variablengewichtung mittels logistischer Regression sind im Anhang A zu finden. Dort wird im weiteren, ein vollständiges parametrisches Modell zur Prognose der Waldbrandgefahr entwickelt.

Die Tabellen 5.6 bis 5.9 geben die Signifikanztest zu den einzelnen Variablen wieder. Die Tabelle 5.6 zeigt die Werte des noch nicht reduzierten logistischen Regressionsmodells an. Für die Stationen Locarno und Piotta (vgl. Tab. 5.7 und 5.9) werden nur die im k NN-Modell mitberücksichtigten Variablen angegeben.

5. DAS k NN-WALDBRAND PROGNOSEMODELL

Tab. 5.6: Signifikanztest der Variablen des vollen Modells mit Hilfe der logistischen Regression für die Messstation Lugano (1991-2002).

Koeffizienten	Estimate	Std. Error	z value	Pr(> z)	Sign.
(Intercept)	-4.22953	0.18069	-23.408	< 2e-16	***
Temperatur	0.43933	0.15297	2.872	0.00408	**
Relative Feuchte	-0.74462	0.14570	-5.110	3.21e-07	***
Druck	0.18416	0.08397	2.193	0.02828	*
Niederschlag	-0.31369	0.41616	-0.754	0.45099	
Globalstrahlung	0.01206	0.20620	0.058	0.95337	
Windspitze	-0.05167	0.22585	-0.229	0.81903	
Sonnenscheindauer	0.08666	0.17998	0.481	0.63019	
Summe Blitze	-0.17052	0.19345	-0.881	0.37806	
Windmittel	0.06558	0.21669	0.303	0.76216	
Summe Niederschlag	-1.31405	0.16845	-7.801	6.16e-15	***
Phänologie	0.60003	0.12104	4.957	7.15e-07	***
Nulldevianz: 1838.1 mit 4382 Freiheitsgraden. Residuendevianz: 1398.9 mit 4371 Freiheitsgraden. AIC: 1422.9					

Tab. 5.7: Signifikanztest der Variablen des reduzierten Modells mit Hilfe der logistischen Regression für die Messstation Lugano (1991-2002).

Koeffizienten	Estimate	Std. Error	z value	Pr(> z)	Sign.
(Intercept)	-4.13462	0.14962	-27.634	< 2e-16	***
Temperatur	0.47515	0.10311	4.608	4.06e-06	***
Relative Feuchte	-0.83023	0.09142	-9.082	< 2e-16	***
Druck	0.21930	0.07416	2.957	0.00310	**
Summe Niederschlag	-1.321796	0.16209	-8.155	3.51e-16	***
Phänologie	0.61063	0.11832	5.161	2.46e-07	***
Nulldevianz: 1838.1 mit 4382 Freiheitsgraden. Residuendevianz: 1402.5 mit 4377 Freiheitsgraden. AIC: 1414.5					

5. DAS k NN-WALDBRAND PROGNOSEMODELL

Tab. 5.8: Signifikanztest der Variablen des reduzierten Modells mit Hilfe der logistischen Regression für die Messstation Locarno (1991-2002).

Koeffizienten	Estimate	Std. Error	z value	Pr(> z)	Sign.
(Intercept)	-3.41910	0.10070	-33.952	< 2e-16	***
Temperatur	0.48643	0.11525	4.221	2.44e-05	***
Relative Feuchte	-0.45853	0.10852	-4.225	2.39e-05	***
Druck	0.21676	0.0827	2.620	0.008786	**
Windspitze	0.26577	0.07510	3.539	0.000401	***
Blitze	0.10413	0.05171	2.013	0.044063	*
Summe Niederschlag	-0.63514	0.12674	-5.012	5.40e-07	**
Phänologie	0.45195	0.11145	4.055	5.01e-05	***
Nulldevianz: 1786.1 mit 4382 Freiheitsgraden. Residuendevianz: 1590.2 mit 4375 Freiheitsgraden. AIC: 1606.2					

Tab. 5.9: Signifikanztest der Variablen des reduzierten Modells mit Hilfe der logistischen Regression für die Messstation Piotta (1991-2002).

Koeffizienten	Estimate	Std. Error	z value	Pr(> z)	Sign.
(Intercept)	-4.12581	0.14636	-28.190	< 2e-16	***
Temperatur	0.43114	0.09950	4.333	1.47e-05	***
Relative Feuchte	-0.71310	0.12245	-5.824	5.76e-09	***
Druck	0.30742	0.11024	2.789	0.005292	**
Blitze	0.17935	0.04902	3.659	0.000254	***
Summe Niederschlag	-0.90213	0.17502	-5.155	2.54e-07	***
Nulldevianz: 1246.6 mit 4382 Freiheitsgraden. Residuendevianz: 1093.1 mit 4377 Freiheitsgraden. AIC: 1105.1.					

Bei der Betrachtung der Signifikanztests fällt auf, dass beispielsweise die Variable *Niederschlag* in keinem der drei Regressionsmodelle signifikant ist. Dies kann damit erklärt werden, dass ein Niederschlagsereignis am Brandtag selbst keinen grossen Einfluss hat. Durch die gegebene zeitliche Auflösung (Tag) ist es sogar möglich, dass der Waldbrand vor dem Niederschlag aufgetreten ist.

Weiter kann festgestellt werden, dass für die einzelnen Regionen unterschiedliche Variablen signifikant sind. Die Phänologie beispielsweise, die für die Regionen Lugano und Locarno hohe Signifikanzwerte aufweist, fällt bei der Variablenreduktion der Region Piotta weg. Dies könnte damit zu tun haben, dass die Phänolo-

5. DAS k NN-WALDBRAND PROGNOSEMODELL

giestation für das k NN-Modell dieser Regionen ihren Standort relativ weit südlich und tief gelegen in der Leventina (Comprovasco) hat und somit nicht ausreichend repräsentativ ist.

Auf den ersten Blick erscheint der Umstand, dass die Variable *Druck* in allen drei Regionen signifikant ist, erstaunlich. Dies kann jedoch damit erklärt werden, dass Brände häufig bei stabilen Hochdrucklagen ausbrechen.

In der Spalte “*Estimate*” der Tabellen 5.7 bis 5.9 sind die β -Werte der Gleichung A.2 der logistischen Regression aufgeführt. Anhand der Vorzeichen dieser Werte, im Rahmen der logistischen Regression Schätzer oder Gewichte genannt, kann aufgrund der negativen Vorzeichen erkannt werden, dass mit kleiner werdenden Variablenwerten deren Einfluss steigt (indirekt proportional). Positive Vorzeichen zeigen abnehmenden Einfluss bei kleiner werdenden Variablenwerten an (direkt proportional).

5.4.3 Skalierung der Variablen

Die für das k NN-Modell ausgewählten Variablen müssen skaliert werden, damit bei der Nachbarsuche die Einheiten und Zahlenwerte der Variablen keinen Einfluss auf das Modell ausüben. Die Skalierung der einzelnen Variablen wird mit Hilfe der in Gleichung 5.3 aufgeführten Skalierungsgleichung durchgeführt. Die Skalierung wird so vorgenommen, dass skalierten Variablen die Standardabweichung 1 erhalten.

$$\hat{x}_i = \frac{x_i - \bar{x}}{\sqrt{\frac{1}{n-1} \sum_i (x_i - \bar{x})^2}} = \frac{x_i - \bar{x}}{sd(x)} \quad (5.3)$$

Damit ist gewährleistet, dass die Abstände zwischen den einzelnen Tagen \mathbf{x} nicht aufgrund unterschiedlicher Variableneinheiten, sondern aufgrund tatsächlicher Unterschiede innerhalb der Variablen definiert sind.

5.5 Variablengewichtung mittels Genetischen Algorithmen (GA)

Nachdem mittels der logistischen Regression die für die Waldbrandprognose signifikanten p Variablen bestimmt worden sind, mussten diese entsprechend ihrem Einfluss gewichtet werden. Dazu standen verschiedene, in Felber & Bartelt (2002) diskutierte, Ansätze zur Verfügung.

Die Variablengewichtung kann mittels Korrelationsmethode (i), logistischer Regression (ii) oder genetischen Algorithmen (GA) (iii) durchgeführt werden. Felber & Bartelt (2002) stellten fest, dass sich die Qualitätsmasse der Modelle nach den drei unterschiedlichen Verfahren gewichteten Variablen nur geringfügig unterscheiden. Das mit der Methode der genetischen Algorithmen (GA) berechnete Modell wies dabei die höchste Prognosequalität auf. Die Auswertungen des k NN-Modells wurden im Rahmen der vorliegenden Arbeit mit GA-gewichteten Variablen berechnet.

Genetische Algorithmen (GA) zählen wie die Diskriminanzanalyse zur Gruppe der stochastischen Optimierungsmethoden. Vorbild bei der Suche nach Optima einer Zielfunktion ist der Evolutionsprozess der Natur. Die Anpassung von Lebewesen an die Erfordernisse ihrer Umwelt kann als Optimierungsprozess interpretiert werden.

Die Theorie der GA wurde von John Holland in den 70er Jahren an der Universität von Michigan entwickelt (Holland, 1975). Holland beschäftigte sich in seiner Forschungsarbeit mit theoretischen Aspekten der Weitergabe genetischer Information von Generation zu Generation. Zeitgleich mit Hollands Untersuchungen wurden an der Technischen Universität Berlin die sogenannten Evolutionsstrategien definiert (Rechenberg, 1973). Beide Methoden wurden später unter dem Oberbegriff der evolutionären Algorithmen zusammengefasst.

Eine Einführung in die Theorie der genetischen Algorithmen sowie in deren praktische Anwendungen gibt Goldberg (1989).

5.5.1 Methode

Um den Ablauf der Optimierung zu beschreiben werden die Begriffe der Genetik respektive der Evolutionsbiologie eingeführt.

In jeder Zelle eines Lebewesens ist der gesamte Bauplan des Organismus enthalten. Im Leben des Organismus bleibt dieser Bauplan im wesentlichen unverändert. Betrachtet man hingegen die Gesamtheit aller Organismen einer Gattung, so verändern sich die Baupläne im Laufe der Zeit.

Neue Baupläne entstehen bei der Fortpflanzung von Organismen (Kreuzung). Darüberhinaus können während der Fortpflanzung spontane Änderungen, sogenannte Mutationen auftreten. Die Überlebensfähigkeit eines Lebewesens richtet sich nun danach, wie gut es mit den vorherrschenden Umweltbedingungen zurecht

5. DAS k NN-WALDBRAND PROGNOSEMODELL

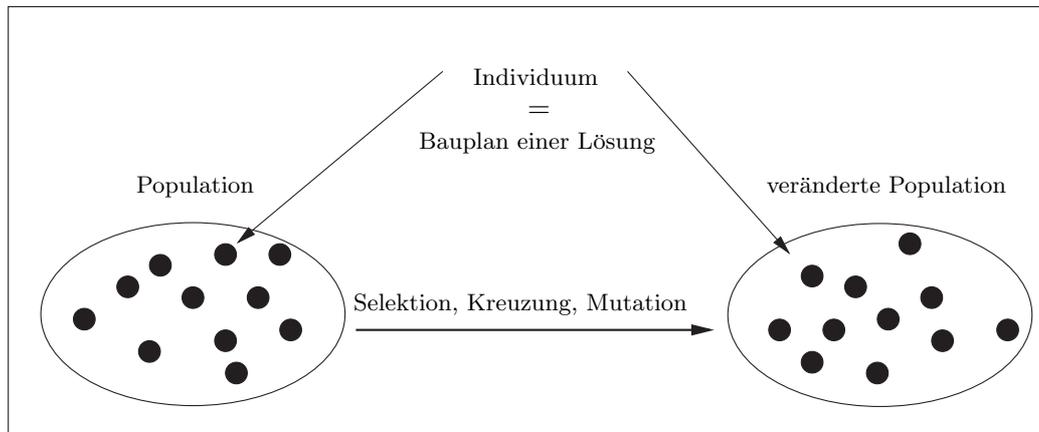


Abb. 5.6: Schema eines genetischen Algorithmus.

kommt. Die am besten Angepassten werden aufgrund der natürlichen Selektion überleben.

Das Modell arbeitet mit einer bestimmten Menge von Bauplänen, die mögliche Lösungen des Problems darstellen. Diese Menge wird in Anlehnung an die Biologie, *Population* genannt. Ein Element der Menge, bezeichnet man als *Individuum*.

- Die natürliche Selektion wird simuliert, indem die schlechtesten Baupläne aus der Population entfernt werden.
- Die Kreuzung wird so imitiert, dass aus je zwei Bauplänen (d.h. Lösungen), neue Baupläne zusammengesetzt werden.
- Mutationen werden durch punktuelle Änderungen von Bauplänen nachgestellt.

Eine Menge von Lösungen wird durch Selektion, Kreuzung und Mutation kontinuierlich verändert, mit dem Ziel einer Optimierung der Lösungen.

Auf das Variablengewichtungsproblem umgesetzt bedeutet dies, dass ein Individuum festlegt, ob eine gewichtete Variable einen grossen Einfluss zur Entstehung eines Waldbrandes hat oder nicht. Eine Population wird nun also als Array von Individuen realisiert (Abb. 5.5.1).

Der Ablauf eines genetischen Algorithmuses (GA) lässt sich in fünf Schritten zusammenfassen:

- **Ausgangspopulation:** Am Anfang wird zufällig eine Population, also eine Menge von Bauplänen erzeugt. Mit den Individuen dieser Population beginnt der Evolutionsprozess. Im Falle der vorliegenden Untersuchung stellen zufällig gewichtete Variablen die Ausgangspopulation dar.
- **Selektion:** In diesem Schritt wird eine natürliche Selektion nachgebildet. Die überlebensfähigsten Lösungen bestehen weiter. Die natürliche Selektion wird in zwei Schritten nachgebildet. Im ersten Schritt wird die Überlebensfähigkeit (*fitness*) der einzelnen Individuen bestimmt. Im zweiten Schritt erfolgt dann die eigentliche Selektion. Je höher die *Fitness* eines Individuums ist, desto grösser ist die Wahrscheinlichkeit, dass es überlebt. Dabei können verschiedene Selektionsverfahren angewendet werden. Für tiefergreifende Einzelheiten dazu wird auf Nissen (1997) verwiesen. Die einzelnen Variablen werden unterschiedlich gewichtet und dabei überleben nur jene Gewichte, bei welchen die *Fitness* höher ist.
- **Kreuzung:** Indem Baupläne miteinander kombiniert werden, sollen neue, bessere Lösungen generiert werden. Auch hier gibt es eine Vielzahl von Kreuzungsvarianten, welche in Nissen (1997) näher erläutert werden. An dieser Stelle werden die nach dem vorhergehenden Prozess gewichteten Variablen in Kombination mit anderen gebracht.
- **Mutation:** Die Mutation verfolgt das Ziel, gänzlich andere Lösungen zu finden. Sie gewährleistet, dass die Vielfalt in der Population erhalten bleibt. An dieser Stelle werden einzelne Gewichte wieder zufällig verändert. Mit Hilfe der *Fitness* wird dann überprüft, ob dies nicht eine bessere Kombination ergibt.
- **Nächste Generation:** Nachdem Selektion, Kreuzung und Mutation durchgeführt worden sind, ist eine neue, veränderte Population entstanden. Im vorliegenden Fall werden die Gewichte nach jeweils 60-80 GA-Modelldurchläufen ausgegeben.

Der Nachteil der Methode liegt darin, dass ein relatives Optimum gefunden wird. Das hat zur Folge, dass eine Gewichtung mit demselben Datensatzes ein zwei-

5. DAS k NN-WALDBRAND PROGNOSEMODELL

tes Mal nicht zu denselben Gewichten führen wird. Auf die k NN-Modellqualität hat dies, wie anhand mehrerer Experimente gezeigt werden konnte, letztendlich jedoch keinen Einfluss (Felber & Bartelt, 2002).

5.6 Erstellung einer k NN-Prognose

Mit den ausgewählten, skalierten und gewichteten Variablen erfolgt die eigentliche Erstellung der k NN-Prognose, die Nachbarsuche.

Wie der Name dieser Methode besagt, handelt es sich um die Suche der k ähnlichsten Fälle. Es muss also noch definiert werden, wie gross k sein soll, das heisst nach wievielen ähnlichen Nachbarn in der Datenbank gesucht werden soll.

An dieser Stelle seien einige Literaturhinweise bezüglich der Wahl von k aufgeführt. Loftsgaarden & Quesenberry (1965) schlagen \sqrt{n} als klassenspezifische Wahl für k vor. Enas & Choi (1986) schlagen aufgrund von Simulationsstudien $n^{\frac{3}{8}}$ für grosse und $n^{\frac{2}{8}}$ für kleine Unterschiede in Klassenkovarianzen vor. Bei einem n von 4344 würde dies 23 respektive 8 Nachbarn ergeben.

In der vorliegenden Arbeit wurde $k = 10$ gewählt. Dies hat vor allem einen praktischen Grund, denn mit $k = 10$ bleibt die Modellausgabe übersichtlich. Der Modellanwender wird zudem in der Praxis keine Zeit haben eine zu grosse Anzahl an Nachbarn zu interpretieren. Es hat sich auch gezeigt, dass auch für $k = 5$ oder $k = 20$ das Verhältnis von positiven η^+ und negativen η^- Nachbarn ungefähr gleich bleibt.

5.6.1 Binäre Prognose (BF)

Bei der binären Prognose, welche einen einfachen, dichotomen Entscheid darstellt, muss ein Schwellenwert DB bestimmt werden, anhand dessen entschieden wird, ob der zu prognostizierende Tag in die Kategorie Ereignistag oder Nichtereignistag eingeteilt wird.

Der Schwellenwert wird anhand verschiedener Modellverifikationsverfahren optimiert. Zur Optimierung stehen mehrere Masse mit unterschiedlichen Gewichtungen der einzelnen Felder der 2×2 -Kontingenzmatrix zur Verfügung. Das in der Folge verwendete k NN-Modell optimiert die Prognose nach dem Heidke-Skill-Score HSS und man erhält daraus den Schwellenwert DB . Wie Abbildung 5.7 zeigt, wird der maximale HSS -Wert bei 3 bis 4 positiven von insgesamt 10 Nach-

barn erreicht. D.h. das Modell wertet einen Prognosetag mit 3 und mehr positiven Nachbarn als Ereignistag. Da der in Abbildung 5.10 prognostizierte Tag 4 positive Nachbarn aufweist, wird er in der binären Prognose als Ereignistag kategorisiert.

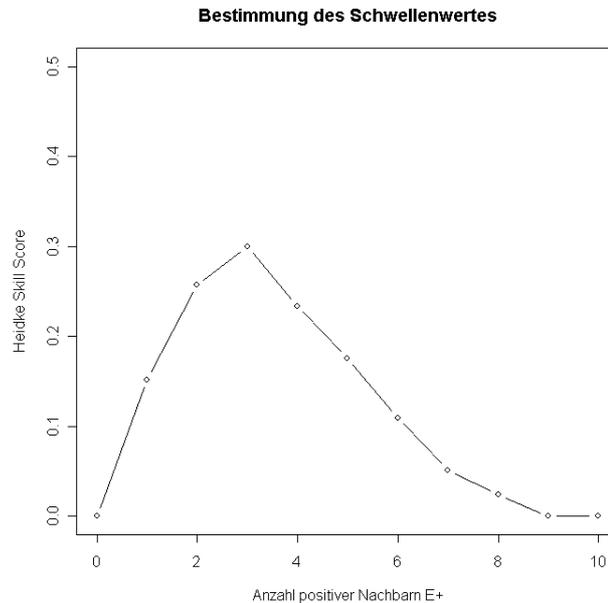


Abb. 5.7: Bestimmung des Schwellenwertes (decision boundary) mittels dem HSS für das k NN-Modell der Region Lugano.

Entsprechend der mit dem k NN-Modell zu untersuchenden Fragestellung, kann dieser Wert auch durch subjektiv festgelegte Kriterien bestimmt werden. In Abhängigkeit der Wahl dieses Grenzwertes werden in der Folge die Felder der Kontingenzmatrix gefüllt.

5.6.2 Probabilistische Prognose (PF)

Die probabilistische Prognose liefert eine Wahrscheinlichkeit für das Eintreten eines Waldbrandereignisses an dem zu prognostizierenden Tag y .

$$p_y = \frac{\eta^+}{k} \quad (5.4)$$

η^+ ist dabei die Anzahl gefundener Nachbarn mit einem Ereignis und k die Anzahl betrachteter Nachbarn insgesamt. p_y ist die Wahrscheinlichkeit, dass am zu prognostizierenden Tag y mindestens ein Waldbrand auftritt.

5. DAS k NN-WALDBRAND PROGNOSEMODELL

Die Problematik der PF liegt darin, die berechneten Wahrscheinlichkeiten p_y einer tatsächlich auftretenden Situation zuzuordnen. Eine mögliche Vereinfachung der Aussage würde darin bestehen, dass man die berechneten Werte klassifiziert. Durch die Einführung unterschiedlicher Gefahrenstufen wird dies auch in den meisten probabilistischen Prognosemodellen vorgenommen (Van Wagner, 1987; Viegas et al., 1999a). Dies ist allerdings nicht ganz unproblematisch, da die Waldbrandgefahr mit der in der Natur anzutreffenden Situation verknüpft wird, während die Wahrscheinlichkeit noch immer auf Ereignisse hinweist, welche im Datensatz zu finden sind.

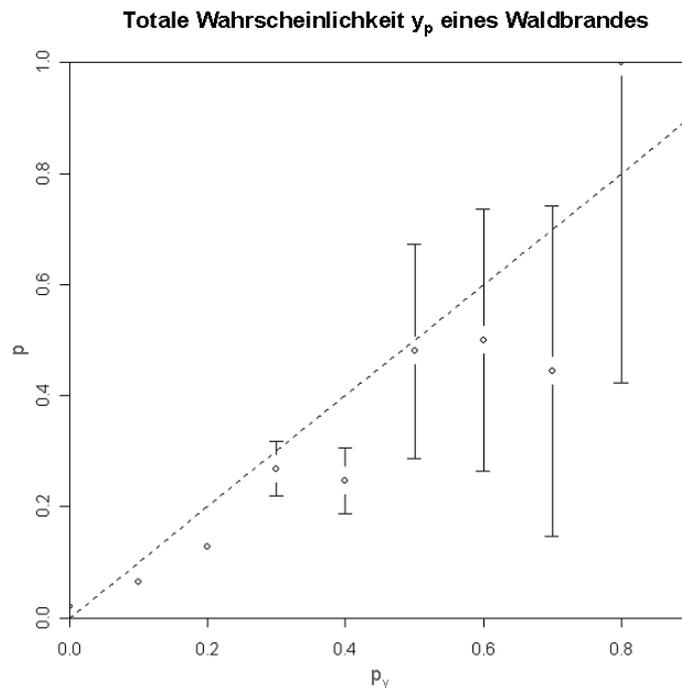


Abb. 5.8: Korrelation zwischen der mit dem k NN-Modell berechneten Wahrscheinlichkeit p_y eines Ereignisses und der tatsächlichen Häufigkeit p des Auftretens eines Waldbrandes. Bei den berechneten Wahrscheinlichkeiten p_y wurde zusätzlich noch die Unsicherheit der Prognose angegeben.

Abbildung 5.8 zeigt das Verhältnis von berechneten Wahrscheinlichkeiten p_y zu den Häufigkeiten p tatsächlich beobachteter Waldbrände. Liegen die Punkte unterhalb der Winkelhalbierenden, so hat das Modell die Gefahr überschätzt. Dieser Umstand kann dadurch erklärt werden, dass das Modell respektive die Variablen-

gewichtung auf die Ereignisse Waldbrand optimiert wurde. Da diese Ereignisse seltener sind als die Nichtereignisse, kommt es zu einer Überschätzung der Gefahr durch das Modell.

Zu den berechneten Wahrscheinlichkeiten p_y wurde jeweils die Prognoseunsicherheit e_p angegeben (vgl. Abb. 5.8).

$$\pm e_p = \frac{\eta^+}{\eta^+\eta^-} * \sqrt{\left(\frac{1}{\eta^+} + \frac{1}{\eta^-}\right)} \quad (5.5)$$

Dies verdeutlicht, dass Prognosen mit grossem p_y selten sind und somit die Vorhersageunsicherheit gross wird.

5.6.3 Deskriptive Prognose (DF)

Die deskriptive Prognose ist die dritte und umfassendste Ausgabemöglichkeit des k NN-Modells. Sie beinhaltet grundsätzlich die gesamte Information, welche in den beiden zuvor angesprochenen Methoden (BF und PF) bereits selektiert worden ist. Eine mögliche Ausgabeform stellt Tabelle 5.10 dar. In Abhängigkeit der Information (Variablen), welche in den beiden Grundlagendatenbanken vorhanden ist, wird diese Liste mehr oder weniger umfassend sein (Spalten).

Tab. 5.10: *Beispieloutput einer deskriptiven Prognose (DF) des k NN-Waldbrand Prognosemodells. Je mehr Information die Waldbranddatenbank beinhaltet, desto umfassender wird diese Ausgabe der DF.*

Datum	E	Betr. Fläche	Exp.	Höhe	Ursache	Waldtyp
17.3.91	1	47ha	S-SW	1070m.ü.M	Blitz	Kastanien
08.5.97	1	68ha	SW-W	620m.ü.M	Unbekannt	Kastanien
11.1.01	1	12ha	S-SW	1710m.ü.M	Brandst.	Föhren
01.2.99	0					
17.2.02	1	107ha	S-SE	280m.ü.M	Eisenbahn	Kastanien
15.5.01	0					
17.2.98	1	22ha	S-SW	1420m.ü.M	Blitz	Lärchen
18.2.98	0					
01.1.00	1	Unbekannt	SE-E	1890m.ü.M	Unbekannt	Kastanien
25.3.97	0					

5. DAS k NN-WALDBRAND PROGNOSEMODELL

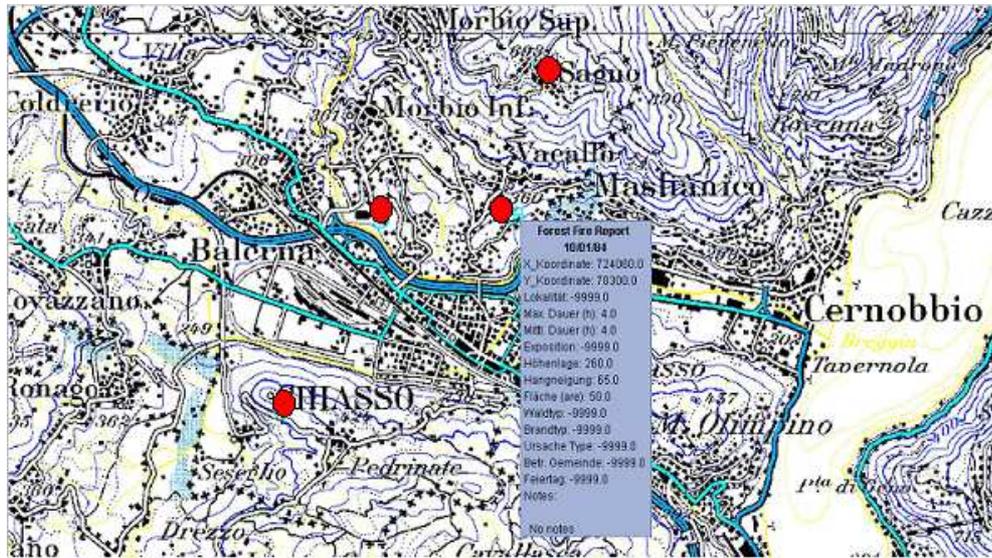


Abb. 5.9: Möglicher k NN-Output in Form einer Karte. Rot eingezeichnete Punkte zeigen Tage mit mindestens einem Ereignis und deren Lokalität an. Weiter ist die gesamte in der Waldbranddatenbank enthaltene Information bezüglich des einzelnen Ereignisses ersichtlich (Cornice plot).

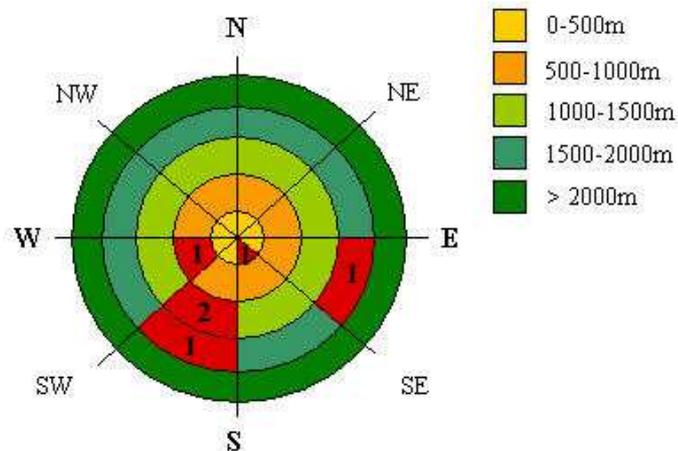


Abb. 5.10: Einfache statistische Auswertung der DF Prognose (Beispiel). Verteilung der in Tab. 5.10 aufgelisteten Waldbrände nach Exposition und Höhenlage.

Grundsätzlich wird zu jedem gefundenen Nachbarn eine möglichst umfangreiche Information mitgeliefert. Die Interpretation dieser deskriptiven Ausgabe erfolgt durch den Anwender, da das Modell nur jene Information liefert, welche in der

5. DAS k NN-WALDBRAND PROGNOSEMODELL

Vergangenheit bei einer ähnlichen Situation in die Datenbank eingegeben worden war. Für einen Anwender sehr hilfreich sind zusätzlich erarbeitete Analysetools wie z.B. Ausgabe als Karte (Abb. 5.9), oder vereinfachte statistische Auswertungen der Brandverteilung (Abb. 5.10).

Der Nachteil der DF gegenüber den beiden anderen Methoden liegt darin, dass die Auswertung und die Interpretation der Ergebnisse nur durch einen Experten vorgenommen werden kann. Um von den vergangenen Beispielen auf die aktuelle Situation schliessen zu können, braucht dieser viel Erfahrung. Dass diese Methode allerdings Erfolg haben kann, zeigte sich in der Lawinenprognose trotz mässigen BF und PF Ergebnissen (Heierli et al., 2003).

6 Verifikation von Prognosemodellen

6.1 Einführung

Nach der Entwicklung eines Modells müssen Fragen hinsichtlich der Anwendbarkeit gestellt werden: Ist das Modell überhaupt anwendbar? Wie genau sind die berechneten Ergebnisse? Repräsentiert das Modell die Realität in angemessener Art und Weise? Sind die richtigen Modellparameter gewählt worden? Gibt es bessere Modelle? Zusammengefasst: *Darf das Modell so verwendet werden?*

Das Konzept zur Beantwortung dieser Fragen wird meist *Validierung* (von *validus* (lat.) = gültig) genannt. In der Wissenschaft ist dieses Konzept jedoch umstritten und nicht einheitlich definiert. Verwirrung entsteht aufgrund der philosophischen Frage, in welchem Masse und ob überhaupt Modelle beziehungsweise wissenschaftliche Theorien geprüft werden können. Popper (1959, 1963) und Scherer (1998) meinen, dass wissenschaftliche Theorien nicht verifiziert sondern bestenfalls falsifiziert werden können. Dennoch hat die Öffentlichkeit ein eigenes, unscharfes Verständnis von Validierung und wird durch die verschiedenen Definitionen verwirrt. Unklarheit entsteht durch wechselweise Verwendung von Ausdrücken wie Validierung, Verifikation, Bewertung, Angemessenheit, etc.

Von *Verifikation* wird im allgemeinen dann gesprochen, wenn das Modell im Hinblick auf dessen Richtigkeit bezüglich der Wiedergabe der Realität überprüft wird. Der Begriff *Validierung* findet dann eine Verwendung, wenn das Modell im Hinblick auf dessen Gültigkeit bezüglich einem Verwendungszweck oder einer Hypothese überprüft wird (Hebel, 2003).

Im Folgenden wird der Ausdruck *Verifikation* (von *verus* (lat.) = wahr) verwendet, da dieser Begriff im Bereich der Qualitätsüberprüfung meteorologischer Prognosemodelle die grösste Verbreitung findet.

6. VERIFIKATION VON PROGNOSEMODELLEN

Anders als in den Bereichen der Risikoabschätzung von Naturgefahren, weist bei den meteorologischen Prognosesystemen die Modellverifikation sowohl in der Entwicklung als auch in der Anwendung eine lange Tradition auf. In den über 100 Jahren seit sich J.P. Finley (1884) mit der Verifikation von meteorologischen Prognosen (Tornadoprognose) beschäftigt hat, sind eine Vielzahl unterschiedlicher Methoden entwickelt worden. Die Bandbreite der zu verifizierenden Prognosen umfasst binäre Prognosen bis hin zu komplexen Ensembleprognosen (verschiedene Wettervariablen zu verschiedenen Zeiten in unterschiedlichen Druckniveaus). Dementsprechend breit gestreut sind die Ansichten, welche Methoden eine gute Verifikation berücksichtigen sollte (Murphy, 1993). Trotz der Vielzahl der entwickelten Verifikationsmethoden kam es bis zu den 80iger Jahren zu keinen signifikanten Weiterentwicklungen (Stephenson & Jolliffe, 2003). Erst mit den Publikationen von Mason (1980, 1982a,b, 1989) und Murphy & Winkler (1987) wurden neue Ansätze entwickelt.

Bei der Verifikation müssen bestimmte Grundsätze berücksichtigt werden. So muss eine Verifikationsmethode auf die Interessen des jeweiligen Nutzers ausgerichtet sein. Dies bedeutet, dass die Ansprüche des Nutzers an das Modell bekannt sein müssen, um zweckmässige Modellresultate zu erhalten, auf deren Basis sinnvolle Entscheidungshilfsmittel entwickelt werden können. Weiter muss berücksichtigt werden, dass die Qualität einer Vorhersage nicht durch ein universelles Mass ausgedrückt werden kann. Jedes einzelne Qualitätsmass gibt nur einen oder einige, aber nicht alle Aspekte der Prognosegüte wieder. Zur Verifikation ist somit immer eine geeignete Auswahl an Gütemassen in Kombination zu verwenden.

Zur Verifikation liegt immer ein Satz aus zueinander gehörenden Vorhersagen und Beobachtungen vor. Diese können - wie im vorliegenden Fall - entweder als binäre- (BF) oder als probabilistische (PF) bzw. als deskriptive Informationen (DF) vorliegen.

6.2 Trainingsdatensatz und Testdatensatz

Für die Bestimmung der Güte eines Prognosemodells muss eine Beziehung zwischen dem Datensatz mit der prognostizierten- und der reellen Grösse hergestellt werden. Dabei muss berücksichtigt werden, dass bei einem Modell nicht nur die Daten, anhand derer das Modell entwickelt wurde, möglichst gut mit der Rea-

lität übereinstimmen, sondern auch Datensätze gleichen Ursprungs, die zum Zeitpunkt der Entwicklung des Modells noch nicht vorlagen. Das Modell muss also die Fähigkeit zur Generalisierung besitzen.

Eine Methode, die sowohl die Genauigkeit als auch das Generalisierungsvermögen eines Modells testet, wenn nur ein Datensatz des zu modellierenden Systems vorliegt, ist die sogenannte *Cross-Validation* Methode (*Bootstrap-Methoden*) (Stahel, 2002). Bei diesem Verfahren muss der Gesamtdatensatz in zwei Teile aufgeteilt werden:

- In einen Trainingsdatensatz, auf dessen Grundlage das Modell erstellt beziehungsweise trainiert wird, und
- in einen Testdatensatz, an welchem der Vorhersagefehler der Modelleinstellungen (Variablenwahl, Variablenengewichtung, etc.) bestimmt respektive minimiert wird.

Trainings- und Testdaten sollten sinnvollerweise unabhängig voneinander sein. Diese Bedingung ist im gegebenen Fall erfüllt, da die meteorologischen Messwerte eines Tages unabhängig von vorhergehenden sind. Im vorliegenden Fall wurde die extreme Variante der *Cross-Validation*, die sogenannte *Leave-one-out* Methode gewählt. Dabei wird der Datensatz der Grösse n wiederholt in einen Trainingsdatensatz der Grösse $n - 1$ und einen Testdatensatz der Grösse 1 aufgeteilt (Evgeniou et al., 2004).

Werden im k NN-Modell die Gewichte der Variable nach der Hinzunahme eines neuen Datenpunktes \mathbf{x} für jeden Tag neu berechnet und in der Prognose des Folgetages berücksichtigt, wird die *Leave-one-out* Methode für jede neue Prognose angewendet. Ist der Datensatz gross genug, kann auf eine tageweise Anpassung der Gewichte verzichtet werden. Im k NN Waldbrandmodell wurden die Daten jeweils am Ende eines jeden Jahres aktualisiert. Die Beweggründe dafür liegen in der Tatsache, dass die Aktualisierung der Waldbranddatenbank mehr Zeit in Anspruch nimmt und daher nicht täglich erweitert werden kann.

6.3 Arten der Verifikation

Ziel einer Verifikation ist es, bestimmte Eigenschaften von Prognosen herauszufinden und diese in Bezug zur Realität zu beurteilen. Gemäss Damrath (2002) gehören dazu folgende Attribute:

Tab. 6.1: Zusammenfassung der wichtigsten Attribute einer Prognose (Damrath, 2002).

Eigenschaft der Vorhersage	Definition	geeignete Gütemaßzahl
mittlerer Fehler	Beziehung zwischen Mittelwert der Vorhersage und der Beobachtung	mittlerer Fehler(BIAS)
Zueinanderpassen	Grad des linearen Zusammenhangs zwischen Vorhersage und Beobachtung	Korrelationsmaße
Genauigkeit	mittlere Übereinstimmung zwischen individuellen Paaren der Vorhersage und der Beobachtung	mittlerer absoluter Fehler, mittlerer quadratischer Fehler, Brier Score usw.
Gewinn gegenüber Alternativen (Skill)	Genauigkeit der Vorhersage im Vergleich zur Genauigkeit einer alternativen bzw. trivialen Vorhersage (Erhaltungsneigung, Klima,...)	Brier Skill Score, Reduktion der Varianz, ...
Vertrauenswürdigkeit	Zusammenhang zwischen bedingten Mittelwerten von Beobachtung und Vorhersage (z. B.: beobachtete und vorhergesagte Mittelwerte der Temperatur bei Strahlungsbedingungen)	Komponenten der Gütemaße, die bedingte Aussagen enthalten
Unterschiedlichkeit der Beobachtungen	Unterschied zwischen den bedingten Mittelwerten und dem Gesamtmittel der Beobachtungen	relevante Komponente des Brier Score
Schärfe	Variabilität der Vorhersage	Varianz der Vorhersage
Unterschiedlichkeit der Vorhersage	Unterschied zwischen den bedingten Mittelwerten und dem Gesamtmittel der Vorhersage	Maße für bedingte Vorhersagefehler
Unsicherheit	Variabilität der Beobachtungen	Varianz der Beobachtungen

Bei Damrath (2002) wird zwischen einer **subjektiven** und einer **objektiven** Verifikation unterschieden. Diese Aufteilung lässt sich gut auf die in Kapitel 4.4 beschriebenen Ausgabemöglichkeiten des k NN-Modells übertragen.

Die deskriptive Prognose DF ist eine **subjektive** Interpretation der ausgegebenen Liste, welche nur diese Art einer Verifikation zulässt. Dabei können sehr unterschiedliche Zustände als Wahrheit aufgefasst werden. Das Urteil zur Qualität der Vorhersage wird individuell unterschiedlich ausfallen. Unterschiedliche Personen werden unterschiedliche Kriterien verwenden. Diese können auch situationsbedingt voneinander differieren. Mit subjektiven Mitteln wird es im allgemeinen möglich sein, für Einzelfälle zweckmässige Hinweise auf Fehler einer Vorhersage

zu erhalten. Es ist allerdings schwierig diese Fehler zu systematisieren, wenn für die subjektive Beurteilung nicht einheitliche Kriterien verwendet werden (Darrath, 2002).

Für das k NN-Lawinenmodell wurde die Qualitätsbewertung der DF in Zusammenarbeit mit einem Modellanwender auf ein Gebiet angewendet. Nach Aussagen des Anwenders war es vor allem die qualitative Information (z.B. Beschreibung der Ereignisse, Art und Anzahl der Hindernisse zwischen der Ablagerung und dem Zielobjekt, etc.), welche die Qualitätsbewertung ermöglichte. Folgende Abbildung 6.1 fasst die Güte der DF zusammen.

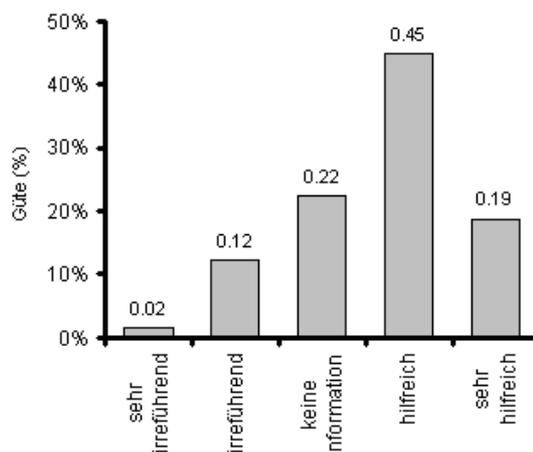


Abb. 6.1: Subjektive Einschätzung der “Nützlichkeit” vom NXD-Lawinenmodell. Die Auswertung wurde nach Informationen eines Nutzers des NXD-Modells im Kanton Wallis durchgeführt. (Heierli et al., 2003).

Mit Hilfe der **objektiven** Verifikation ist es möglich Vorhersagen und Beobachtungen miteinander zu vergleichen und aufbauend statistische Masszahlen zu berechnen. Der Vorteil der objektiven Prognose besteht darin, dass sie von individuellen Urteilen frei ist und systematische Vorhersagefehler gut erkannt werden können. Zur Beurteilung der Prognosequalität reicht eine einzige Masszahl im allgemeinen jedoch nicht aus. Je nach Typ der Vorhersagen und des vorherzusagenden Ereignisses ist es notwendig unterschiedliche Arten von Qualitätsmassen zu berechnen. Durch die Mehrzahl der benötigten Masse wird die Verifikation allerdings nicht immer trivial.

6.4 Verifikation von binären Prognosen

Um die Qualität von binären Prognosen (BF) zu berechnen, sollten die Daten in einer 2×2 -Kontingenzmatrix organisiert werden. Tabelle 6.2 zeigt die vier möglichen Kombinationen von Prognose-Ereignis Paaren in der Kontingenzmatrix. a ist die Anzahl jener Fälle, in denen das interessierende Ereignis sowohl beobachtet als auch prognostiziert wurde. d ist die Anzahl der Fälle, in denen das Ereignis nicht beobachtet und auch nicht vorhergesagt wurde. Die Summe $a + d$ ergibt die Anzahl der korrekten Vorhersagen, während b und c die Anzahl der irrtümlichen Vorhersagen darstellen. Die relativen Häufigkeiten p der einzelnen Felder (a bis d) können bestimmt werden, indem die entsprechenden Werte von a bis d durch die Gesamtsumme aller Tage (n) dividiert werden (Tab. 6.3).

Tab. 6.2: Kontingenzmatrix kategoriemer Prognosen binärer Ereignisse.

Prognose	Ereigniseintritt		Total
	Ja	Nein	
Ja	a	b	$a+b$
Nein	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d=n$

Tab. 6.3: Kontingenzmatrix kategoriemer Prognosen binärer Ereignisse. Joint und marginal Wahrscheinlichkeiten.

Prognose	Ereigniseintritt		
	Ja	Nein	
Ja	$p(\hat{x} = 1, x = 1)$	$p(\hat{x} = 1, x = 0)$	$p(\hat{x} = 1)$
Nein	$p(\hat{x} = 0, x = 1)$	$p(\hat{x} = 0, x = 0)$	$p(\hat{x} = 0)$
Total	$p(x = 1)$	$p(x = 0)$	1

Aus den vier Elementen der Kontingenzmatrix lassen sich eine Reihe von Gütemassen berechnen, mit denen unterschiedliche Eigenschaften von Prognosen charakterisiert werden können. Diese Gütemasse wurden meist in der englischsprachigen Literatur eingeführt, sodass im allgemeinen auch die englischen Bezeichnungen gebräuchlich sind.

(a) Probability of detection (POD)

$$POD = \frac{a}{a+c} = \hat{p}(\hat{x} = 1|x = 1) \quad (6.1)$$

Wertebereich: 0 bis 1

Bester Wert: 1

Eigenschaften: Es werden nur richtige JA-Aussagen erfasst, falsche Alarme werden ignoriert. Durch zu häufiges Vorhersagen des JA-Fälle kann *POD* maximiert werden. In der Literatur sind auch die Bezeichnungen *hit rate* (Stephenson & Jolliffe, 2003) oder wie in Wilks (1995) fälschlicherweise sogar *proportion correct* zu finden. Da eine Optimierung des Prognoseskills von der Maximierung des *POD*'s und der gleichzeitigen Minimierung von *F* abhängt, ist die *POD* als alleiniges Qualitätsmass unzureichend.

(b) False alarm ratio (FAR)

$$FAR = \frac{b}{a+b} = \hat{p}(x = 0|\hat{x} = 1) \quad (6.2)$$

Wertebereich: 0 bis 1

Bester Wert: 0

Eigenschaften: Es werden nur falsche Alarme berücksichtigt. Das Gütemass kann somit durch eine zu geringe Häufigkeit von JA-Aussagen maximiert werden. Werden *POD* und *F* als konstant betrachtet, nimmt der Wert von *FAR* bei steigender Grundwahrscheinlichkeit *s* ab.

POD und *FAR* sind eigenständig keine perfekten Gütemasse, weil sie jeweils nur einen Teilaspekt der Vorhersagequalität (*forecast skill*) berücksichtigen.

Anhand der beiden Masse *POD* und *FAR*, kann die Güte des Modells subjektiv abgeschätzt werden. Die beiden Grössen müssen allerdings gemeinsam berücksichtigt werden. Eine Erhöhung des *POD* ist einfach möglich, dies allerdings nur

6. VERIFIKATION VON PROGNOSEMODELLEN

auf Kosten des FAR . Demgegenüber kann FAR bis zum Wert der Grundwahrscheinlichkeit s reduziert werden, dies aber ebenfalls nur auf Kosten des POD . Das Paar muss daher so kombiniert werden, dass POD "gut" und FAR "akzeptabel" wird. Die Abbildung 7.1 stellt das Verhältnis POD zu FAR für die Region Lugano dar.

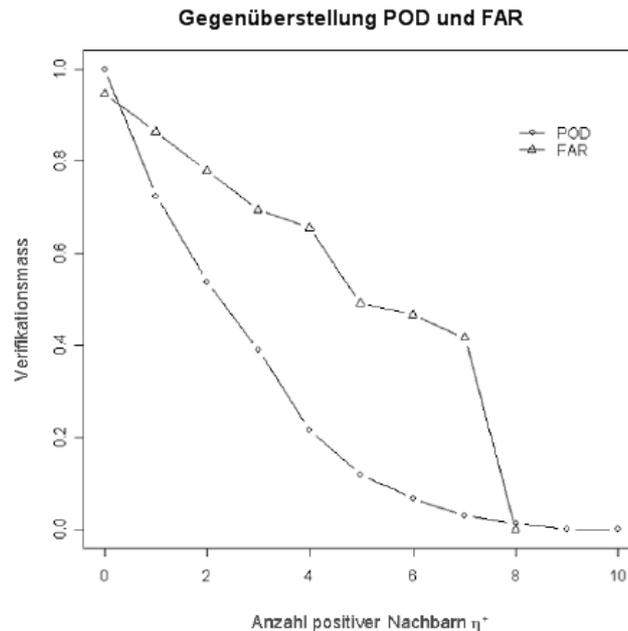


Abb. 6.2: Gegenüberstellung POD und FAR der Region Lugano (1991-2002). (Beispiel).

Obwohl die Kombination von POD und FAR ein zweckmässiges Gütemass darstellt und mit deren Hilfe eine praxisorientierte Aussage gemacht werden kann (wie oft wurde zuviel gewarnt, wie oft zuwenig), reicht sie nicht aus, um ein vollständiges Bild der Qualität des Prognosemodells zu bekommen. Jede 2×2 -Kontingenzmatrix besitzt drei Freiheitsgrade. Das bedeutet, es fehlt noch ein weiteres Mass um eine umfassende Information bezüglich der Qualität zu erhalten. Dieses Mass, beispielsweise der HSS oder TSS , sollte daher unabhängig von POD und FAR sein.

(c) False alarm rate (F)

$$F = \frac{b}{b+d} = \hat{p}(\hat{x} = 1|x = 0) \quad (6.3)$$

Wertebereich: 0 bis 1

Bester Wert: 0

Eigenschaften: Die *false alarm rate* F ist oft unter dem Begriff *POFD* (*probability of false detection*) zu finden. Wichtig ist, dass F nicht mit FAR verwechselt wird.

(d) Prozentzahl korrekter Vorhersagen (PEC) (percent correct score)

$$PEC = \frac{a+d}{n} = \hat{p}(\hat{x} = 1) \quad (6.4)$$

Wertebereich: 0 bis 1

Bester Wert: 1

Eigenschaften: Es dominiert die häufigste Kategorie. Sind die Ereignisse selten, liegt PEC meist nahe bei 1 und täuscht gute Vorhersageleistungen vor. PEC kann auch durch POD und F ausgedrückt werden (Gl. 6.5).

$$PEC = (1-s)F + sPOD \implies POD = -\left(\frac{1-s}{s}\right)F + \frac{PEC}{s} \quad (6.5)$$

(e) Critical success index (CSI) (threat score)

$$CSI = \frac{a}{a+b+c} \quad (6.6)$$

Wertebereich: 0 bis 1

Bester Wert: 1

Eigenschaften: Der CSI ist zur allgemeinen Charakterisierung der Vorhersagen besser geeignet als POD und FAR , da sowohl falsche Alarme als auch fehlende richtige Vorhersagen miteinbezogen werden. Allerdings können die Ergebnisse von Stichproben mit unterschiedlichen Probenumfang, das heisst bei einer nicht gleichmässigen Aufteilung von Ereignis- und Nichtereignistagen, nicht miteinander verglichen werden.

(f) Frequency bias (FBI)

$$FBI = \frac{a + b}{a + c} \quad (6.7)$$

Wertebereich: 0 bis ∞

Bester Wert: 1

Eigenschaften: Dieses Mass liefert keine direkte Aussage zur Güte der Vorhersage, sondern gibt nur das Verhältnis der prognostizierten zur beobachteten Häufigkeit des JA-Ereignisses an.

Perfekte Prognosen mit $POD = 1$ und $F = 0$ sind immer *unbiased* (erwartungstreu) ($FBI = 1$). Bei Prognosen mit keinem *Skill* $POD = F$ ist der *Bias* (systematischer Fehler) gleich POD/s (oder F/s). Ist der Bias >1 , bedeutet dies, dass mehr Prognosen gemacht wurden, als Ereignisse auftraten, was als *overforecasting* bezeichnet wird. Wünschenswert wäre ein Bias = 1, d.h. das Ereignis wurde in gleichem Masse prognostiziert, wie es effektiv auftrat.

Die bisher beschriebenen Gütemasse sind vergleichbar mit dem *mittleren absoluten Fehler* (*MAE*) oder der *Wurzel aus dem mittleren quadratischen Fehler* (*RMSE*) für Fehlercharakteristiken kontinuierlicher Werte. Diese drücken die Güte der Vorhersage an sich ohne Vergleich zu einer alternativen Prognose aus (Damrath, 2002). Analog zur *Reduktion der Varianz* (*RV*) können für binäre Werte auch *Skill scores* (*SS*) gewählt werden.

Forecast Skill

Forecast Skills geben eine relative Genauigkeit der Vorhersage in Bezug auf einen Satz von Standard- respektive Referenzprognosen an. Der *Skill Score* *SS* (Gl. 6.8) wird dabei als prozentuale Verbesserung gegenüber dieser Referenzprognose interpretiert.

$$SS = \frac{A - A_{\text{ref}}}{A_{\text{perf}} - A_{\text{ref}}} \quad (6.8)$$

A stellt dabei ein bestimmtes Mass für die Genauigkeit der Prognose dar. Die Wahl der Referenzprognose (A_{ref}) ist eine Funktion der Zeit. Für längere Prognosezeiträume wird die Klimatologie, für kürzere Zeiträume, wie in unserem Fall,

die Persistenz, als Referenzprognose genommen. Ein SS von 0% ($A = A_{\text{ref}}$) bedeutet, dass gegenüber einer Referenzprognose (z.B. Persistenz- oder zufälligen Prognose) keine Verbesserung erzielt wurde. Ein SS von 100% bedeutet perfekte Prognose ($A = A_{\text{perf}}$).

(g) Heidke Skill Score HSS

$$HSS = \frac{\frac{(a+d)}{n} - \frac{[(a+b)(a+c)+(b+d)(c+d)]}{n^2}}{1 - \frac{[(a+b)(a+c)+(b+d)(c+d)]}{n^2}} = \frac{2(ad - bc)}{(a+c)(c+d) + (a+b)(b+d)} \quad (6.9)$$

Wertebereich: -1 bis 1

Bester Wert: 1

Eigenschaften: Eine grosse Verbreitung findet der auf POD basierende *Heidke Skill Score* (HSS). Auch dieses Gütemass hängt von der Grundwahrscheinlichkeit s ab und ist somit kein unabhängiges Qualitätsmass.

(h) True Skill Statistic

$$TSS = \frac{(a * d) - (b * c)}{(a + c) * (b + d)} = \frac{a}{a + c} + \frac{d}{b + d} - 1 \quad (6.10)$$

Wertebereich: -1 bis 1

Bester Wert: 1

Eigenschaften: Der TSS ist die Summe aus POD für das JA-Ereignis und POD für das NEIN-Ereignis. Aus Normierungsgründen wird 1 subtrahiert. Eine zufallsverteilte Serie liefert einen TSS -Wert von 0. Bei Werten über 0 ist eine Vorhersage also besser als Zufall.

Weitere häufig verwendete Skills sind in Stephenson & Jolliffe (2003) und Wilks (1995) beschrieben.

6.4.1 Relative Operating Characteristic (ROC)

Mit der Einführung der *signal detection theory* (Mason, 1980) respektive der daraus entwickelten *relative operating characteristic* (ROC) wurde ein Gütemass entwickelt, mit dessen Hilfe die Güte eines Modells graphisch dargestellt und analysiert werden kann. Dazu dient ein ROC -Diagramm, bei welchem F auf der Abszisse gegen POD auf der Ordinate aufgetragen wird.

6. VERIFIKATION VON PROGNOSEMODELLEN

Sind die Variablen wie im Fall der vorliegenden Arbeit nicht normalverteilt, sind die Werte des *ROC*-Diagramms nur eingeschränkt aussagekräftig.

Das Skillmass *ROC* ist die Fläche A_z (Gl. 6.11) unter der Kurve r (Abb. 6.3). Die berechnete Fläche A_z liegt im Bereich $[0,1]$. A_z Werte von 0.5 bedeuten keinen Skill, die perfekte Kurve würde durch die Punkte $(0,0)$, $(0,1)$ und $(1,1)$ verlaufen, was eine Fläche $A_z = 1$ ergibt

$$A_z = \int_0^1 r \quad (6.11)$$

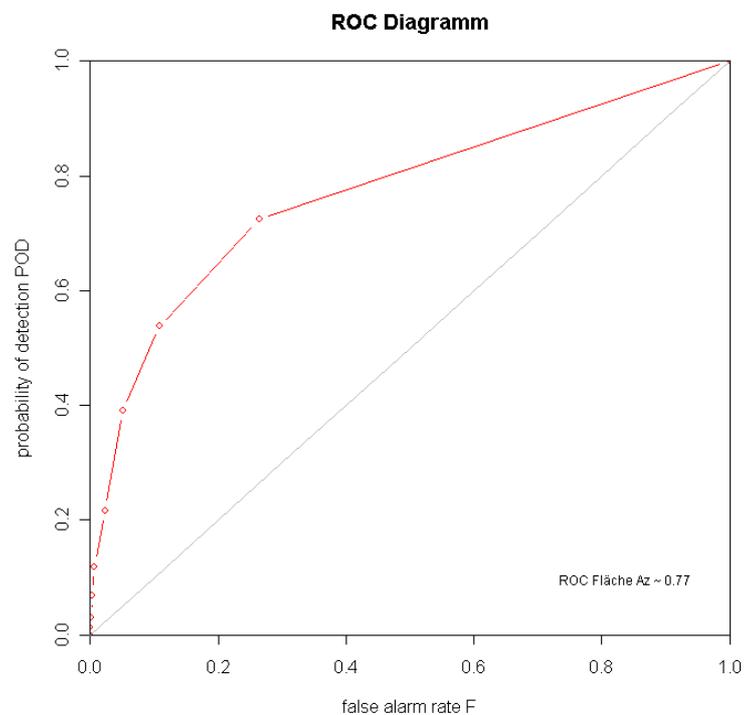


Abb. 6.3: *ROC-Diagramm für die Waldbrandprognose des kNN-Modells der Region Lugano (1991-2002). (Beispiel).*

6.5 Verifikation von probabilistischen Prognosen

Wie wissen wir, ob eine Wahrscheinlichkeitsvorhersage richtig oder falsch ist?

Sobald eine Vorhersage in Form von Wahrscheinlichkeiten ausgedrückt wird, werden alle möglichen Ergebnisse vorausgesagt. Aber: Das Vertrauen beziehungsweise die Vertrauenswürdigkeit in die Prognose kann unterschiedlich eingeschätzt werden. Im Gegensatz zur binären Prognose kann eine einzelne Wahrscheinlichkeitsvorhersage niemals “richtig” oder “falsch” sein, es sei denn, die Wahrscheinlichkeit sei 0 oder 1.

Neben den genannten “klassischen” *measure-oriented* Modellverifikation wurden in jüngster Zeit neuere, *distribution-oriented* Ansätze, entwickelt (Murphy & Winkler, 1987). Damit berechnete Gütemasse dienen dazu die Qualität der probabilistischen Prognose zu überprüfen. Nach Murphy & Winkler (1987) spielt die *joint distribution* (Gl. 6.12) bei der Prognoseverifikation eine grundlegende Rolle. Bei einer 2×2 Kontingenzmatrix, müssen drei Parameter bekannt sein um eine *joint distribution* zu rekonstruieren.

$$p(\hat{x}, x) \tag{6.12}$$

Die *joint distribution* kann in eine Bedingte- (*conditional*) und in eine Randverteilung (*marginal*) faktorisiert werden:

$$p(\hat{x}, x) = p(x|\hat{x})p(\hat{x}) \tag{6.13}$$

$$p(\hat{x}, x) = p(\hat{x}|x)p(x) \tag{6.14}$$

Aus der Sicht des Prognostikers ist die *calibration-refinement factorization* (Gl. 6.13) die gebräuchlichere. Sie beinhaltet Aussagen darüber, wie gut jede Prognose kalibriert ist respektive wie oft jeder der möglichen Prognosewerte verwendet wird.

Die zweite mögliche Faktorisierung der bedingten Wahrscheinlichkeit von Prognosen und Ereignissen ist die *likelihood-base rate factorization* (Gl. 6.14). Detailliertere Informationen bezüglich der beiden Ansätze sind in Murphy & Winkler (1987) und Stephenson & Jolliffe (2003) aufgeführt.

Ein weiteres Mass für die Verifikation probabilistischer Vorhersagen ist der **Brier Score** (*BS*) (Gl. 6.15) (Damrath, 2002).

6. VERIFIKATION VON PROGNOSEMODELLEN

$$BS = \frac{1}{n} \sum_{j=1}^n (\hat{p}_j - x_j)^2 \quad (6.15)$$

\hat{p}_j ist der Wert der Wahrscheinlichkeitsvorhersage, x_j der Beobachtungswert. Eine von Murphy (1973) vorgeschlagene Aufteilung von BS führt zu

$$BS = \frac{1}{n} \sum_{k=1}^L \sum_{j=1}^{Mk} [(\hat{p}_k - \bar{x}_k)^2 - (\bar{x}_k - \bar{x})^2 + (\bar{x} - x_{jk})^2]. \quad (6.16)$$

Dazu wird das mögliche Spektrum der Wahrscheinlichkeiten in L Kategorien eingeteilt. Mit \hat{x}_k wird der mittlere Wert der beobachteten Häufigkeit in der k -ten Kategorie bezeichnet. \hat{x} ist der Mittelwert der beobachteten Häufigkeiten in der gesamten Datenbank. Mk sind die Anzahl Fälle in der Kategorie k . Es ist zu beachten, dass dieses k nicht mit dem k der k NN-Methode verwechselt wird. In diesem Fall steht die Variable k für die Kategorien, in welche die Fälle eingeteilt wurden. Der erste Term beschreibt die tatsächliche Differenz zwischen Vorhersage und der mittleren Häufigkeit in der jeweiligen Kategorie. Er entspricht dem in Tabelle 6.1 umschriebenen Begriff "Genauigkeit" (*reliability*). Der zweite Term beschreibt die in Tabelle 6.1 genannten Unterschiede zwischen bedingten Mittelwerten und dem Gesamtmittel (*resolution*). Schliesslich entspricht der dritte Term der in der Tabelle 6.1 mit Varianz (*uncertainty*) bezeichneten Kategorie (Stephenson & Jolliffe, 2003). Die *uncertainty* hängt nur von der Eintrittswahrscheinlichkeit des Ereignisses ab und ist somit unabhängig von der Vorhersage.

Der *Brier Score* (Gl. 6.16) beinhaltet somit aufgrund seiner quadratischen Struktur Informationen bezüglich der Genauigkeit (*reliability*), Unterschiedlichkeit der Beobachtungen (*resolution*) und Varianz (*uncertainty*).

$$BS = BS_{reliability} - BS_{resolution} + BS_{uncertainty} \quad (6.17)$$

Bei einer perfekten Prognose wird BS 0.

Mit Hilfe eines Zuverlässigkeitsdiagrammes (*attributes diagram*), können die aufgeführten Komponenten des BS (Gl. 6.17) auch geometrisch dargestellt werden.

6.5.1 Zuverlässigkeitsdiagramm (attributes diagram)

Bei einem Zuverlässigkeitsdiagramm wird die prognostizierte Wahrscheinlichkeit p_y gegen die beobachtete relative Häufigkeit p aufgetragen (Abb. 6.4). Dabei veranschaulicht die Diagonale das Verhalten eines perfekten Modells. Die schattierte Fläche gibt den Bereich an, in welchem ein Modell einen positiven Skill gegenüber der Referenzprognose hat. Die Kurve, welche durch die berechneten Werte gebildet wird, folgt bei idealer Prognose der Diagonale, sollte jedoch auf keinen Fall unterhalb der grau gekennzeichneten Fläche verlaufen.

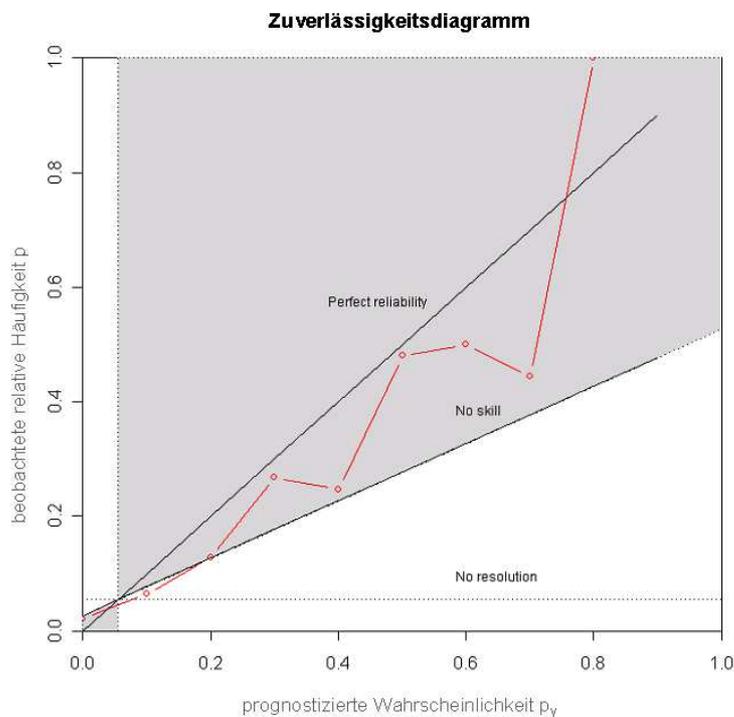


Abb. 6.4: Zuverlässigkeitsdiagramm (attributes diagram) für die Waldbrandprognose des k NN-Modells der Region Lugano (1991-2002). (Beispiel).

6.6 Verifikation von deskriptiven Prognosen

Die grossen Vorteile des k NN-Modells liegen in der deskriptiven Prognose. Mit den Begriffen Prognose oder Warnung ist jedoch hierbei grosse Vorsicht geboten. Prinzipiell liefert das k NN-Modell die Information darüber, was bei ähnlichen

6. VERIFIKATION VON PROGNOSEMODELLEN

meteorologischen Bedingungen in der Vergangenheit passiert ist. Dies stellt im Grunde genommen eine Aussage über eine potentiell mögliche Gefahr dar. Eine objektive Verifikation dieser Information ist kaum möglich, denn grundsätzlich ist die Güte der *DB* dem subjektiven Urteil des jeweiligen Anwenders unterworfen. Sehr häufig basieren menschliche Entscheidungen auf dem Prinzip der *Nearest Neighbors*. Intuitiv angewendete Entscheide wie: “...das letzte Mal war es auch gut so...” oder “...ähnlich wie vor 5 Jahren, sollten wir auch jetzt...” sind rational oft nicht erklärbar. Im Prinzip führen Warndienste solche Überlegungen täglich aus. Meist sind die Entscheidungen klar, wenn es beispielsweise regnet oder seit zwei Wochen nicht geschneit hat. Aus Erfahrung weiss man, dass somit keine unmittelbare Gefährdung vorliegt. Schwieriger wird es in Fällen, die nicht mehr eindeutig sind. 70cm Neuschnee oder eine seit Tagen anhaltende Föhn-situation rufen sofort eine auf Erfahrung basierende Einschätzung einer möglichen Gefährdung hervor. Das *k*NN-Modell hilft in solchen Fällen sich besser an vergangene Situationen zu erinnern. Der Vergleich von gegenwärtiger Situation und Vergangenheit - allenfalls unter Berücksichtigung weiterer Umstände - soll dem Anwender helfen, die momentane Gefährdungslage besser einzuschätzen. Die Entscheidung selbst wird somit anhand subjektiver Eindrücke gefällt, allerdings mit Hilfe einer einfachen statistischen Auswertung der vergangenen Situationen erhärtet. Mit der Auflistung der Ereignisse der Vergangenheit wird die Interpretation aber bewusst dem Experten überlassen.

7 Resultate

7.1 Einführung

Im Rahmen der vorliegenden Arbeit wurde auf eine detaillierte Auflistung einzelner Zwischenergebnisse verzichtet. Dies hätte Verständnis und Lesbarkeit stark beeinträchtigt und einen sinnvollen Umfang der Arbeit überstiegen. Aufgenommen wurden zusammenfassende Darstellungen, Auswertungsergebnisse und Analysen in dem Mass, in dem diese einen ausreichend umfassenden Überblick über die Modellergebnisse geben.

7.2 Binäre Prognose (BF)

Die binäre Prognose (BF), welche einen dichotomen Ja-Nein-Entscheid trifft, stellt eine mögliche Ausgabeform des k NN-Modells dar (Kap. 4.3). Unabhängig von der Variablenauswahl und deren Gewichtung hängt die Prognosegüte vom gewählten Schwellenwert (DB) (*decision boundary*) ab. Die in der Folge berechneten Ergebnisse basieren allesamt auf dem k NN-Modell, bei welchem $DB = 3$ angenommen wurde.

Anhand des PEC lässt sich die Prozentzahl der korrekten Vorhersagen einfach bestimmen. Die Werte liegen zwischen $PEC = 0.87$ für Lugano und Locarno und $PEC = 0.92$ für Piotta. Die hohe Anzahl korrekter Prognosen wird grösstenteils durch korrekt prognostizierte Nichtereignisse beeinflusst. Somit ist dieses Mass, obwohl bei der Verifikation von Prognosemodellen oft angegeben, als alleinige Angabe unbrauchbar und täuscht oftmals eine zu hohe Qualität vor.

POD und FAR stellen weitere, oft verwendete Qualitätsmasse von Prognosemodellen dar (Stephenson & Jolliffe, 2003). Ein optimales Verhältnis zwischen POD und FAR ist dann erreicht, wenn die Summe der beiden Masse $POD + FAR = 1$

7. RESULTATE

ergeben. Für $DB = 3$ sind die Ergebnisse von POD und FAR in den Abbildungen 7.1, 7.2 und 7.3 dargestellt.

Analysen von k NN-Lawinenmodellbenutzern haben ergeben, dass POD -Werte, in Abhängigkeit der Fragestellung (z.B. Strassensperrung oder Skigebietssicherung), zwischen 60% und 100% und FAR -Werte zwischen 0% und 100% ideal wären (Heierli et al., 2003). Die Abbildungen 7.1 bis 7.3 zeigen aber, dass diese Werte bei Annahme einer realistischen FAR bei Weitem nicht erreicht werden können. Auf zu viele, oder noch schlimmer, zu wenige Warnungen reagiert der Mensch sehr sensibel und lehnt ein solches Prognosemodell als ungenau ab.

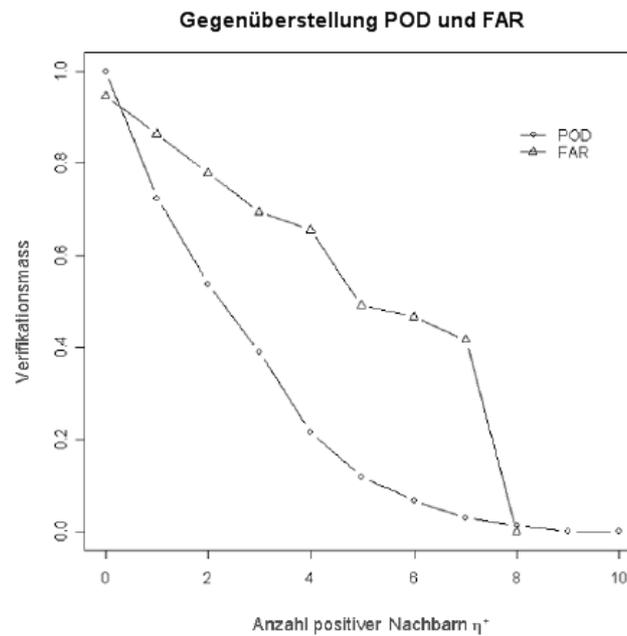


Abb. 7.1: Gegenüberstellung POD und FAR der Region Lugano (1991-2002).

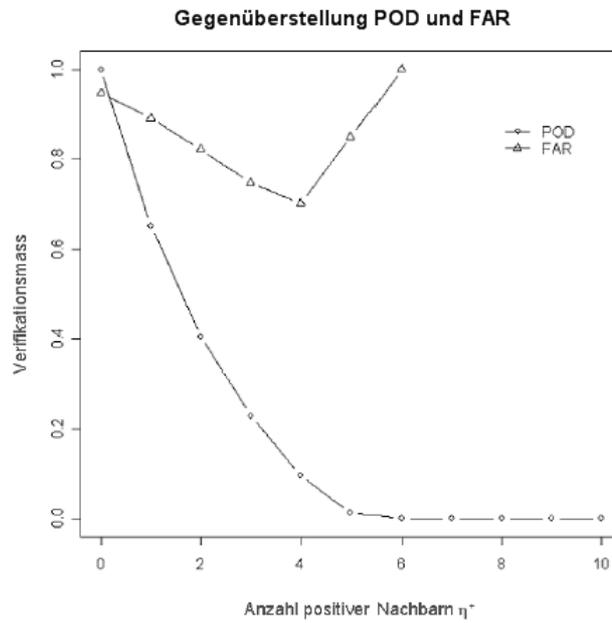


Abb. 7.2: Gegenüberstellung POD und FAR der Region Locarno (1991-2002).

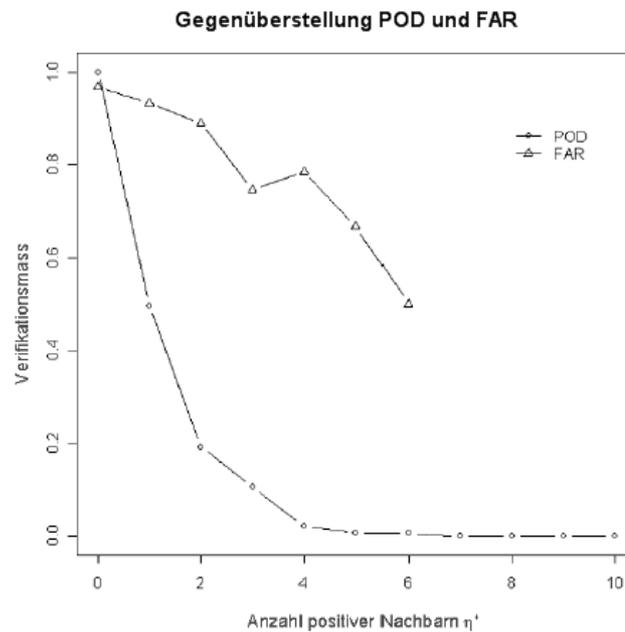


Abb. 7.3: Gegenüberstellung POD und FAR der Region Piotta (1991-2002).

7. RESULTATE

Abbildung 7.4 zeigt die berechneten *HSS*-Werte für die drei Regionen. Es fällt auf, dass in allen drei Untersuchungsgebieten das Maximum bei einem Schwellenwert von drei positiven Nachbarn η^+ zu finden ist. Die *HSS*-Werte schwanken allerdings zwischen 0.13 oder 13% für Piotta, 0.20 respektive 20% für Locarno bis 0.30 respektive 30% für Lugano. Die mit dem *k*NN-Modell erstellte Prognose stellt somit eine klare Verbesserung gegenüber einer zufälligen Prognose dar. Diese ist allerdings im nördlichsten der drei Untersuchungsgebiete gering. Auch der *HSS* ist - obwohl ein weit verbreitetes Qualitätsmass - von der Grundwahrscheinlichkeit *s* abhängig. Vergleiche mit anderen Prognosemodellen sind somit ebenfalls nicht möglich.

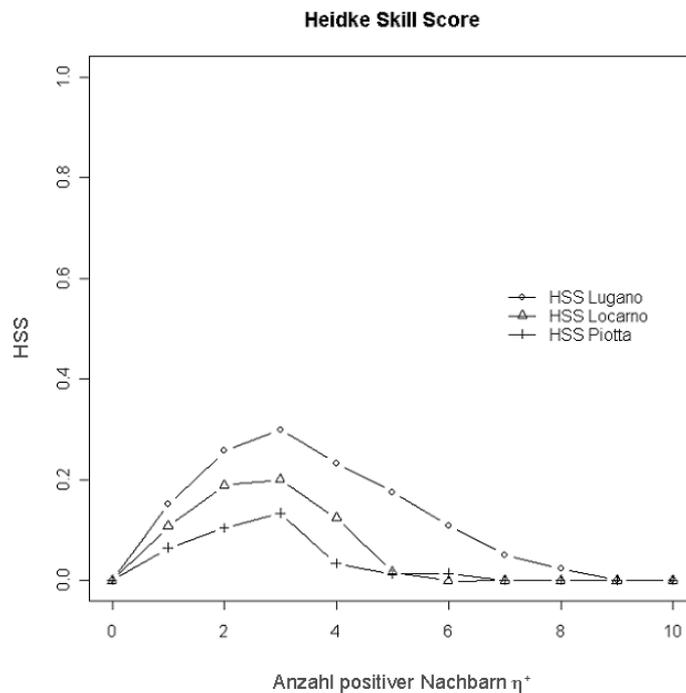


Abb. 7.4: Heidke Skill Score der *k*NN-Modelle für die Prognose der Regionen Lugano, Locarno und Piotta (1991-2002).

7.2.1 Relative Operating Characteristic (ROC)

Bei der Qualitätsanalyse mittels *ROC* wird F gegen POD aufgetragen. Abbildung 7.5 zeigt die *ROC*-Kurve für die drei Regionen. Wiederum erzielte das Prognosemodell für die südlichste Region mit einer Fläche $A_z = 0.77$ die besten Resultate.

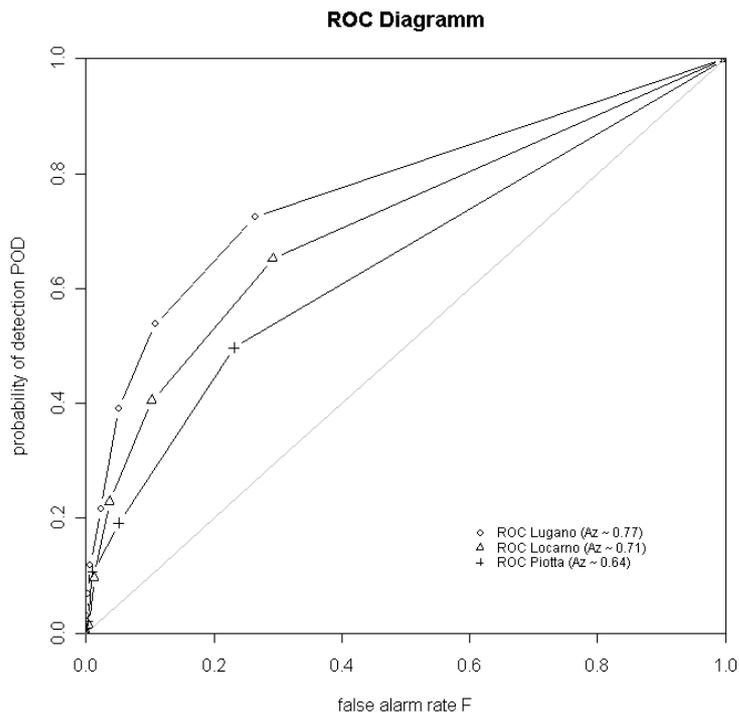


Abb. 7.5: ROC-Diagramm für die Prognosen der drei Regionen Lugano, Locarno und Piotta (1991-2002).

7.3 Probabilistische Prognose (PF)

Mit der probabilistischen Prognose (PF) wird eine Wahrscheinlichkeit p_y für das Eintreten eines Ereignisses an einem zu prognostizierenden Tag y berechnet (Kap. 4.3).

Mit dem *Brier Score* (BS), dem am häufigsten verwendeten Qualitätsmass probabilistischer Prognosen, sind für die drei Untersuchungsgebiete folgende Werte berechnet worden:

7. RESULTATE

- Lugano: $BS = 0.093$
- Locarno: $BS = 0.095$
- Piotta: $BS = 0.095$

Auch hier zeigt sich, dass das k NN-Modell für die südlichste Region das beste Resultat liefert. Mit den Absolutwerten des BS alleine kann aber prinzipiell keine ausreichende Aussage bezüglich der Modellqualität gemacht werden. Ein Qualitätsvergleich mit einem anderen Prognosemodell ist nur dann zulässig, wenn für beide Modelle der gleiche Datensatz getestet wird. Dies wurde im vorliegenden Fall mit dem logistischen Regressionsmodell (LR) exemplarisch durchgeführt. Für denselben Datensatz (Region Lugano) wurde mit dem logistischen Prognosemodell ein BS -Wert von 0.14 erreicht. Die Berechnung der Waldbrandgefahr mit dem k NN-Modell führte somit zu besseren Ergebnissen.

7.3.1 Zuverlässigkeitsdiagramm (attributes diagram)

Das Zuverlässigkeitsdiagramm, welches die Terme des BS graphisch darstellt (vgl. Gl. 6.16), erlaubt einen guten Überblick und erleichtert den Vergleich zwischen unterschiedlichen Modellen.

Obwohl das k NN-Modell für die Region Lugano (Abb. 7.6) die Wahrscheinlichkeit p_y für einen Waldbrand allgemein überschätzt, folgt die Kurve der Diagonalen recht gut. Die maximale Abweichung beträgt 20%; im Durchschnitt liegt sie bei Werten um 10%, was etwa der Wahrscheinlichkeitsklassenbreite entspricht. Die Prognoseunsicherheit nimmt bei grösseren Wahrscheinlichkeiten ($p_y > 0.6$) deutlich zu, da solche Fälle zu selten auftreten (vgl. Abb. 5.8).

Für die beiden anderen Regionen ergaben die Auswertungen der Zuverlässigkeitsdiagramme (Abb. 7.7 und 7.8) ähnliche Resultate wie bereits bei der BF . Die Qualität des k NN-Modells für diese beiden Gebiete ist auch bei der PF unzureichend.

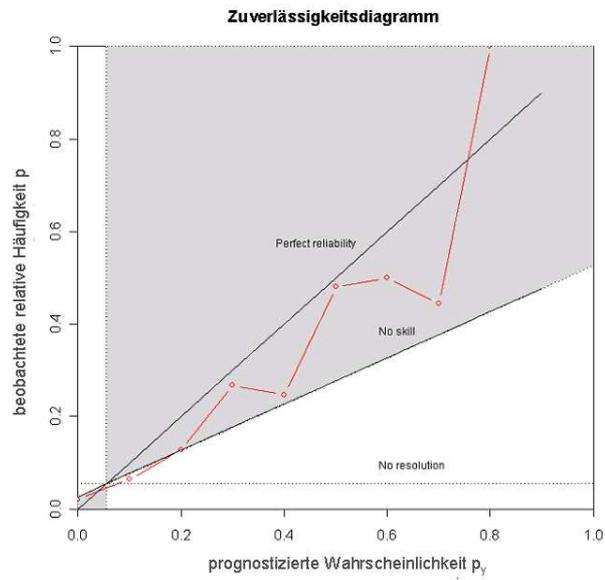


Abb. 7.6: Zuverlässigkeitsdiagramm der Region Lugano (1991-2002).

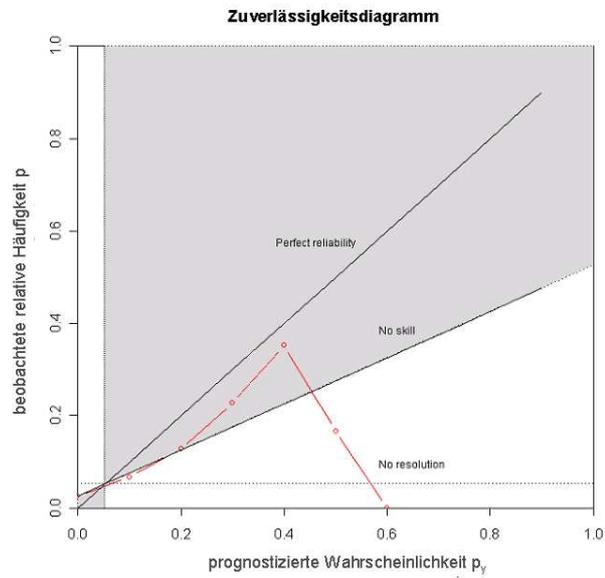


Abb. 7.7: Zuverlässigkeitsdiagramm der Region Locarno (1991-2002).

7. RESULTATE

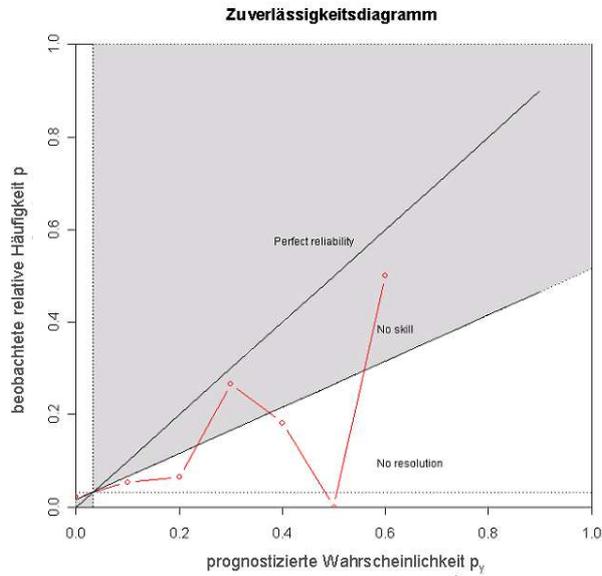


Abb. 7.8: Zuverlässigkeitsdiagramm der Region Piotta (1991-2002).

7.4 Deskriptive Prognose (DF)

Die Verifikation und somit auch die Beurteilung der Resultate gestaltet sich bei der deskriptiven Prognose (DF) schwierig. Für das im Rahmen dieser Arbeit entwickelte k NN-Waldbrand Prognosemodell liegen noch keine Informationen bezüglich der Prognosequalität von Modellanwendern vor. Trotzdem kann ein Versuch unternommen werden, den Wert der deskriptiven Prognose einzuschätzen.

Gegenüber der BF und der PF weist die DF einen grossen Vorteil auf. Dadurch, dass alle nächsten Nachbarn inklusive den Ereignissen und den dazugehörigen Detailinformationen bei jeder Nachbarnsuche aufgelistet werden (Abb. 7.1), geht keine in der Datenbank enthaltene Information verloren. Die räumliche Verteilung der aufgetretenen Brände (Abb. 5.9) kann durch einfache, statistische Auswertungen der η^+ Zusatzinformation einfach und übersichtlich dargestellt werden (Abb. 7.9). Neben der Karteninformation (vgl. Abb. 5.9) kann sich der Anwender mit Hilfe dieses Höhen- und Expositionsdiagrammes somit rasch einen Überblick über die räumliche Verteilung der vergangenen Brände verschaffen.

Die Tabelle 7.1 und Abbildung 7.9 spiegeln die Situation vom 26.03.1994 wieder.

Tab. 7.1: Output der deskriptiven Prognose (DF) des *k*NN-Waldbrand Prognosemodells für den 26.03.1994. Je mehr Informationen die Waldbranddatenbank beinhaltet, desto umfassender wird die Ausgabe der DF.

Datum	E	Betr. Fläche	Exp.	Höhe	Ursache	Waldtyp
17.3.91	1	47ha	S-SW	1070m.ü.M	Blitz	Kastanien
08.5.97	1	68ha	SW-W	620m.ü.M	Unbekannt	Kastanien
11.1.01	1	12ha	S-SW	1710m.ü.M	Brandst.	Föhren
01.2.99	0					
17.2.02	1	107ha	S-SE	280m.ü.M	Eisenbahn	Kastanien
15.5.01	0					
17.2.98	1	22ha	S-SW	1420m.ü.M	Blitz	Lärchen
18.2.98	0					
01.1.00	1	Unbekannt	SE-E	1890m.ü.M	Unbekannt	Kastanien
25.3.97	0					

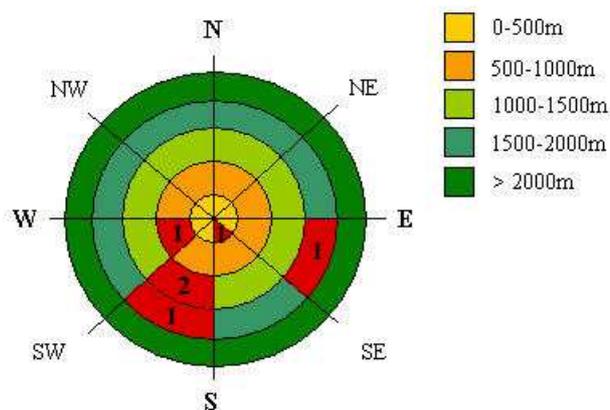


Abb. 7.9: Einfache statistische Auswertung der deskriptiven Prognose (DF) vom 26.03.1994 (vgl. Tab. 7.1). Die rot markierten Felder zeigen die durch Brände betroffenen Expositionen und Höhenlagen. Vermerkt ist zusätzlich auch die Anzahl der Brände (schwarze Ziffern).

7.5 Fallbeispiele

Im Folgenden werden für die drei Untersuchungsgebiete sowohl die berechneten Prognosen wie auch die jeweils erzielten Prognosequalitäten vorgestellt.

7. RESULTATE

Abbildung 7.10 zeigt sowohl die berechnete Anzahl positiver Nachbarn (η^+) als auch die Tage mit effektiv aufgetretenen Waldbränden für den gesamten Untersuchungszeitraum der südlichsten Region Lugano. Den berechneten Werten liegt das volle Modell (vgl. Tab. 5.6) zugrunde. Eine deutliche “Struktur” einer Häufung von Waldbränden im Winterhalbjahr ist sowohl an der Anzahl der tatsächlich aufgetretenen Brände als auch an der Anzahl der berechneten positiven Nachbarn deutlich zu erkennen. Die Abbildung 7.11 zeigt eine ähnliche Struktur mit dem Unterschied, dass den berechneten Nachbarn das reduzierte Modell (vgl. Tab. 5.7) zugrunde liegt. Dies hat zur Folge, dass die Anzahl der gefundenen Nachbarn leicht tiefer ist (volles Modell: 2429 η^+ ; reduziertes Modell: 2271 η^+). Anhand der leicht veränderten Struktur kann erkannt werden, dass nicht nur weniger, sondern auch andere nächste Nachbarn gefunden werden. Dies kann mit der unterschiedlichen Anzahl an Dimensionen des Ereignisraumes Ω und mit der unterschiedlichen Gewichtung der einzelnen Variablen erklärt werden.

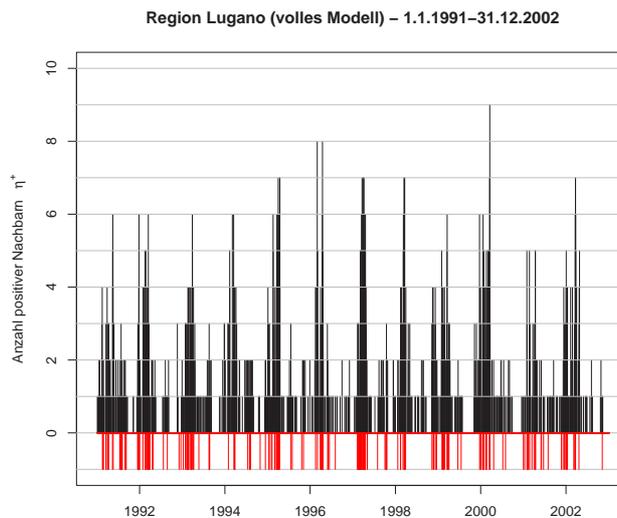


Abb. 7.10: Berechnete Anzahl positiver Nachbarn η^+ und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1991-31.12.2002 (volles Modell). Schwarz: berechnete positive Nachbarn; rot: tatsächlich aufgetretene Ereignisse.

Auch für die beiden anderen Untersuchungsgebiete Locarno und Piotta (Abb. 7.12 und 7.13) sind ähnliche Muster erkennbar. Im Gegensatz zu Abbildung 7.10 ist aber die Anzahl gefundener positiver Nachbarn η^+ kleiner (Locarno: 2224 η^+ ;

Piotta: 1465 η^+). Dies hängt allerdings auch mit der Gesamtanzahl der Waldbrände zusammen, welche in der nördlichsten Region deutlich am kleinsten ist. Betrachtet man aber das Verhältnis zwischen der Anzahl der gefundenen positiven Nachbarn η^+ und der Anzahl der effektiv aufgetretenen Waldbränden, so wird ersichtlich, dass praktisch kein Unterschied zwischen den drei Untersuchungsregionen festgestellt werden kann (vgl. Tab. 7.2).

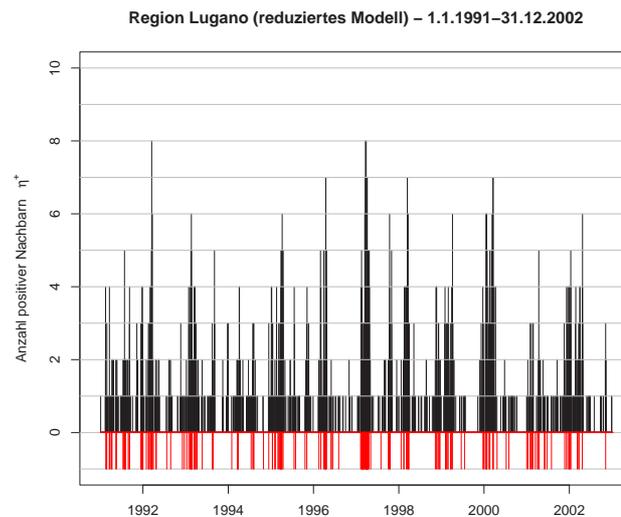


Abb. 7.11: Berechnete Anzahl positiver Nachbarn η^+ und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1991-31.12.2002 (reduziertes Modell). Schwarz: berechnete positive Nachbarn; rot: tatsächlich aufgetretene Ereignisse.

Tab. 7.2: Zusammenfassung der Anzahl aufgetretener Waldbrände und der Anzahl berechneter positiver Nachbarn η^+ in den drei Untersuchungsgebieten.

Region	Anzahl aufgetretener Brände	Anzahl η^+	Verhältnis
Lugano	236	2429 (volles Modell) 2271 (red. Modell)	0.097 0.104
Locarno	227	2224 (red. Modell)	0.102
Piotta	141	1465 (red. Modell)	0.096

7. RESULTATE

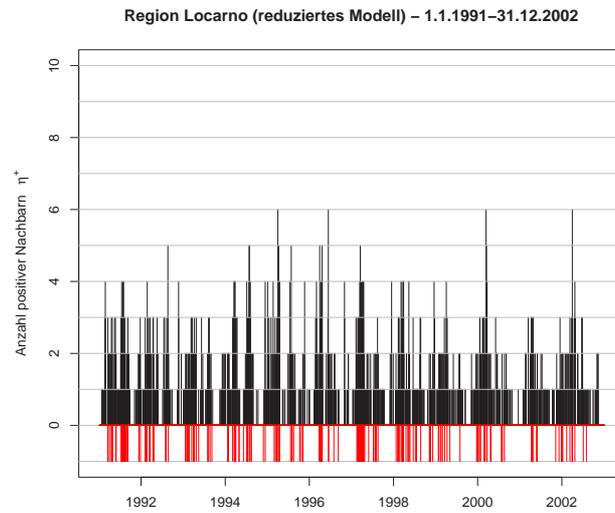


Abb. 7.12: Berechnete Anzahl positiver Nachbarn η^+ und tatsächlich aufgetretene Waldbrände der Region Locarno für den Zeitraum 01.01.1991-31.12.2002 (reduziertes Modell). Schwarz: berechnete positive Nachbarn; rot: tatsächlich aufgetretene Ereignisse.

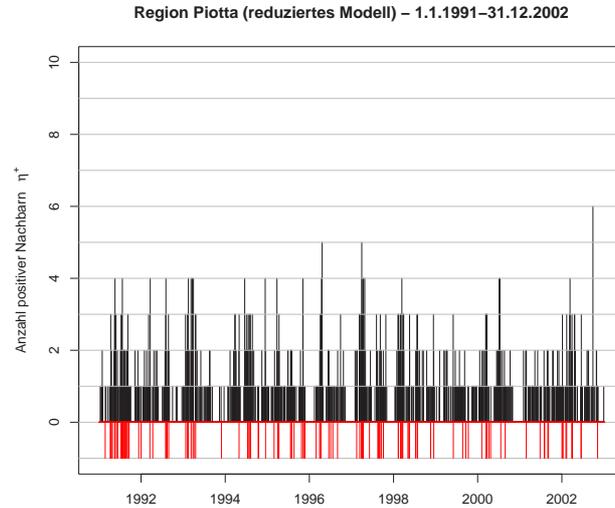


Abb. 7.13: Berechnete Anzahl positiver Nachbarn η^+ und tatsächlich aufgetretene Waldbrände der Region Piotta für den Zeitraum 01.01.1991-31.12.2002 (reduziertes Modell). Schwarz: berechnete positive Nachbarn; rot: tatsächlich aufgetretene Ereignisse.

In Kapitel 5.6.1 wurde erläutert, wie der optimale Schwellenwert (DB) der Anzahl positiver Nachbarn η^+ für die binäre Prognose bestimmt worden ist. Abbildung 7.14 zeigt nur jene η^+ an, für welche $DB \geq 3$ gilt. Im Vergleich zu Abbildung 7.10 kann erkannt werden, dass die Anzahl der “Fehlalarme” deutlich abgenommen hat; dies allerdings auch auf Kosten einiger Fälle, welche nun nicht mehr prognostiziert werden (vgl. Kap. 6.4). Weiter wird ersichtlich, dass in einem relativ kurzen Zeitraum gehäuft auftretende Brände (Winterhalbjahr) vom Modell gut prognostiziert werden können. Hingegen sind die “sporadischen”, meist im Sommerhalbjahr auftretenden Feuer schlechter prognostizierbar. Dies hängt damit zusammen, dass die längeren und vor allem stabileren winterlichen Trockenperioden vom Modell einfacher erfasst werden können. Im Sommerhalbjahr erschweren die schneller wechselnden Witterungsbedingungen eine bessere Prognose der Waldbrände.

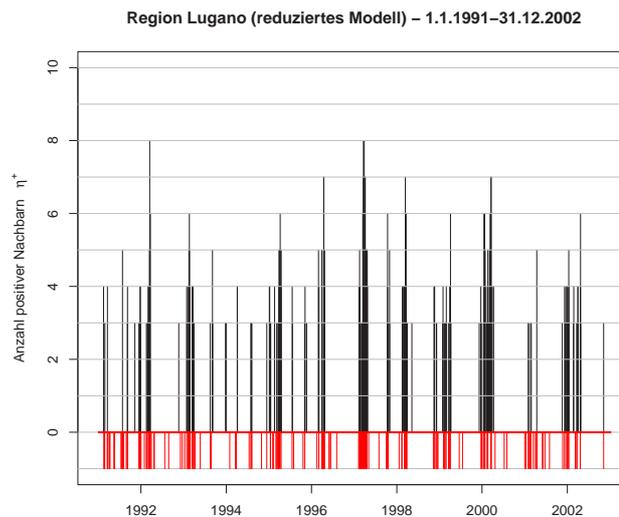


Abb. 7.14: Berechnete Anzahl positiver Nachbarn $\eta^+ \geq 3$ und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1991-31.12.2002 (reduziertes Modell). Schwarz: berechnete positive Nachbarn; rot: tatsächlich aufgetretene Ereignisse.

Für eine detailliertere Analyse der Prognose wird in der Folge anhand von zwei weiteren Beispielen ein kürzerer Zeitraum in einem der Untersuchungsgebiete genauer betrachtet.

7. RESULTATE

Abbildung 7.15 zeigt die Waldbrandverhältnisse und die entsprechenden Prognosen des Jahres 1997 für die Region Lugano. Wiederum fällt die deutliche Zweiteilung in Winter- und Sommerwaldbrände auf. Dieser Umstand konnte unter anderem bereits mit Hilfe der Signifikanz der Variable *Phänologie* (vgl. Tab. 5.6) beschrieben werden. Mit der Festlegung des Schwellenwertes $DB \geq 3$ (vgl. Abb. 7.16) kann die Prognosequalität in diesem Fall deutlich verbessert werden. Dies ist darauf zurückzuführen, dass die Brände im Jahr 1997 grösstenteils bei “eindeutig” potentiell gefährlichen Situationen aufgetreten sind. Der bereits erwähnte Umstand, dass gehäuft auftretende Brände besser prognostiziert werden können, wird in den beiden Abbildungen 7.15 und 7.16 wiederum ersichtlich.

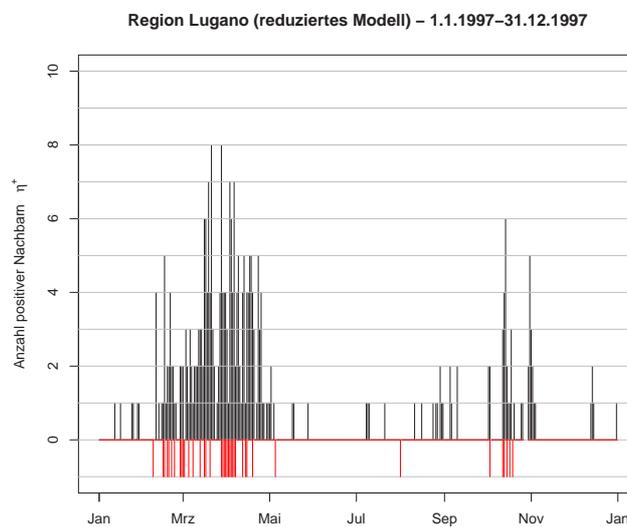


Abb. 7.15: Berechnete Anzahl positiver Nachbarn und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1997-31.12.1997 (reduziertes Modell). Schwarz: berechnete positive Nachbarn; rot: tatsächlich aufgetretene Ereignisse.

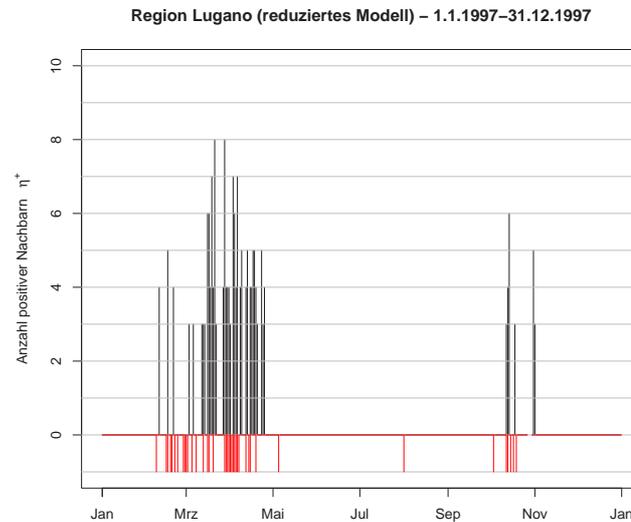


Abb. 7.16: Berechnete Anzahl positiver Nachbarn $\eta^+ \geq 3$ und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1997–31.12.1997 (reduziertes Modell). Schwarz: berechnete positive Nachbarn; rot: tatsächlich aufgetretene Ereignisse.

Abbildung 7.17 ermöglicht eine noch genauere Betrachtung der Waldbrandsituation im April 1997. Bis zum 19. April 1997 sind an 11 Tagen in der Region Lugano Feuer aufgetreten. Sowohl das volle, als auch das reduzierte Modell “beurteilten” die gefährliche Situation richtig, indem in beiden Fällen zumindest immer ein positiver Nachbar η^+ gefunden wurde. Die Beurteilung des Modells fällt auch bei einem Schwellenwert von $DB \geq 3$ positiv aus. Ob ein Feuer am Tag x oder am Tag $x + 1$ ausbricht, ist in solchen Situationen häufig zufällig; entscheidend ist, dass das Potential für ein Feuer vorhanden ist und dies vom Modell auch als solches erkannt wird. Unzureichend ist die Prognosequalität für das letzte Drittel des Monats April 1997. Eine Erklärung für die Waldbrandsituation im April 1997 und die Veränderung nach dem 19. April kann in der damals vorherrschenden meteorologischen Situation gefunden werden. Der betrachtete Zeitraum war durch extreme Trockenheit geprägt. Am Monatsbeginn lag die 30-Tagesniederschlagssumme bei 1.6mm. Es überwog hauptsächlich Sonnenschein bei einer mittleren Temperatur von rund 17°C .

7. RESULTATE

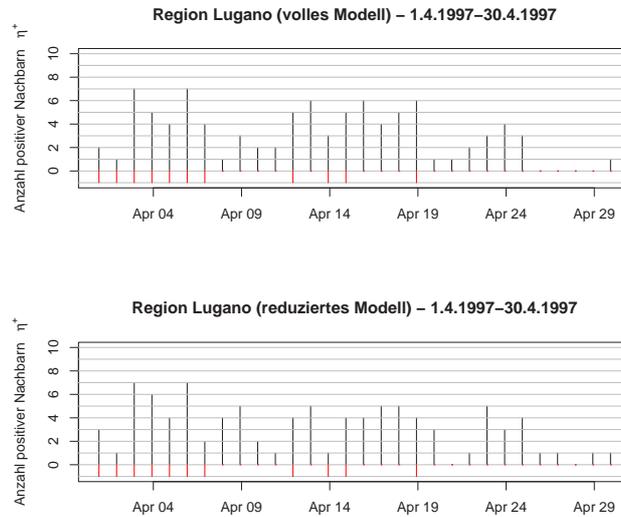


Abb. 7.17: Berechnete Anzahl positiver Nachbarn und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.04.1997-30.04.1997 (volles und reduziertes Modell). Schwarz: berechnete positive Nachbarn; rot: tatsächlich aufgetretene Ereignisse.

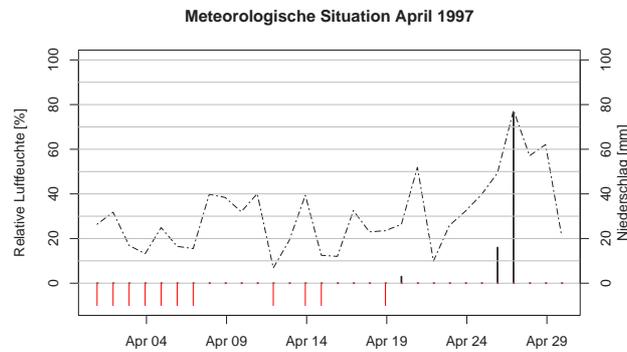


Abb. 7.18: Meteorologische Situation an der Station Lugano im April 1997. Neben den Waldbränden ist der Verlauf der relativen Luftfeuchte [%] und der gefallen Niederschläge [mm] eingezeichnet.

In Abbildung 7.18 sind mit der relativen Luftfeuchte und den Niederschlagssummen die meteorologischen Variablen dargestellt, die eine Veränderung der potentiellen Waldbrandgefährdung zumindest teilweise erklären können. Eine Zunahme der relativen Luftfeuchte nach dem 19. April führte ebenso wie die drei Niederschlagsereignisse zu einer langsamen Entspannung der kritischen Situation.

7.6 Sensitivitätsanalyse

Mit einer Sensitivitätsanalyse soll der Einfluss der Variabilität einzelner Parameter auf das Modellergebnis untersucht werden. Bezüglich der k NN-Waldbrandprognosemodellergebnisse ist die Betrachtung folgender variabler Größen von Interesse:

- (a) Sensitivitätsanalyse bezüglich der Anzahl meteorologischer Variablen: Wie wirkt sich die Veränderung der Anzahl Variablen auf das Waldbrandprognosemodell aus (Vergleich reduziertes versus volles Modell)?
- (b) Sensitivitätsanalyse bezüglich der Prognoseunsicherheit der einzelnen meteorologischen Parameter. Da die Modelleingangsgrößen prognostizierte meteorologische Werte sind, sind sie einer Prognoseunsicherheit unterworfen. Welchen Einfluss hat diese Unsicherheit auf die gefundenen nächsten Nachbarn?
- (c) Sensitivitätsanalyse bezüglich räumlichen und zeitlichen Veränderungen: Wie wirkt sich eine Veränderung der Gebietsgröße, respektive des beobachteten Zeitraums, auf die Modellergebnisse aus?
- (d) Sensitivitätsanalyse bezüglich der gewählten Anzahl Nachbarn k .

Zu (a): Die Grundlagen einer ersten Selektion der meteorologischen Variablen, basierend auf empirischen Annahmen und nach Signifikanzberechnungen mittels der logistischen Regression, wurden bereits in den Kapiteln 5.4.1 und 5.4.2 erläutert. Basierend auf diesen zuvor erarbeiteten Grundlagen, konnte in Kapitel 7.5 gezeigt werden, dass sowohl unterschiedliche Nachbarn, als auch eine unterschiedliche Anzahl dieser gefunden werden, wenn das volle und das reduzierte Modell eingesetzt werden. Die beiden Abbildungen 7.19 und 7.20 zeigen diese Unterschiede in den gefundenen positiven Nachbarn η^+ für den gesamten Untersuchungszeitraum, respektive im Detail für das Jahr 1997. Beide Abbildungen

7. RESULTATE

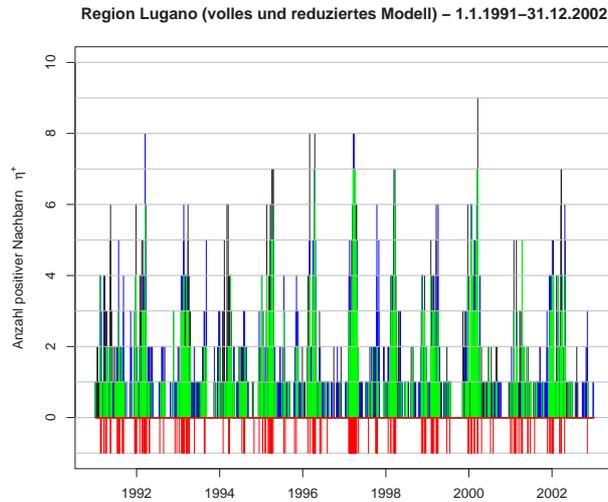


Abb. 7.19: Berechnete Anzahl positiver Nachbarn η^+ und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1991-31.12.2003. Schwarz: Nachbarn des vollen Modells; blau: Nachbarn des reduzierten Modells; grün: “gemeinsame” Nachbarn, rot: tatsächliche Ereignisse.

zeigen, dass bei $\eta^+ = 1$ nur selten beide Modelle (grün) dieselben Tage finden. Ist die positive Anzahl an gefundenen Nachbarn aber grösser, werden oft dieselben Tage von beiden Modellen gefunden. Aufgrund der Tatsache, dass beispielsweise bei der binären Prognose nur die Regel $\eta^+ \geq 3$ oder $\eta^+ < 3$ zählt, können diese Fälle schlussendlich oft vernachlässigt werden. Es lässt sich zusammenfassend sagen, dass, obwohl beim vollen und beim reduzierten Modell nicht immer dieselbe Anzahl und nicht dieselben positiven Nachbarn gefunden werden, der Einfluss auf die Prognosequalität meist gering ist.

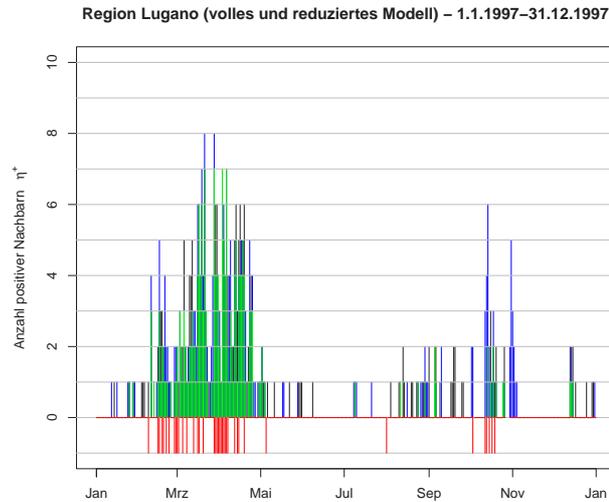


Abb. 7.20: Berechnete Anzahl positiver Nachbarn η^+ und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1997-31.12.1997. Schwarz: Nachbarn des vollen Modells; blau: Nachbarn des reduzierten Modells; grün: “gemeinsame” Nachbarn; rot: tatsächliche Ereignisse.

Zu **(b)**: Um den Einfluss der Prognoseunsicherheit der einzelnen meteorologischen Parameter bezüglich der Waldbrandvorhersage abzuschätzen, gibt es zwei Möglichkeiten:

- Die Monte-Carlo-Methode, welche allerdings eine sehr zeit- und rechenaufwändige Möglichkeit darstellt. Bei jeder Variablenänderung entsprechend der jeweiligen Prognoseunsicherheit muss für den gesamten Zeitraum die Waldbrandprognose neu gerechnet und verifiziert werden, denn nur so kann die Sensitivität einer einzelnen prognostizierten Variablen genau bestimmt werden.
- Eine rein “empirische” Abschätzung des Einflusses. Diese Methode ist bezüglich Zeit- und Rechenaufwand viel schneller und einfacher durchzuführen.

Im Folgenden werden die Ergebnisse einer Sensitivitätsanalyse, welche für das Modell mit der reduzierten Anzahl an Variablen (vgl. Tab 5.6) für die Region Lugano durchgeführt worden ist, vorgestellt. In Tabelle 7.3 werden die für die Abschätzung der Sensitivität wichtigsten Größen zusammengefasst. Der *Wir-*

7. RESULTATE

Wirkungsfaktor berechnet sich aus dem *Gewicht* (vgl. Tab 5.6) dividiert durch den betrachteten Zeitraum (z.B. Temperatur: 1-Tagesmaximum; Relative Feuchte: 1-Tagesminimum; Luftdruck: 1-Tagesmittel; Niederschlag: 30-Tagessumme, etc.). Die Prognoseunsicherheit ist ein geschätzter Wert. Die Varianz (Var), Standardabweichung (Sd), Mittelwert (Mw) und Median sind die für den Untersuchungszeitraum 01.01.1991-31.12.2002 berechneten Grössen. Die Variable *Phänologie* kann für die folgenden Untersuchungen weggelassen werden, da sie keiner Prognoseunsicherheit unterliegt.

Je kleiner der berechnete Wirkungsfaktor ist, desto geringer ist die Sensitivität des Modells auf die entsprechende Variable. Dabei wird die Prognoseunsicherheit allerdings noch nicht mitberücksichtigt. Unter Berücksichtigung der in Tabelle 7.3 aufgeführten Werte, kann anschliessend festgestellt werden, dass:

- bei der Temperatur die Prognoseunsicherheit $< Sd$ ist,
- bei der relativen Luftfeuchte die Prognoseunsicherheit $\approx Sd$ ist,
- beim Luftdruck die Prognoseunsicherheit $\approx Sd$ ist und
- bei der 30-Tagesniederschlagssumme die Prognoseunsicherheit $> Sd$ ist. (Angegebener SD-Wert auf 30-Tagessumme bezogen, Prognoseunsicherheit auf 1-Tagesprognose).

Die Kombination aus Wirkungsfaktor und dem "Verhältnis" Prognoseunsicherheit/Sd bewertet die Sensitivität des Modells auf die Veränderung einer Variablen gesamthaft. Die Sensitivität lässt sich wie folgt berechnen:

$$\text{Sensitivität } (S) = \text{Wirkungsfaktor} * \text{Prognoseunsicherheit/Sd}$$

Der grosse Wirkungsfaktor (0.83) und das ungünstige Prognoseunsicherheit/Sd-Verhältnis (1) der relativen Luftfeuchte führt dazu, dass das Waldbrandprognosemodell auf diese Grösse am sensitivsten reagiert ($S = 0.83$). Bei der Lufttemperatur wirkt sich das günstige Prognoseunsicherheit/Sd-Verhältnis positiv auf die Sensitivität ($S = 0.19$) aus. Eine ebenfalls kleine Sensitivität ($S = 0.13$) auf das Modell weist der Luftdruck auf; dies ist auf den geringen Wirkungsfaktor zurückzuführen. Beim Niederschlag ist zwar das Prognoseunsicherheit/Sd-

Verhältnis gross (5.26; Sd-Wert muss zuerst durch 30 (Anzahl Tage) dividiert werden), die Sensitivität ($S = 0.21$) auf das Modell ist aber vergleichbar mit der Lufttemperatur; dies aufgrund des sehr kleinen Wirkungsfaktors.

Zusammenfassend lässt sich also sagen, dass das Prognosemodell auf die Variable “relative Luftfeuchte” zirka vier mal sensitiver als auf die übrigen Variablen reagiert, wobei der “Luftdruck” die am wenigsten heikle Grösse darstellt.

Tab. 7.3: Zusammenfassung der wichtigsten Grössen zur Bestimmung der Sensitivität des Modells in Abhängigkeit jeder einzelnen Variablen.

Variable	Gewicht	Prognoseunsicherh.	Wirkungsfaktor	Varianz	Standardabw.	Mittelwert	Median
Temperatur	0.48	$\pm 3^\circ\text{C}$	0.48	55.6	7.6	16.7	16.6
Rel. Feuchte	0.83	$\pm 20\%$	0.83	365.5	19.1	53.8	55.1
Luftdruck	0.22	$\pm 5\text{hPa}$	0.22	55.9	7.5	982.0	982.3
Σ N'schlag	1.33	$\pm 20\text{mm}$	0.04	12983.0	114.0	137.8	112.0
Phänologie	0.61	-	-	-	-	-	-

Zu (c): Im Folgenden soll sowohl der Einfluss der Gebietsgrösse, als auch der Einfluss des beobachteten Zeitraums auf die Anzahl gefundener positiver Nachbarn η^+ untersucht werden. In Kapitel 5.2 konnte bereits aufgezeigt werden, dass die Grundwahrscheinlichkeit s einen sehr starken Einfluss haben kann. Weiter soll überprüft werden, ob neben der Grundwahrscheinlichkeit s noch weitere Faktoren die Modellergebnisse quantitativ und qualitativ beeinflussen.

In Abbildung 7.21 wird, wie bereits in Abbildung 7.15, die berechnete Anzahl positiver Nachbarn gegenüber den tatsächlich aufgetretenen Waldbränden für die Region Lugano im Zeitraum 01.01.1997-31.12.1997 dargestellt, mit dem Unter-

7. RESULTATE

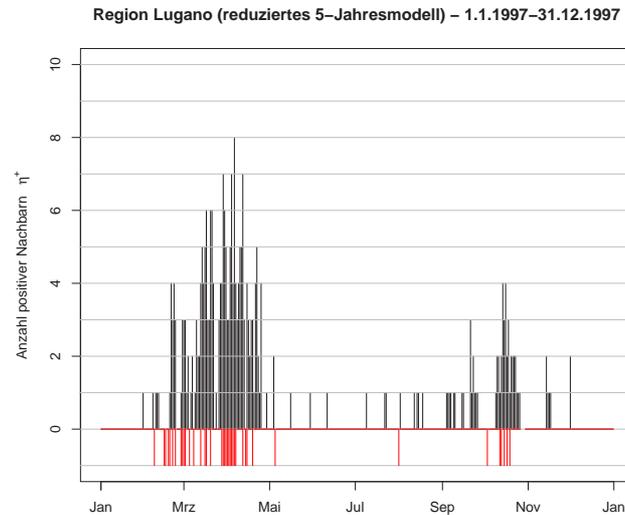


Abb. 7.21: Berechnete Anzahl positiver Nachbarn und tatsächlich aufgetretene Waldbrände der Region Lugano für den Zeitraum 01.01.1997–31.12.1997 (reduziertes Modell). Schwarz: berechnete positive Nachbarn; rot: tatsächlich aufgetretene Ereignisse. Im Gegensatz zu den bisherigen Modellberechnungen wurden in diesem Fall nur 5 Jahre mitberücksichtigt.

schied, dass dem Modell der Abbildung 7.21 nur von 5 und nicht von 12 Jahren Daten zugrunde liegen. Vergleicht man die beiden Abbildungen kann wiederum festgestellt werden, dass, obwohl nicht dieselbe Anzahl und nicht dieselben Nachbarn gefunden werden, die Prognosequalität in etwa gleich ist (12-Jahresmodell: $300 \eta^+$; 5-Jahresmodell: $287 \eta^+$). Es ist allerdings zu beachten, dass bei einem zu kurzen Beobachtungszeitraum, zum Beispiel einem Jahr, zwar immer $k=10$ nächste Nachbarn gefunden werden können, deren Ähnlichkeit zur aktuellen Situation aber nur noch sehr gering ist, weshalb sie sich für eine Prognose nur noch bedingt eignen.

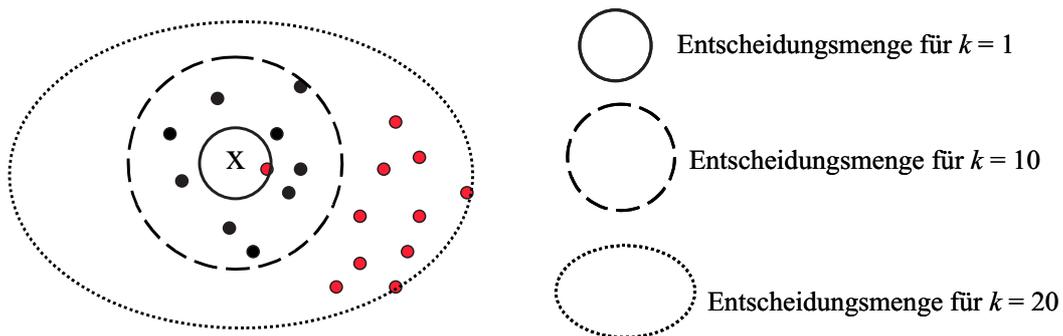
Der betrachtete Zeitraum spielt für die Prognosequalität somit eine Rolle. Gute Ergebnisse können ab einem Beobachtungszeitraum von etwa drei Jahren erwartet werden.

Zu (d): Die Wahl der Anzahl der zu betrachtenden Nachbarn k ist bereits in Kapitel 5.6 ausführlich besprochen worden. Anhand der Abbildung 7.22 soll die Problematik bezüglich der Sensitivität des Prognosemodells nochmals kurz angesprochen werden. Für $k = 1$ ist die Sensitivität des Modells gegenüber Ausreißern sehr gross, demgegenüber werden bei einem zu grossen k zu viele Objekte aus “anderen” Klassen in die Entscheidungsfindung miteinbezogen. Die “geeignestste” Anzahl nächster Nachbarn wurde auf $k = 10$ festgelegt. Die Begründung dafür lieferten einerseits die Berechnungen in Kapitel 5.6, andererseits sollte die Modellausgabe für die praktische Anwendung übersichtlich bleiben.

Wahl des Parameters k

„zu kleines“ k : hohe Sensitivität gegenüber Ausreißern

„zu großes“ k : viele Objekte aus anderen Clustern (Klassen) in der Entscheidungsmenge.



x: zu klassifizieren

Abb. 7.22: Darstellung der Problematik der Wahl von k . Die roten Punkte stammen aus einer anderen Klasse.

8 Diskussion

8.1 Kritische Betrachtung der Ergebnisse

Um die Qualität des entwickelten Waldbrandprognosemodells abschliessend beurteilen zu können, ist eine kritische Betrachtung der Ergebnisse und der Methode nötig. Dabei ist zu berücksichtigen, dass die durchgeführten Modellverifikationen vorsichtig und differenziert beurteilt werden müssen.

8.1.1 Fazit - Verifikation

Die Durchführung von Verifikationsberechnungen des Prognosemodells ist für die Einschätzung der Qualität unabdingbar. Bereits in Kapitel 6 wurde gezeigt, dass eine Vielzahl an unterschiedlichen Verifikationsmassen zur Verfügung stehen. Es ist zu berücksichtigen, dass die Berechnungen und insbesondere die Interpretationen der berechneten Werte grosser Vorsicht bedürfen.

Die Qualität einer Prognose kann nicht durch ein universelles Mass ausgedrückt werden. Zur abschliessenden Beurteilung ist somit immer eine geeignete Auswahl an Gütemassen in Kombination zu verwenden. Die Vielzahl an Verifikationsmassen ist vor allem bei der binären Prognose (BF) gross.

Die Kombination *POD* und *FAR* stellt ein zweckmässiges, rasch und einfach überblickbares Gütemass dar. Um ein abschliessendes Urteil bezüglich der Prognosequalität zu erhalten, muss aber mindestens ein weiteres Mass in die Beurteilung miteinbezogen werden.

Ein grosses Problem bei einer Vielzahl von Verifikationsmassen ist ihre Abhängigkeit von der Grundwahrscheinlichkeit s . Dieser Umstand schränkt die Aussagekraft von *CSI*, *FAR* oder auch *HSS* dahingehend ein, dass keine objektiven Aussagen bezüglich der Modellqualität gemacht werden können. Die Qualität von unterschiedlichen Prognosemodellen kann mit diesen Massen somit nur dann verglichen werden, wenn derselbe Datensatz untersucht wird.

8. DISKUSSION

Mit der Einführung der *signal detection theory* (Mason, 1980) wurde ein neues, robusteres Verifikationsmass eingeführt. Der Nachteil dieser Grösse ist, dass sie normalverteilte Variablen voraussetzt, was im vorliegenden Fall nicht im vollen Umfang gegeben ist.

Um die Qualität der probabilistischen Prognose (PF) zu überprüfen, stehen, basierend auf den *distribution-oriented* Ansätzen, weitere Gütemasse zur Verfügung. Der *Brier Score* (BS) und das Zuverlässigkeitsdiagramm (graphische Darstellung seiner Komponenten) sind die am weitesten verbreiteten Masse. Mit dem *Brier Score*, welcher eine Aussage bezüglich *reliability*, *resolution* und *uncertainty* beinhaltet, kann anhand eines einzelnen Wertes die Prognosequalität eines Modells bestimmt und in Zusammenhang mit anderen *BS*-Werten auch interpretiert werden. Wie bei den Verifikationsmassen der *BF* ist ein Qualitätsvergleich von unterschiedlichen Modellen mit dem *BS* aber nur bei der Berechnung mit demselben Datensatz zulässig.

Eine objektive Verifikation der deskriptiven Prognose (DF) ist nicht möglich. Die Interpretation der Resultate ist sehr stark subjektiv geprägt und deshalb kaum verifizierbar.

8.1.2 Fazit - binäre Prognose

Ausgehend von den in Kapitel 6.4 erläuterten Verifikationsmassen, wurden zusätzlich zu den in Kapitel 7.2 präsentierten Ergebnissen weitere Modellauswertungen für die binäre Prognose (BF) vorgenommen. Die Resultate bewegen sich alle in derselben Grössenordnung, sodass für die binäre Prognose folgendes Fazit gezogen werden kann:

Die *BF* stellt eine einfache und übersichtliche Ausgabeform der *k*NN-Methode dar. Da per Definition (Kap. 4.3) nur ein Ja-Nein-Entscheid zugelassen wird, ist eine klare Aussage bezüglich der aktuellen Waldbrandgefahr möglich. Dem Anwender wird weder eine Möglichkeit einer Interpretation, noch eine Angabe bezüglich der Prognoseunsicherheit gegeben.

Die Prognosequalität ist für alle drei Regionen, insbesondere für die nördlichste Region (Piotta), ungenügend. Die *BF* als alleinige Ausgabeform reicht somit

nicht zur Prognose der aktuellen Waldbrandgefahr aus. Die Gründe dafür liegen hauptsächlich in der sehr niedrigen Grundwahrscheinlichkeit s , in der Komplexität der Topographie und der zu erklärenden Prozesse.

Obwohl die Qualität der binären Prognose somit zum Teil grossen Vorbehalten unterliegt, könnte es beispielsweise Sinn machen, bei bereits einem gefundenen positiven Nachbarn die Bevölkerung auf ein gewisses, erhöhtes Gefährdungspotential aufmerksam zu machen. Basierend auf der Anzahl der gefundenen positiven Nachbarn η^+ , könnte auf einer international existierenden 5-stufigen Gefahrenskala die aktuelle Gefährdung bestimmt werden. Da “hinter” einem gefundenen η^+ immer ein “tatsächlicher” Waldbrand steht und diese Information der Datenbank entnommen werden kann, können zusätzlich zur Gefahrenskala auch Informationen bezüglich potentiell gefährdeter Expositionen, Höhenlagen, Waldtypen, etc. angegeben werden (vgl. deskriptive Prognose (DF)).

8.1.3 Fazit - probabilistische Prognose

Die Auswertungen der probabilistischen Prognose (PF) des k NN-Modells für die drei Untersuchungsgebiete ergaben analoge Resultate zu den binären Prognosen.

Das k NN-Modell liefert für die südlichste Region brauchbare bis gute Ergebnisse, währenddem die Resultate für die beiden anderen Regionen unbefriedigend sind. Dies konnte sowohl mit dem *Brier Score*, als auch mit der graphischen Auswertung der Zuverlässigkeitsdiagramme gezeigt werden. Die Ursache für die unbefriedigenden Ergebnisse in den beiden nördlichen Regionen sind vorwiegend in der Seltenheit der Ereignisse und der zunehmend komplexen Topographie zu finden.

Die probabilistische Prognose (PF) kann dann hilfreich sein, wenn sie in einen geeigneten Zusammenhang gestellt wird. Kann aber beispielsweise das Ereignis nicht genügend gut definiert werden (Lage oder Zeitpunkt eines Brandausbruches), ist die Interpretation von Wahrscheinlichkeitswerten schwierig und nicht mehr objektiv. Wird versucht, das Ereignis genauer zu definieren, tritt verstärkt die Problematik der tiefen Grundwahrscheinlichkeit s auf.

Im Gegensatz zur deskriptiven und bedingt auch zur binären Prognose kann zwischen der berechneten Wahrscheinlichkeit y_p und tatsächlich aufgetretenen Ereignissen kein Zusammenhang mehr hergestellt werden. Dieser Verlust an In-

8. DISKUSSION

formation schränkt die Aussagekraft der probabilistischen Prognose stark ein. Dennoch wäre es grundsätzlich einfach möglich, die berechneten Wahrscheinlichkeitswerte in eine Gefahrenstufe, respektive in ein Gefährdungsindex, umzuwandeln. Bei den meisten aktuell eingesetzten Prognosemodellen ist dies auch der Fall (Viegas et al., 1999b).

8.1.4 Fazit - deskriptive Prognose

Wie in Kapitel 6.6 bereits gezeigt werden konnte, ist eine Verifikation der deskriptiven Prognosen (DF) und somit der Beurteilung der Modellqualität nur sehr schwer durchführbar und bedingt einige Jahre an Modelleinsatz.

Anhand der Auswertungen der deskriptiven Prognose im Bereich der Lawinprognose konnte aber gezeigt werden, dass diese Ausgabeform meist zu guten Ergebnissen führt (Kap. 6.3) (Heierli et al., 2003). Der Vorteil der *DF* liegt darin, dass im Gegensatz zur *BF* oder *PF* kein Informationsverlust stattfindet. Die gesamte in den Grunddatenbanken enthaltene Information wird dem Anwender zur Verfügung gestellt. In Kombination mit der binären Prognose kann diese zusätzlich "gewonnene" Information sehr wertvoll sein; so beispielsweise bei der zuvor erwähnten Einteilung in Gefahrenstufen, wo neben den Stufen dann noch detailliertere Informationen bezüglich vergangener Ereignisse angegeben werden können (vgl. Abb. 7.9).

Um die Qualität des *k*NN-Waldbrandmodells - und somit den direkten Nutzen - definitiv beurteilen zu können, muss eine langfristige, durch den Anwender vorgenommene, empirische Beurteilung des Modells durchgeführt werden. Dies bedingt, dass das Modell zunächst über mindestens zwei Jahre täglich eingesetzt wird. Gegenüber anderen Prognosemodellen, darunter auch der binären und probabilistischen Prognose, stellt dies einen Nachteil dar.

Sind die Ergebnisse gut und ergibt das *k*NN-Modell, insbesondere die deskriptive Prognose (DF), einen Nutzen, lohnt sich dieser Verifikationsaufwand.

Zusammenfassend lässt sich sagen, dass:

- die in das Prognosemodell integrierten statistischen Verifikationsmasse die Modellqualität der binären und probabilistischen Prognose gut beschreiben können,
- die berechneten *BF*- und *PF*-Waldbrandprognosen, ausgenommen für die Region Lugano, zu unzureichenden Ergebnissen führen,
- diese unzureichende Modellqualität in diesen beiden Regionen hauptsächlich auf die tiefe Grundwahrscheinlichkeit s und die komplexe Topographie des Geländes zurückzuführen ist, und
- die deskriptive Prognose trotz fehlender Verifikation für die Abschätzung der aktuellen Waldbrandsituation sehr wichtige und nützliche Informationen liefert.

8.1.5 Fazit - Grenzen und Möglichkeiten der k NN-Methode

Mit Hilfe der k NN-Methode ist ein in der Waldbrandforschung neuartiges Modell zur Prognose der Waldbrandgefahr entwickelt worden, welches sowohl den Anforderungen der Wissenschaft, als auch der Praxis genügt. Wie alle Methoden weist auch diese sowohl Vor- als auch Nachteile auf:

Ein grosser Vorteil der k NN-Methode liegt darin, dass in Zukunft ein Prognosemodell zur Verfügung stehen wird, welches in Form der *DF* neben den “klassischen”, herkömmlichen Wahrscheinlichkeitswerten oder Indices weitere Informationen bezüglich der aktuellen Gefährdung beinhaltet.

Wird es in Zukunft möglich sein, das Modell in Regionen mit höheren Grundwahrscheinlichkeitswerten s und einfacherer Topographie einzusetzen, sind auch bessere Prognoseergebnisse zu erwarten.

Die Basis für eine räumliche und allenfalls auch thematische Ausweitung des k NN-Modells ist im Rahmen dieser Arbeit geschaffen worden. Weiter wurden auch die Grundlagen erarbeitet, welche eine relativ rasche und einfache Abschätzung bezüglich einer sinnvollen Modellentwicklung zur Prognose weiterer Naturgefahren zulassen.

8. DISKUSSION

Ein Nachteil der Methodik liegt darin, dass für den Neueinsatz eines Modells Grunddatensätze (meteorologischer- und Waldbranddatensatz) von mindestens drei Jahren vorhanden sein müssen. Weiter wird vorausgesetzt, dass diese Daten homogen und keinem zeitlichen Trend unterworfen sind. Diese “Vorlaufzeit” wird bei anderen, beispielsweise deterministischen Modellen, nicht vorausgesetzt. Im Gegensatz zu diesen setzt das k NN-Modell jedoch keine Kenntnisse der physikalischen Prozesse voraus. Kennt man die Physik der ablaufenden Prozesse ausreichend, können mit deterministischen Modellen eindeutig bessere Ergebnisse erzielt werden.

Diverse durchgeführte Sensitivitätsanalysen haben gezeigt, dass das Modell auf Veränderungen der unterschiedlichen Parameter verschieden reagiert.

Eine Veränderung der Anzahl berücksichtigter Variablen wirkt sich nur beschränkt auf die Prognosequalität des Modells aus, dies obwohl beim vollen und reduzierten Modell nicht immer dieselbe Anzahl und nicht dieselben positiven Nachbarn η^+ gefunden werden. Je grösser die Anzahl der betrachteten η^+ ist, desto “stabiler” reagiert das Modell auf die Veränderung der Anzahl Variablen. Dies bedeutet, dass eindeutig gefährliche Situationen mit Hilfe der berücksichtigten Variablen vom Modell gut erkannt und auch prognostiziert werden können.

Anhand von empirischen Abschätzungen konnte auch der Einfluss der Prognoseunsicherheit der meteorologischen Variablen bestimmt werden. Die Kombination aus Wirkungsfaktor und dem Prognoseunsicherheit/Sd-Verhältnis stellte dabei die Grundlage der Sensitivitätsberechnungen dar. Dabei konnte aufgezeigt werden, dass das Modell vor allem auf die relative Luftfeuchte sensitiv reagiert.

Weiter beeinflussen sowohl die Gebietsgrösse und somit die darin “enthaltene” Grundwahrscheinlichkeit s , sowie der zur Nachbarsuche zur Verfügung stehende Zeitraum die Qualität der Prognose. Ist s zu klein ($< \text{ca. } 5\%$), stösst das statistische Modell an seine Grenzen. Beim betrachteten Zeitraum ist wichtig, dass die Variablen keinem zeitlichen Trend unterliegen; trotzdem sollten Messreihen von mindestens drei Jahren berücksichtigt werden.

Die Wahl der optimalen Anzahl zu berücksichtigender Nachbarn k kann numerisch bestimmt werden (vgl. Kap. 5.6). Damit die Modellausgabe weiter übersichtlich bleibt, wurde im vorliegenden Fall $k = 10$ gewählt.

Das k NN-Modell soll in naher Zukunft für die drei Regionen im Tessin ope-

rationell eingesetzt werden. Basierend auf einem von Purves et al. (2002) entwickelten und vom Autor adaptierten Programm (*Cornice*)⁵, können ab dem Sommer 2005 täglich Waldbrandprognosen durch den operationellen Dienst der zuständigen Behörden vorgenommen werden. Die Ergebnisse müssen nach einer zirka zweijährigen Einsatzzeit analysiert werden. Darauf basierend, kann der Entschluss einer Fortführung oder eines Abbruchs des Projektes getroffen werden.

8.2 Vergleich mit anderen Modellen

In der Vergangenheit wurde eine Vielzahl von diversen Waldbrandprognosemodellen entwickelt. Die Prognosemodelle werden teilweise von staatlichen Wetterdiensten (z.B. Deutschland: M-68 Modell (DWD, 2005)), teilweise von Universitäten (Universität Zürich: Prognosemodell für den Kanton Graubünden (Allgöwer, 2001)) oder (halb)privaten Unternehmen und Organisationen (USDA, Forest Service USA: diverse Modelle und Monitoring Systeme (USDA, 2005)) entwickelt und betrieben. In Kapitel 3 (Tab. 3.1) ist eine Übersicht über einige wichtige, teilweise im operationellen Einsatz stehende Prognosemodelle gegeben. Ein umfassender und aktueller Überblick über diverse in der Waldbrandforschung entwickelte Produkte, Organisationen, Konferenzen und vieles mehr ist unter GFMC (2005) zu finden.

Es ist sehr schwierig, die Qualität der unterschiedlichen Modelle miteinander zu vergleichen. Einerseits gibt es sehr viele im Einsatz stehende Modelle, andererseits sind sie aufgrund der verschiedenen Anforderungen, welche sie erfüllen sollten, oftmals unterschiedlich konzipiert. Weiter gilt ebenfalls zu berücksichtigen, dass eine Vielzahl entwickelter Modelle nicht über den wissenschaftlichen Aspekt “hinauskommen” und nie in der Praxis eingesetzt werden.

Allgemein lässt sich sagen, dass die meisten empirischen Prognosemodelle, welche in Südeuropa im Einsatz sind, auf dem Kanadischen-Feuer-Wetter Index Ansatz (Van Wagner, 1987) basieren. Oftmals unterscheiden sich diese Modelle nur in der Gewichtung der Variablen, weiteren kleineren Nuancen oder den unterschiedlichen Darstellungsformen und Optionen.

⁵Von R. Purves und K. Morris entwickeltes und in der regionalen Lawinenprognose in Schottland eingesetztes *k*NN-Modell (Purves et al., 2002)

8. DISKUSSION

Teilweise sind diese Modelle mit einem Feuerausbreitungsmodul gekoppelt und können somit Aussagen bezüglich einer allfälligen Ausbreitungsrichtung und -geschwindigkeit machen.

In den von Waldbränden stark betroffenen Regionen (Südfrankreich, Portugal, etc.) ist die Feuererkennung und die Ausbreitungsmodellierung wichtiger als das Erkennen einer potentiell gefährlichen Situation, zumal diese im Sommer oft klar erkennbar und teilweise über lange Zeiträume gegeben ist. Der Einsatz eines Modells zur Abschätzung der aktuellen Waldbrandgefährdung (z.B. k NN-Modell) macht in diesen Gebieten in den entsprechenden Zeiten somit nur beschränkt Sinn.

Eine abschliessende Beurteilung über die Qualität einzelner Modelle darf nur dann gemacht werden, wenn sie für dasselbe Gebiet und denselben Zeitraum miteinander verglichen werden. Faktoren wie Datenbeschaffung, Unterhaltskosten, Zeitaufwand, Personal, etc. müssen dann ebenfalls mitberücksichtigt und beurteilt werden.

8.3 Bedeutung der Ergebnisse in der Waldbrandforschung

Mit der vorliegenden Arbeit wurde ein neues Modell zur Prognose der aktuellen Waldbrandgefahr entwickelt. Im Gegensatz zu vielen, den FWI-Indices (Van Wagner, 1987) zugrundeliegenden Modellen, basiert das k NN-Modell auf einem “eigenständigen”, statistischen Ansatz. Die drei unterschiedlichen Ausgabeformen des Prognosemodells sind dabei im Vergleich mit den “herkömmlichen” Prognosen neuartig. Trotz einer Einschätzung der Situation nach objektiven Kriterien (Euklid’sches Abstandsmass) verlangt die Modellinterpretation grosses Wissen bezüglich Waldbränden. Subjektive Kriterien des Experten spielen in der Gefahrenbeurteilung weiterhin eine sehr grosse Rolle und können nicht vollends durch berechnete Grössen ersetzt werden.

Die “Simplizität” des Modells lässt eine einfache räumliche Ausweitung zu, sodass das Modell schnell und günstig in anderen Waldbrandregionen eingesetzt werden kann.

Die Modellverifikationen haben gezeigt, dass die Prognosequalität in der südlich-

sten Region des Untersuchungsgebiets am höchsten ist. Dieser Umstand ist damit zu erklären, dass einerseits die Grundwahrscheinlichkeit s am höchsten und andererseits die Topographie weniger komplex und somit die räumliche Extrapolation der meteorologischen Variablen besser ist. Ein Modelleinsatz in einem Gebiet mit einer einfachen Topographie (z.B. Region Aquitaine (F)) lässt somit noch bessere Ergebnisse erwarten. Es wäre somit wünschenswert, das Modell auch in anderen Waldbrandgebieten einzusetzen und die Qualität des Modells auch dort zu überprüfen.

8.4 Künftige Modellentwicklung

Die mit einer möglichen Klimaveränderung erwarteten Veränderungen der meteorologischen Parameter (insbesondere die Kopplung von höheren Lufttemperaturen mit abnehmenden Niederschlägen im Sommer) sind zweifellos mit erheblichen ökologischen und sozioökonomischen Auswirkungen verbunden. Dieses mögliche Szenario würde ebenfalls zu einer Veränderung des Feuerregimes führen.

Eine Zunahme an potentiell gefährlichen Situationen und somit der Bedarf an einem Prognosemodell würde verstärkt auftreten. Obwohl das k NN-Modell zeitlich homogene Messreihen voraussetzt, kann eine langsame Veränderung dahingehend berücksichtigt werden, dass der Untersuchungszeitraum mitangepasst und allenfalls verkürzt wird.

Eine Kombination mit weiteren, aus der Fernerkundung gewonnenen Daten könnte zu einer deutlichen Qualitätsverbesserung des Prognosemodells führen. So könnte beispielsweise die Einführung neuer Variablen wie der Bodenfeuchte oder des NDVI's (Normalized Difference Vegetation Index) die Situation des vorhandenen Brandgutes besser beschreiben.

Eine Verknüpfung mit einem GIS-System könnte die Darstellungsmöglichkeiten des Outputs und weitere, einfache räumliche Analysen erleichtern.

Für die Zukunft wird eine Web basierte Version des k NN-Modells angestrebt. Basierend auf einer umfassenden Naturgefahrenbank (Lawinen, Waldbrand, Murgang, etc.), könnte das Modell für verschiedene Fragestellungen eingesetzt werden. Es wäre auch möglich, basierend auf vergangenen Ereignissen, Gefahrenkarten neu zu berechnen.

8. DISKUSSION

Eine Kombination der unterschiedlichen potentiellen Gefahren ist im Sinne eines integralen Risikomanagements für Naturgefahren unumgänglich und könnte durch den Einsatz eines gemeinsamen k NN-Modells ermöglicht, respektive erleichtert werden.

9 Schlussfolgerungen und Perspektiven

Ziel dieser Arbeit war die Entwicklung eines Modells zur Prognose der aktuellen Waldbrandgefahr auf der schweizerischen Alpensüdseite (Kanton Tessin). Dabei galt zu beachten, dass das Prognosemodell sowohl den wissenschaftlichen als auch den praxisrelevanten Anforderungen genügen sollte.

Die Aufgabenstellung erforderte somit zuerst eine detaillierte Betrachtung des Phänomens Waldbrand und seiner Entstehungsfaktoren. Dabei konnte beispielsweise erkannt werden, dass die Entwicklung der Waldbrände im Verlaufe der Zeit hauptsächlich durch sozioökonomische Einflüsse gesteuert werden. Weiter fiel auf, dass die Anzahl der Brände nach der Einführung von gesetzlichen Massnahmen ab 1987 deutlich abnahm. Anthropogene Einwirkungen dieser Art auf das natürliche System mussten bei der Entwicklung des Prognosemodells berücksichtigt werden.

Das Ziel der Modellierung war die Voraussage von komplexen und teilweise unverständlichen Prozessen und deren Konsequenzen. Anhand der Anforderungen an das zu entwickelnde Waldbrandprognosemodell und unter Berücksichtigung von diversen Einschränkungen musste eine Modellwahl vorgenommen werden.

Basierend auf einem k -Nearest Neighbors (k NN) Ansatz, welcher bereits erfolgreich im Bereich der Lawinenprognose eingesetzt wird, konnte für die Naturgefahr Waldbrand ein Prognosemodell entwickelt werden, welches die in Kapitel 1.2 geforderten Bedingungen erfüllte.

Das k NN-Modell ermöglicht es, für einen bestimmten Tag \mathbf{y} drei unterschiedliche Ausgabeformen von Prognosen zu erstellen:

- bei der binären Prognose (BF) wird in Abhängigkeit eines Schwellenwertes (DB) eine dichotome Entscheidung gefällt. Dieser Ja-Nein-Entscheid stellt den einfachsten der möglichen Fälle dar;

9. SCHLUSSFOLGERUNGEN UND PERSPEKTIVEN

- bei der probabilistischen Prognose (PF) wird die Wahrscheinlichkeit p_y für das Eintreten eines Ereignisses berechnet, dies allerdings auf Kosten eines grossen Anteils an Informationen;
- bei der deskriptiven Prognose (DF) wird die gesamte Information, welche in einer Ereignisdatenbank über k Nachbarn von \mathbf{y} enthalten ist, ausgegeben.

Zusammenfassend kann festgehalten werden, dass, obwohl die Qualität der binären Prognose in gewissen Untersuchungsgebieten unzureichend war, mit ihrer Hilfe auf ein erhöhtes Gefährdungspotential aufmerksam gemacht werden kann. Auch die probabilistische Prognose (PF) genügt den gestellten Anforderungen an das Prognosemodell nicht. Diese Schwächen liegen allerdings nicht in der k NN-Methode selbst, sondern sind ein allgemeines Problem statistischer Modelle. Eine Kombination niedriger Grundwahrscheinlichkeitswerte s und komplexer physikalischer Prozesse ergeben meist unbefriedigende Modellergebnisse.

Der wesentliche Vorteil der k NN-Methode liegt in der deskriptiven Prognose (DF). Bei der Auflistung von analogen Gefährdungslagen fällt das Modell keinen Entscheid und überlässt dem Anwender die gesamte Verantwortung. Es entsteht somit eine stark subjektiv geprägte Gefahreinschätzung; das Modell dient dabei dem Anwender als unterstützende Entscheidungshilfe.

Die Durchführung von Verifikationsberechnungen ist für die Einschätzung der Modellqualität unabdingbar. Anders als im Bereich der meteorologischen Prognosen sind bei der Risikoabschätzung von Naturgefahren systematische Modellverifikationen nicht weit verbreitet.

Die im Rahmen der vorliegenden Dissertation erarbeiteten Grundlagen sollten u.a. auch bei künftigen Modellvergleichen und bei der Einschätzung der erreichten Qualität eines Modells als Basis dienen.

Mit Hilfe der k NN-Methode ist ein in der Waldbrandprognose neuartiges Modell zur Prognose der Waldbrandgefahr entwickelt worden, welches sowohl den Anforderungen der Wissenschaft, als auch der Praxis entspricht. Eine zentrale Herausforderung der Anwendung und der Weiterentwicklung des k NN-Modells besteht nun darin, das Modell so weiterzuentwickeln, dass neben einer Qualitätsverbesserung auch die genannten Bedingungen weiter erfüllt werden können.

9. SCHLUSSFOLGERUNGEN UND PERSPEKTIVEN

Basierend auf einer umfassenden Naturgefahrenbank, soll die Methodik in Zukunft zudem sowohl räumlich als auch thematisch weiterentwickelt werden.

10 Danksagung

Die hier vorliegende Arbeit konnte nur Dank der grossen Hilfe vieler Personen durchgeführt werden. Herzlich bedanken möchte ich mich bei:

- Prof. Dr. E. Parlow für die Bereitschaft die Leitung dieser Arbeit zu übernehmen
- Dr. Perry Bartelt für die Betreuung der Dissertation am SLF
- Prof. Dr. Dieter Scherer für die Übernahme des Koreferats
- Dr. Othmar Buser für die vielen fruchtbaren Diskussionen
- Marco Conedera von der Sottostazione Sud delle Alpi für die wertvollen Informationen bezüglich der Waldbrände auf der Alpensüdseite
- Bernd Hebel und Florian Imbery für die aufwändigen Korrekturen und wertvollen Tipps
- Ross Purves für die Zuverfügungstellung von *Cornice* und weiteren wertvollen Tipps
- Felix Tiefenbacher für die grosse Hilfe und die gute Zeit während den drei gemeinsamen Jahren in Davos
- Casa Letizia ...
- Mathieu, Sven, Katharina, Marc-André, Monique, Dani, Norbi, Martin und Joachim für die Unterstützung und die tolle gemeinsame Zeit

Einen besonderen Dank gilt meinen Eltern und Yvette, ohne deren Unterstützung diese Dissertation nicht durchgeführt hätte werden können.

Diese Arbeit wurde finanziell durch den Schweizerischen Nationalfonds SNF im Rahmen des Projektes Nr. 21-45452.95 und durch das SLF Davos unterstützt.

A Logistisches Regressionsmodell

Wie im Kapitel 4 bereits erwähnt, liesse sich das Klassifikationsproblem, anstelle eines nichtparametrischen- auch mit einem parametrischen Verfahren lösen. Eine logistische Regression (LR) stellt die geeignete Methode zur Lösung eines Klassifikationsproblems mit einer binären Zielgrösse dar.

Im Folgenden werden die Modellgleichungen der logistischen Regression hergeleitet (Stahel, 2002). Die Methode dient einerseits dazu, die Variablenbestimmung für das k NN-Modell vorzunehmen. Andererseits kann basierend auf denselben Grundlagen ein eigenständiges Modell zur Prognose der Waldbrandgefahr entwickelt werden.

Da es sich bei der logistischen Regression um ein parametrisches Verfahren handelt, bei welcher die Verteilung der Daten bekannt sein muss, wird im Folgenden davon ausgegangen, dass die Waldbrandereignisse Y_i (Gl. A.1)

$$Y_i = B(n, \pi) \tag{A.1}$$

binomialverteilt sind, wobei n die Anzahl Tage in der Datenbank darstellt, und die Wahrscheinlichkeit π durch die Grundwahrscheinlichkeit s (*base rate*) gegeben ist. Da für die Regressionsfunktion nur Werte zwischen 0 und 1 zulässig sind, wird ein linearer Prädiktor gebildet (Gl. A.2):

$$\eta_i = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \dots + \beta_p x_i^{(p)} = \mathbf{x}_i^\dagger \beta, \tag{A.2}$$

wobei $\mathbf{x}_i = (1, x_i^{(1)}, \dots, x_i^{(p)})^\dagger$ die zum Tag i gehörenden Werte der p Variablen und $\beta = (\beta_0, \dots, \beta_p)^\dagger$ die dazugehörenden, für alle Tage i geltenden Gewichte darstellt.

Auf diesen linearen Prädiktor wird nun die Transformation $g^+ : \mathbb{R} \mapsto [0, 1]$ angewendet, welche folgende Bedingung erfüllt:

$$P\langle Y_i = 1 \rangle = \mathcal{E}\langle Y_i \rangle = g^+\langle \eta_i \rangle. \tag{A.3}$$

Gewöhnlich wird für die Transformation g^+ die logistische Funktion

$$g^+\langle \eta_i \rangle = \frac{\exp\langle \eta_i \rangle}{1 + \exp\langle \eta_i \rangle}. \tag{A.4}$$

A. LOGISTISCHES REGRESSIONSMODELL

gewählt.

Die klassischen Voraussetzungen der linearen Regression haben folgende Analogien zum Modell der logistischen Regression:

1. Die Y_i müssen unabhängig sein.
2. Y_i sind binomialverteilt. Diese Annahme ist trivialerweise erfüllt, da Y_i nur zwei Werte 0 oder 1 annehmen kann.
3. Für den Erwartungswert von Y_i , der definitionsgemäss mit $P(Y_i = 1)$ identisch ist, wird angenommen:

$$\mathcal{E}(Y_i) = P(Y_i = 1) = \frac{\exp\langle\eta_i\rangle}{1 + \exp\langle\eta_i\rangle} \quad (\text{A.5})$$

4. Für die Varianz der binomialverteilten Variablen Y_i gilt:

$$\text{Var}(Y_i) = P(Y_i = 1)P(Y_i = 0).$$

Die Varianz hängt also vom Erwartungswert ab (Stahel, 2002).

Mit dem Modell können die Parameter β für einen gegebenen Datensatz geschätzt werden. Weiter berechnet das Modell die Signifikanz (P -Wert) der Variablen. Diese Signifikanz wurde im k NN-Modell für die Auswahl der Variablen berücksichtigt.

Für ein Prognosetag $\mathbf{y} = (1, y^{(1)}, \dots, y^{(p)})^\dagger$ kann nun der entsprechende lineare Prädiktor $\eta_y = \mathbf{y}^\dagger$ gebildet werden. Damit liefert das hier dargestellte logistische Modell mit

$$p_{LR} = g^+\langle\eta_y\rangle = \frac{\exp\langle\eta_y\rangle}{1 + \exp\langle\eta_y\rangle} \quad (\text{A.6})$$

die Wahrscheinlichkeit dafür, dass sich am Tag \mathbf{y} ein Waldbrand ereignet. Neben dem k NN-Modell ist dies eine zweite, jedoch parametrische Lösung des Klassifikationsproblems. Dieses logistische Modell kann analog zum k NN-Modell mit den angeführten Methoden Verifiziert werden (vgl. Kap. 6).

Abbildung A.1 stellt die Auswertung der Wahrscheinlichkeitsberechnungen mittels LR-Methode für die Station Lugano des Untersuchungszeitraumes 1991-2002 dar. Darin wird die Häufigkeit von Waldbrandtagen respektive Nicht-Waldbrandtagen für unterschiedlich prognostizierte Wahrscheinlichkeiten p_{LR} gezeigt. Beim

A. LOGISTISCHES REGRESSIONSMODELL

dritten Plot wird die prognostizierte Wahrscheinlichkeit p^* mit dem tatsächlichen Eintritt P von Ereignissen korreliert. Wenn für ein Tag z.B. eine Wahrscheinlichkeit von 0.4 prognostiziert worden ist, kam es in 40% aller Fälle zu einem Ereignis. Liegen die Punkte unter der Winkelhalbierenden (1:1-Gerade), hat das Modell die Gefahr überschätzt; Liegen sie darüber, wurde unterschätzt. Die Ereignisse werden bei grossen Wahrscheinlichkeiten ($P > 0.5$) so selten, dass betreffend einer Eintrittswahrscheinlichkeit p^* keine gesicherten Aussagen mehr möglich sind.

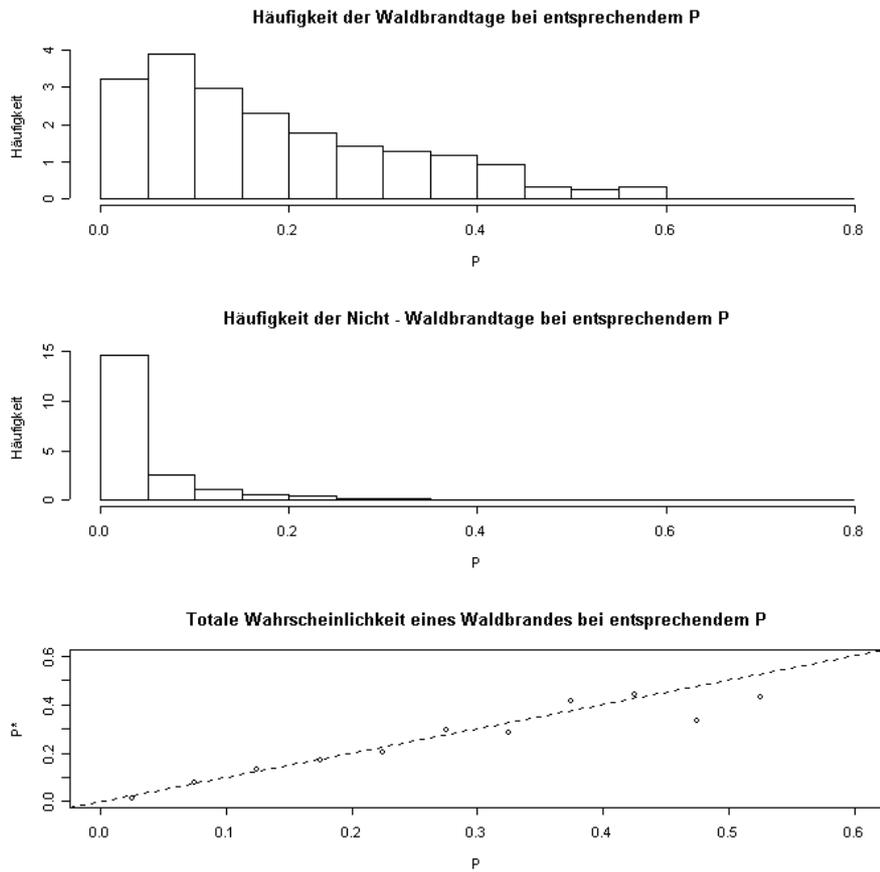


Abb. A.1: Verifikation des logistischen Prognosemodells.

A. LOGISTISCHES REGRESSIONSMODELL

Fazit

Ein Vergleich der logistischen Vorhersage mit der Prognosequalität der probabilistischen Prognose des k NN-Modells zeigt, dass die Resultate des LR-Modells von ähnlicher Qualität sind.

B Gebietszuordnung der Gemeinden

- Region 1: Messstation Lugano
- Region 2: Messstation Locarno
- Region 3: Messstation Piotta

PLZ	Gemeinde	Kanton	Region
5160	BRUSINO ARSIZIO	TI	1
5242	BALERNA	TI	1
5147	BARBENGO	TI	1
5148	BEDANO	TI	1
5149	BEDIGLIORA	TI	1
5243	BESAZIO	TI	1
5150	BIDOGNO	TI	1
5151	BIOGGIO	TI	1
5153	BIRONICO	TI	1
5141	AGNO	TI	1
5142	AGRA	TI	1
5143	ARANNO	TI	1
5144	AROGNO	TI	1
5145	AROSIO	TI	1
5241	ARZO	TI	1
5146	ASTANO	TI	1
5250	CHIASSO	TI	1
5174	CIMADERA	TI	1
5175	CIMO	TI	1
5247	CAPOLAGO	TI	1
5246	CANEGGIO	TI	1
5167	CANOBBIO	TI	1
5168	CARABBIA	TI	1

B. GEBIETSZUORDNUNG DER GEMEINDEN

PLZ	Gemeinde	Kanton	Region
5169	CARABIETTA	TI	1
5170	CARONA	TI	1
5248	CASIMA	TI	1
5171	CASLANO	TI	1
5249	CASTEL SAN PIETRO	TI	1
5244	BRUZELLA	TI	1
5245	CABBIO	TI	1
5161	CADEMARIO	TI	1
5162	CADEMPINO	TI	1
5163	CADRO	TI	1
5164	CAGIALLO	TI	1
5165	CAMIGNOLO	TI	1
5155	BOGNO	TI	1
5154	BISSONE	TI	1
5156	BOSCO LUGANESE	TI	1
5158	BREGANZONA	TI	1
5159	BRENO	TI	1
5173	CERTARA	TI	1
5231	VEZIA	TI	1
5232	VEZIO	TI	1
5233	VICO MORCOTE	TI	1
5234	VIGANELLO	TI	1
5235	VILLA LUGANESE	TI	1
5267	TREMONA	TI	1
5268	VACALLO	TI	1
5228	VAGLIO	TI	1
5229	VALCOLLA	TI	1
5230	VERNATE	TI	1
5266	STABIO	TI	1
5222	SESSA	TI	1
5223	SIGIRINO	TI	1
5224	SONVICO	TI	1
5225	SORENGO	TI	1
5226	TESSERETE	TI	1
5227	TORRICELLA-TAVERNE	TI	1
5219	ROVIO	TI	1
5264	SAGNO	TI	1
5220	SALA CAPRIASCA	TI	1
5265	SALORINO	TI	1
5221	SAVOSA	TI	1

B. GEBIETSZUORDNUNG DER GEMEINDEN

PLZ	Gemeinde	Kanton	Region
5262	RANCATE	TI	1
5215	PREGASSONA	TI	1
5216	PURA	TI	1
5263	RIVA SAN VITALE	TI	1
5217	RIVERA	TI	1
5218	ROVEREDO (TI)	TI	1
5210	PARADISO	TI	1
5211	PAZZALLO	TI	1
5212	PONTE CAPRIASCA	TI	1
5213	PONTE TRESA	TI	1
5214	PORZA	TI	1
5205	MUZZANO	TI	1
5204	MUGENA	TI	1
5259	MUGGIO	TI	1
5206	NEGGIO	TI	1
5207	NOVAGGIO	TI	1
5260	NOVAZZANO	TI	1
5208	ORIGLIO	TI	1
5209	PAMBIO-NORANCO	TI	1
5203	MORCOTE	TI	1
5255	MERIDE	TI	1
5199	MEZZOVICO-VIRA	TI	1
5200	MIGLIEGLIA	TI	1
5201	MONTAGNOLA	TI	1
5256	MONTE	TI	1
5202	MONTEGGIO	TI	1
5257	MORBIO INFERIORE	TI	1
5258	MORBIO SUPERIORE	TI	1
5193	MAGLIASO	TI	1
5194	MANNO	TI	1
5195	MAROGGIA	TI	1
5196	MASSAGNO	TI	1
5197	MELANO	TI	1
5198	MELIDE	TI	1
5254	MENDRISIO	TI	1
5190	LOPAGNO	TI	1
5191	LUGAGGIA	TI	1
5192	LUGANO	TI	1
5110	INDEMINI	TI	1
5188	ISEO	TI	1

B. GEBIETSZUORDNUNG DER GEMEINDEN

PLZ	Gemeinde	Kanton	Region
5189	LAMONE	TI	1
5253	LIGORNETTO	TI	1
5252	GENESTRERIO	TI	1
5185	GENTILINO	TI	1
5186	GRANCIA	TI	1
5187	GRAVESANO	TI	1
5184	GANDRIA	TI	1
5178	CROGLIO-CASTELROTTO	TI	1
5179	CUREGGIA	TI	1
5180	CUREGLIA	TI	1
5181	CURIO	TI	1
5182	DAVESCO-SORAGNO	TI	1
5183	FESCOGGIA	TI	1
5251	COLDRERIO	TI	1
5176	COMANO	TI	1
5238	COMUN. CORTICIASCA-VALCOLLA	TI	1
5236	COMUN. SALA CAPR-BIDOGN-CORTIC	TI	1
5237	COMUN. SALA CAPRIASCA-VAGLIO	TI	1
5177	CORTICIASCA	TI	1
5302	AVEGNO	TI	2
5002	BELLINZONA	TI	2
5281	BIASCA	TI	2
5301	AURIGENO	TI	2
5001	ARBEDO-CASTIONE	TI	2
5091	ASCONA	TI	2
5092	AURESSIO	TI	2
5282	CLARO	TI	2
5003	CADENAZZO	TI	2
5004	CAMORINO	TI	2
5094	BORGNONE	TI	2
5097	BRISSAGO	TI	2
5134	VIRA (GAMBAROGNO)	TI	2
5135	VOGORNO	TI	2
5018	SANT'ANTONIO	TI	2
5096	BRIONE S/MINUSIO	TI	2
5322	SOMEIO	TI	2
5133	VERSCIO	TI	2
5130	TEGNA	TI	2
5131	TENERO-CONTRA	TI	2
5127	SAN NAZZARO	TI	2

B. GEBIETSZUORDNUNG DER GEMEINDEN

PLZ	Gemeinde	Kanton	Region
5128	SANT'ABBONDIO	TI	2
5017	SANT'ANTONINO	TI	2
5019	SEMENTINA	TI	2
5015	PREONZO	TI	2
5016	ROBASACCO	TI	2
5125	RONCO SOPRA ASCONA	TI	2
5014	PIANEZZO	TI	2
5123	PIAZZOGNA	TI	2
5120	MURALTO	TI	2
5121	ORSELINA	TI	2
5286	OSOGNA	TI	2
5122	PALAGNEDRA	TI	2
5118	MINUSIO	TI	2
5319	MOGHEGNO	TI	2
5012	MOLENO	TI	2
5013	MONTE CARASSO	TI	2
5317	MAGGIA	TI	2
5011	MEDEGLIA	TI	2
5117	MERGOSCIA	TI	2
5113	LOCARNO	TI	2
5316	LODANO	TI	2
5285	LODRINO	TI	2
5115	LOSONE	TI	2
5010	LUMINO	TI	2
5116	MAGADINO	TI	2
5008	GUDO	TI	2
5111	INTRAGNA	TI	2
5284	IRAGNA	TI	2
5009	ISONE	TI	2
5112	LAVERTEZZO	TI	2
5007	GORDUNO	TI	2
5106	GERRA (GAMBAROGNO)	TI	2
5005	GIUBIASCO	TI	2
5313	GIUMAGLIO	TI	2
5006	GNOSCA	TI	2
5314	GORDEVIO	TI	2
5108	GORDOLA	TI	2
5283	CRESCIANO	TI	2
5104	CUGNASCO	TI	2
5020	COMUN. MEDEGL.-ROBASSA.	TI	2

B. GEBIETSZUORDNUNG DER GEMEINDEN

PLZ	Gemeinde	Kanton	Region
5101	CONTONE	TI	2
5102	CORIPPO	TI	2
5311	COGLIO	TI	2
5098	CAVIANO	TI	2
5099	CAVIGLIANO	TI	2
5063	BEDRETTO	TI	3
5093	BERZONA	TI	3
5303	BIGNASCO	TI	3
5061	AIROLO	TI	3
5062	ANZONICO	TI	3
5031	AQUILA	TI	3
5306	BRONTALLO	TI	3
5310	CEVIO	TI	3
5069	CHIGGIOGNA	TI	3
5070	CHIRONICO	TI	3
5067	CAMPELLO	TI	3
5032	CAMPO (BLENIO)	TI	3
5307	CAMPO (VALLEMAGGIA)	TI	3
5033	CASTRO	TI	3
5065	CALONICO	TI	3
5066	CALPIOGNA	TI	3
5064	BODIO	TI	3
5304	BOSCO/GURIN	TI	3
5095	BRIONE (VERZASCA)	TI	3
5305	BROGLIO	TI	3
5132	VERGELETTTO	TI	3
5081	SOBRIO	TI	3
5129	SONOGNO	TI	3
5047	TORRE	TI	3
5046	SEMIONE	TI	3
5045	PRUGIASCO	TI	3
5079	QUINTO	TI	3
5080	ROSSURA	TI	3
5321	PRATO-SORNICO	TI	3
5320	PECCIA	TI	3
5076	PERSONICO	TI	3
5077	POLLEGIO	TI	3
5044	PONTO VALENTINO	TI	3
5078	PRATO (LEVENTINA)	TI	3
5119	MOSOGNO	TI	3

B. GEBIETSZUORDNUNG DER GEMEINDEN

PLZ	Gemeinde	Kanton	Region
5043	OLIVONE	TI	3
5075	OSCO	TI	3
5074	MAIRENGO	TI	3
5041	MALVAGLIA	TI	3
5042	MAROLTA	TI	3
5318	MENZONIO	TI	3
5315	LINESCIO	TI	3
5114	LOCO	TI	3
5039	LOTTIGNA	TI	3
5040	LUDIANO	TI	3
5109	GRESSO	TI	3
5037	LARGARIO	TI	3
5038	LEONTICA	TI	3
5107	GERRA (VERZASCA)	TI	3
5036	GHIRONE	TI	3
5073	GIORNICO	TI	3
5071	DALPE	TI	3
5035	DONGIO	TI	3
5072	FAIDO	TI	3
5105	FRASCO	TI	3
5312	FUSIO	TI	3
5136	ONSERNONE	TI	3
5034	CORZONESO	TI	3
5068	CAVAGNAGO	TI	3
5308	CAVERGNO	TI	3
5309	CERENTINO	TI	3

Literaturverzeichnis

- Adriaans, P. & Zantige, D. (1996a). *Data Mining*. Addison-Wesley.
- Adriaans, P. & Zantige, D. (1996b). *Data Mining*. Addison Wesley.
- Alexakis, E. (2000). L'influence du climat et l'homme sur l'évolution des feux de forêt au Tessin. Master's thesis, Université de Genève.
- Allgöwer, B. (2001). *Wildfire Research in Switzerland*. Available at: <http://www.geo.unizh.ch/gis/research/edmg/fire>. Accessed 07.03.05.
- Bellmann, R. (1961). *Adaptive Control Processes*. Princeton University Press.
- Bossel, H. (1992). *Modellbildung und Simulation*. Verlag Vieweg.
- Bovio, G., Quaglino, A., & Nosenzo, A. (1994). Individuazione di un indice di previsione per il pericolo di incendi boschivi. *Monte e Boschi* (4).
- Brabec, B. (2001). *Computergestützte regionale Lawinenprognose*. Ph. D. thesis, Eidg. Tech. Hochschule Zürich, ETHZ.
- Burgan, R. & Rothermel, R. (1984). Behave: fire behavior prediction and fuel modeling system-fuel subsystem. Technical report, U.S. Department of Agriculture, Forest Service.
- Buser, O. (1983). Avalanche forecast with the method of nearest neighbours: An interactive approach. *Cold Regions Science and Technology* (8), 155–163.
- Buser, O., Bütler, M., & Good, W. (1986). Avalanche forecast by the nearest neighbor method. In *Avalanche Formation, Movement and Effects*, Volume 162, Davos, pp. 557–569. IAHS.
- BUWAL (1998). Begriffsdefinition zu den Themen: Geomorphologie, Naturgefahren, Forstwesen, Sicherheit und Risiko. Technical report, BUWAL.

LITERATURVERZEICHNIS

- Ceschi, I. (1977). Gli incendi boschivi in Canton Ticino. Technical Report 65, Bollettino Società Ticinese Scienze Naturali.
- Cleveland, R., Cleveland, W., McRae, J., & Terpening, I. (1990). StI: A seasonal-trend decomposition procedure based on loess. *Journal of Statistics* 6, 3–73.
- Conedera, M. (2003). Waldbrände. In *Extremereignisse und Klimaänderung*, pp. 57–61. OcCC.
- Conedera, M., Marcozzi, M., Jud, B., Mandallaz, D., Chatelain, F., Frank, C., Kienast, F., Ambrosetti, P., & Corti, G. (1996). *Incendi boschivi al Sud delle Alpi: passato, presente e possibili sviluppi futuri*. vdf Hochschulverlag.
- Conedera, M., Marxer, P., Ambrosetti, P., G., D. B., & Spinedi, F. (1998). The 1997 forest fire season in Switzerland. *International Forest Fire News* (18), 85–88.
- Conedera, M., Marxer, P., Moretti, M., & Tinner, W. (1997). Waldbrandforschung an der FNP Sottostazione Sud delle Alpi. *Schweizer Wald* (12), 18–23.
- Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on information Theory* 13, 21–27.
- Damrath, U. (2002). Verifikation von Wettervorhersagen. In D. Wetterdienst (Ed.), *Die neue Modellkette des DWD II*, Volume 1/2.
- Dasarathy, B. (1991). *Nearest Neighbor Norms: NN Pattern Classification Techniques*. IEEE Computer Society Press.
- Deeming, J., Burgan, R., & Cohen, J. (1977). The national fire danger rating system - 1978. Gen. tech. report int-39, USDA Forest Service.
- DWD (2005). *Waldbrandgefahrenindex M-68*. Deutscher Wetterdienst. Available at: <http://www.agrowetter.de/Agrarwetter/waldix.htm>. Accessed 07.03.05.
- Ellenberg, H. (1996). *Vegetation Mitteleuropas mit den Alpen* (6 ed.). Eugen Ulmer Verlag.
- Enas, G. & Choi, S. (1986). Choice of the smoothing parameter and efficiency of k-nearest neighbour classification. *Computers and Mathematics with Applications* 2.

- Evgeniou, T., Pontil, M., & Elisseeff, A. (2004). Leave one out error, stability, and generalization of voting combinations of classifiers. *Mach. Learn.* 55(1), 71–97.
- FAO (1999). State of the world’s forest. Technical report, Food and Agriculture Organization of the United Nations.
- Felber, A. & Bartelt, P. (2002). Definition of forest fire hazard variables of the nearest neighbour forecasting method. In D. X. Viegas (Ed.), *IV International Conference on Forest Fire Research*, Luso – Coimbra, Portugal. in press.
- Finley, J. (1884). Tornado prediction. *American Meteorological Journal* 1, 85–88.
- Fix, E. & Hodges, J. (1951). Discriminatory analysis: Nonparametric discrimination: Consistency properties. Technical report, USAF School of Aviation Medicine.
- Fix, E. & Hodges, J. (1952). Discriminatory analysis: Nonparametric discrimination: Small sample performance. Technical report, USAF School of Aviation Medicine.
- Friedman, J. (1993). *From Statistics to Neural Networks: Theory and Pattern Recognition applications*, Volume 136 of *Series F, Computer and System Sciences*, Chapter An overview of predictive learning and functional approximation, pp. 1–61. Springer.
- Garcia Diez, E., Rivas Soriano, L., De Pablo, F., & Garcia Diez, A. (1999). Prediction of the daily number of forest fires. *International Journal of Wildland Fire* 9(3).
- Gassner, M., Birkeland, K., Etter, H., & Leonard, T. (2000). Nxd 2000: An improved avalanche forecasting program based on the nearest neighbor method. In *International Snow Science Workshop*, Big Sky, Montana, USA.
- GFMC, I. (2005). *Global Fire Monitoring Center*. The GFMC is an Activity of the UN International Strategy for Disaster Reduction ISDR. Available at: <http://www.fire.uni-freiburg.de>. Accessed 07.03.05.
- Goldberg, D. (1989). *Genetic Algorithms in Search Optimization and Machine Learning*. Addison-Wesley Publishing.

LITERATURVERZEICHNIS

- Goncalvez, Z. & Lourenco, L. (1990). Meteorological index of forest fire risk in the portuguese mainland territory. In *I. International Conference on Forest Fire Research*, Coimbra, Portugal.
- Hebel, B. (2003). *Validierung numerischer Erosionsmodelle in Einzelhang- und Einzugsgebiet-Dimension*. Ph. D. thesis, Universität Basel.
- Heierli, J., Purves, R., Felber, A., & Kowalski, J. (2003). Verification of nearest neighbours interpretations in avalanche forecasting. *Annals of Glaciology* 38.
- Holland, J. (1975). *Adaption in Natural and Artificial Systems*. University of Michigan.
- Hürzeler, M. (2003). *Zeitreihenanalyse: Unterlagen für den Nachdiplomkurs in angewandter Statistik an der ETH Zürich*. Skript.
- Jeklin, R. & Schöb, T. (1993). *Waldbrandbekämpfung im Gebirge*. Verlag Bündner Wald.
- Jenkins, M. & Hebertson, E. (2000). Incorporating a human risk factor in avalanche danger rating systems. In *ISSW*.
- Kristensen, K. & Larsson, C. (1994). An avalanche forecasting system program based on a modified nearest neighbor method. In *International Snow Science Workshop*, Snowbird, Utah, USA.
- Langhart, R. (1999). *Räumliche und zeitliche charakterisierung von waldbrandherden im kanton graubünden*. Master's thesis, Universität Zürich.
- Loftsgaarden, D. & Quesenberry, C. (1965). A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics* (36), 1049–1051.
- Mandallaz, D. & Ye, R. (1996). *Statistical model for the prediction of forest fires. Project minerve iii - final report*, Swiss Institute of Technology Zurich, Zurich, Switzerland.
- Mani, P. (1995). Erfassung der Wildbachgefahr mit Hilfe von GIS-basierten Modellen. *Schweiz. Zeitschr. f. Forstwesen* (9), 727–739.

- Marxer, P. (2003). *Oberflächenabfluss und Bodenerosion auf Brandflächen des Kastanienwaldgürtels der Südschweiz mit einer Anleitung zur Bewertung der post-fire Erosionsanfälligkeit*. Ph. D. thesis, Universität Basel.
- Mason, I. (1980). Decision-theoretic evaluation of probabilistic predictions. In *WMO Symposium on Probabilistic and Statistical Methods in Weather Forecasting*.
- Mason, I. (1982a). A model for assessment of weather forecasts. *Austral. Met. Mag.* 30, 291–303.
- Mason, I. (1982b). On scores for yes/no forecasts. In *Preprints of papers delivered at the Ninth AMS Conference on Weather Forecasting and Analysis*.
- Mason, I. (1989). Dependence of the critical success index on sample climate and threshold probability. *Austral. Met. Mag.* 37, 75–81.
- Matthews, R. (1996a). Base-rate errors and rain forecasts. *Nature* 382, 766.
- Matthews, R. (1996b). Why is weather forecast still under a cloud. *Mathematics today* 32, 168–170.
- Meister, R. & Brabec, B. (1999). Nex-mod. Techn. ber., Eidg. Inst. für Schnee- und Lawinenforschung SLF, Davos.
- Mitchell, T. (1997). *Machine Learning*. McGraw Hill.
- Mitchie, D., Spiegelhalter, D., & Taylor, C. (1994). *Machine Learning, Neural and Statistical Classification*. Ellis Horwood.
- Moretti, M., Patocchi, N., Conedera, M., Duelli, P., & Edwards, P. (1998). The influence of single and repeated fires on invertebrates in chestnut forests in southern Switzerland. In L. Trabaud (Ed.), *Fire Management and Landscape Ecology*, pp. 237–245. International Association of Wildland Fire.
- Murphy, A. (1973). A new vector partition of the probability score. *Journal of Applied Meteorology* (12), 595–600.
- Murphy, A. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting* 8, 281–293.

LITERATURVERZEICHNIS

- Murphy, A. & Winkler, R. (1987). A general framework for forecast verification. *Mon. Weather Rev.* 115, 1330–1338.
- Nissen, V. (1997). *Einführung in die Evolutionären Algorithmen. Optimierung nach dem Vorbild der Evolution.* Viewag Verlag.
- Nothnagel, M. (1999). Klassifikationsverfahren der Diskriminanzanalyse. Master's thesis, Institut für Mathematik, Humboldt-Universität, Berlin.
- Popper, K. (1959). *The Logic of Scientific Discovery.* Harper, New York.
- Popper, K. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge.* Harper, New York.
- Purves, R., Morrison, K., Moss, G., & Wright, B. (2002). Cornice - development of a nearest neighbours model applied in backcountry avalanche forecasting in scotland. ISSW.
- Rebetez, M. (1999, March). Twentieth century trends in droughts in southern switzerland. *Geophysical research letters* 26(6), 755–758.
- Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* Frommann-Holzboog.
- Ripley, B. (1996). *Pattern recognition and neural networks.* Cambridge University Press.
- Scherer, D. (1998). Regionale Systemanalyse. Theorie und Beispiele. Habilitationsschrift, Philosophisch Naturwissenschaftliche Fakultät der Universität Basel.
- Schneebeli, M. & Laternser, M. (2004). A probabilistic model to evaluate the optimal density of stations measuring snowfall. *Journal of Applied Meteorology* 43(5), 711–719.
- Schöning, R., Bachmann, A., & Maissen, U. (1998). Incendi: Unterstützung der Waldbrandwarnung im Kanton Graubünden. *ArcAktuell* (3).
- Schönwiese, C. (1991). *Praktische Statistik für Meteorologen und Geowissenschaftler*, Volume 2. Aufl. Gebrüder Bornträger.

- Sol, B. (1989). Risque numerique meteorologique d'incendies de foret en région mediterraneene. Note de travaux, SMIRSE.
- Stahel, W. (2002). Lineare Regression: Block Rg-1 des Nachdiplomkurses Statistik der ETHZ. Vorlesungsskript, ETH Zürich.
- Stephenson, D. & Jolliffe, I. (2003). *Forecast Verification. A Practinioneer's Guide in Atmospheric Science*. Wiley.
- Tinner, W., Conedera, M., Ammann, B., Gäggeler, H., Geyde, S., Jones, R., & Sägesser, B. (1998). Pollen and charcoal in lake sediments compared with historically documented forest fires in Southern Switzerland since 1920. *The Holocene* (8), 32–42.
- Tinner, W., Hubschmid, P., Wehrli, M., Ammann, B., & Conedera, M. (1999). Long-term forest fire ecology and dynamics in Southern Switzerland. *Journal of Ecology* (87), 273–289.
- USDA (2005). *Fire and Aviation Management*. USDA, Forest Service. Available at: <http://www.fs.fed.us/fire>. Accessed 07.03.05.
- Van Wagner, C. (1987). Development and structure of the canadian forest fire weather index system. Report, Canadian Forestry Service.
- Viegas, D., Bovio, G., Ferreira, A., Nosenzo, A., & Sol, B. (1999a). Comparative study of various methods of fire danger evaluation in southern europe. *International Journal of Wildland Fire* 9(4), 235–246.
- Viegas, X., Bovio, G., Ferreira, A., Nosenzo, A., & Sol, B. (1999b). Comparative Study of Various Methods of Fire Danger Evaluation in Southern Europe. *International Journal of Wildland Fire* 9(4), 235–246.
- Weiss, S. & Indurkha, N. (1998). *Predictive Data Mining*. Morgan Kaufmann Publishers, Inc.
- Wilks, D. (1995). *Statistical Methods in the Atmospheric Sciences*. Academic Press.
- Winkler, O. (1971). *Anleitung zur Verhütung und Bekämpfung von Waldbränden*, Volume 2. Auflage. Verlag der Forstwirtschaftlichen Zentralstelle der Schweiz.

Curriculum Vitae

Persönliches

Name	Andreas Felber
Geburtsdatum	15. Mai 1975
Geburtsort	Basel, Schweiz
Nationalität	Schweiz
aktuelle Wohnadresse	Flurstrasse 4 7270 Davos Platz

Ausbildung

- 1995	Matura im Mathematisch-Naturwissenschaftlichen Gymnasium in Basel
1995 - 2001	Studium der Geographie, Meteorologie und Ethnologie an der Universität Basel
1999 - 2000	Studium von Freizeit und Tourismus an der Universität Bern
2000 - 2001	Diplomarbeit in der Meteorologie bei Prof. Dr. D. Scherer, Universität Basel
1999 - 2000	Wissenschaftlicher Hilfsassistent bei Prof. Dr. E. Parlow, Universität Basel
2001 - 2003	Nachdiplomkurs in angewandter Statistik an der ETH Zürich
2002 - 2003	Nachdiplomkurs Räumliche Informationssysteme an der ETH Zürich
2001 - 2005	Dissertation am Institut für Schnee- und Lawinenforschung, SLF in Davos