

DER EINSATZ VERSCHIEDENER KLASSIFIKATOREN ZUR ERKENNUNG FLACHSCHALIGER RUTSCHUNGEN

MODELLGEBIET SACHSELN / SCHWEIZ

IMPLEMENTATION OF DIFFERENT CLASSIFICATION METHODS FOR DETECTING SOIL SLIPS

MODEL AREA SACHSELN / SWITZERLAND

Maik Hamberger¹, Michael Moser²

ZUSAMMENFASSUNG

Am 15. August 1997 verursachten Starkniederschläge mit einer Intensität von 150 mm in zwei Stunden mehr als 700 kleinflächige und flachschalige Rutschungen, die zu großen Vermurungen in den Talbereichen in der Umgebung von Sachseln und des Melchtales in der Schweiz führten. In den Folgejahren wurden im Rahmen verschiedener Projekte umfangreiche Aufnahmen durchgeführt, um den Einfluss verschiedener Parameter auf die Rutschaktivität eines Hanges zu ermitteln. Die Ergebnisse dieser ersten Evaluierung zeigten, dass die Rutschungen nicht durch einen Parameter - morphologischer, geotechnischer, geologischer, hydrologischer oder vegetationskundlicher Art -, sondern eine Kombination vieler Faktoren ausgelöst wurden. Mit drei unterschiedlichen Klassifikationsverfahren, der linearen Diskriminanzanalyse, den Support Vector Machines und den Klassifikationsbäumen, wurde versucht, Parameterkonstellationen zu finden, die effizient zwischen Rutschungsbereichen und Nichtrutschungsbereichen unterscheiden. Die Ergebnisse der durchgeführten Klassifikationen belegen, dass es zum einen möglich ist, einen Großteil (81 %) der Rutschungen mit verschiedenen Modellen zu erkennen, zum anderen aber auch die Möglichkeit besteht, nicht nur ca. 74 % der Rutschungen sondern ebensoviel Nichtrutschungen richtig zu klassifizieren.

Keywords: Hangbewegungen, Rutschungserkennung, GIS, lineare Diskriminanzanalyse, Support Vector Machines, Klassifikationsbäume

ABSTRACT

On August 15th, 1997 a thunderstorm, which had a total rainfall of 150 mm over a two hour period, was recorded in Central Switzerland and resulted in more than 700 soil slips and huge

¹ Dipl.-Geol., Trumer Schutzbauten GmbH, Weißenbach 106, 5431 Kuchl, Österreich (Tel.: +43-6244-20325; Fax: +43-6244-20325-11; email: m.hamberger@trumerschutzbauten.com)

² Prof. Dr., Lehrstuhl für Angewandte Geologie, Schlossgarten 5, 91054 Erlangen, Deutschland (Tel.: +49-9131-85-22697; Fax: +49-9131-85-29294; email: moser@geol.uni-erlangen.de)

damages in the valleys of Sachseln and Melchtal. These soil slips were evaluated and analysed in the following years. The first results of the analysed soil slips highlighted that morphological, geotechnical, geological, hydrological and botanical factors have a crucial influence on the slope stability. Thus it was determined that the occurrence of a soil slip is not reliant on a single parameter, but rather on a set of parameters. Using three different classification methods - linear discriminant analysis, support vector machines and classification trees - the set of parameters that best separate soil slips areas from non-soil-slip areas were investigated, while at the same time attempting to find the simplest combination thereof. The results of the soil slip recognition demonstrate that it is possible to recognize, on the one hand, most of the soil slips (81 %) using different models and, on the other hand, up to 74 % of both soil slip cells and non-soil-slip cells (an important difference).

Keywords: Soil slips, soil slips recognition, GIS, linear discriminant analysis, support vector machines, classification trees

EINLEITUNG

Naturgefahren haben schon immer die oft sehr begrenzten Siedlungsräume in den Tälern der Alpen bedroht. Naturkatastrophen, bei denen eine Vielzahl von Menschen sowie menschliche Güter betroffen sind, sind in den letzten Jahrzehnten stark angestiegen. Betrachtet man alleine den Zeitraum seit 1970, so hat sich die Zahl der weltweit erfassten Naturkatastrophen mehr als vervierfacht (Zanetti et al., 2004). In der Schweiz wurden die schwerwiegendsten Naturkatastrophen durch Massenbewegungen und Lawinen ausgelöst. Bei den zehn schwersten Naturkatastrophen der Schweiz der letzten 50 Jahre wurden bei insgesamt sieben Ereignissen 273 Menschen durch Massenbewegungen und Lawinen getötet. Die Zahl der durch diese Katastrophen betroffenen Personen lag bei über 3500 (CRED, 2004). Seit 1972 werden in der Schweiz unterschiedliche Naturgefahren in Karten erfasst, um gefährdete Gebiete nicht für die Überbauung und Besiedelung freizugeben. Mitte der 70er Jahre entwirft Kienholz (1977) einen Leitfaden für die Aufnahme von Naturgefahren und deren Darstellung in geomorphologischen Karten.

Obwohl seit nunmehr über 30 Jahren Verordnungen zur Erhebung von Naturgefahren existieren, gibt es auch 2007 noch viele Bereiche im Alpenraum, in denen noch keine Naturgefahrenbeurteilung vorgenommen wurde. Dazu zählen auch Gebiete, die eine dichte Besiedelung aufweisen. Das Problem der flächendeckenden Gefahrenbeurteilung liegt im damit verbundenen hohen Zeitaufwand und der Komplexität verschiedener Naturgefahren, die

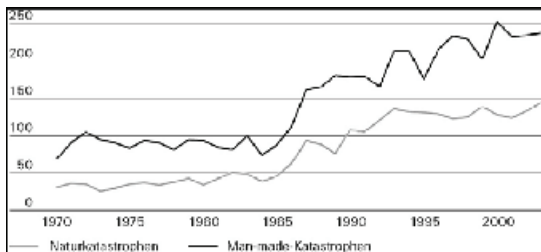


Abb. 1: Anzahl der weltweit erfassten Naturkatastrophen und Man-made-Katastrophen seit 1970 (Zanetti et al., 2004).

Fig. 1: Worldwide number of natural disasters and Man-made-catastrophes since 1970 (Zanetti et al., 2004).

durch unterschiedliche Faktoren ausgelöst werden können. Die Veränderungen an alpinen Hängen im Laufe der Zeit erfordern es, Gefahrenkarten nach einigen Jahren neu zu überarbeiten und zu verifizieren - was ebenfalls sehr zeitaufwendig ist.

Zudem zeigt eine Literaturrecherche der letzten Jahre, dass das Interesse immer größer wird, ein Verfahren zu entwickeln, welches mit einem

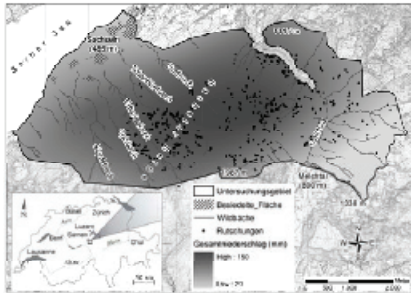


Abb. 2: Verteilung der Niederschläge und Rutschungen im Modellgebiet.

Fig. 2: Distribution of precipitation and soil slips of the model area.



Abb. 3: Rutschungen vom 15.08.1997 in Sachseln/Schweiz.

Fig. 3: Soil slips of Sachseln/Switzerland from August 15th, 1997.

möglichst geringen Aufwand effizient auf die einzelnen Typen von Naturgefahren hinweist. Dazu werden vermehrt Methoden aus dem Bereich der Statistik und Mustererkennung herangezogen.

Am Beispiel Sachseln/Schweiz wird dargestellt, wie mit Klassifikationssystemen flachschalige Rutschungen erkannt werden können, die durch Starkniederschläge ausgelöst wurden. Das Projekt wurde in einer interdisziplinären Zusammenarbeit mit dem Lehrstuhl für Mustererkennung der Universität Erlangen-Nürnberg durchgeführt.

Im Raum Sachseln hat sich am 15. August 1997 eine extrem stationäre Gewitterzelle entladen, in deren Folge über 700 meist kleinflächige und flachschalige Rutschungen entstanden sind. Der Großteil des Anbruchsmaterials wurde über die Gerinne in die Täler hinab transportiert und lagerte sich in den Ortschaften der Tallagen ab. In den Orten Sachseln und Melchtal waren erhebliche Sachschäden zu verzeichnen. Der Gesamtschaden des nur zwei Stunden anhaltenden Starkniederschlags lag bei ca. 100 Mio. CHF; Menschenleben forderte dieses Ereignis glücklicherweise nicht.

GRUNDLAGEN

Evaluierung

Nach den Starkniederschlägen wurden die Schäden im Raum Sachseln und Melchtal nur wenige Tage später erhoben. Intensive Untersuchungen wurden in den nachfolgenden Jahren durch folgende Institutionen durchgeführt:

- 1997 - 1998 Ereignisdokumentation Sachseln (BWW (Bundesamt für Wasserwirtschaft), 1998): Beschreibung der Hydrometeorologie, Spitzenabflüsse und Feststoffbilanzen, der Massenbewegungen und Schäden im Raum Sachseln
- 1998 - 2000 Diplomarbeit an der Universität Erlangen (Hamberger, 2000): Evaluierung der Rutschungen und geotechnische Kartierung im südlichen Melchtal
- 1998 - 2001 Diplomarbeit an der Universität Erlangen (Werner, 2001): Evaluierung der Rutschungen und geotechnische Kartierung im nördlichen Melchtal

- 1998 - 2001 Projekt "Vegetationswirkungen und Rutschungen" der Eidgenössischen Forschungsanstalt WSL: Untersuchungen zum Einfluss der Vegetation auf oberflächennahe Rutschprozesse (Rickli, 2001)
- 1998 - 2000 Diplomarbeit an der ETH-Zürich: Geologische Aufnahmen im Bereich der Sachsler Berge
- 2001 - 2003 Geotechnische Erhebungen für Dissertation an der Universität Erlangen: Geotechnische Kartierung im Bereich der Sachsler Berge

Ergebnisse der Evaluierung

Im Rahmen dieser Arbeiten wurden flächendeckend die Parameter erhoben und analysiert, die sich am 15. August 1997 eventuell negativ bzw. rutschungsfördernd auf die Stabilität ausgewirkt haben. Die Auswertung der Evaluationen von insgesamt 553 Rutschungen führte u.a. zu folgenden Ergebnissen:

- Nach dem Klassifikationsschema von Cruden & Varnes (1996) entstanden neben *translational slides* (60 %) und *rotational slides* (20 %) auch *compound slides* (20 %).
- Die größtenteils oberflächlichen Rutschungen waren im Durchschnitt 18 m lang, 13 m breit und 0,9 m tief. Der Durchschnittswert der Anbruchflächen lag bei 191 m², der Wert der Anbruchsvolumina bei 115 m³.
- Rutschungen traten bei Hangneigungen zwischen 17 und 55° auf, wobei die höchsten Rutschaktivitäten (Rutschungen/ha) zwischen 33° und 44° ermittelt wurden.
- Die Auswertung der Hangmorphologie im Rutschungsbereich zeigt, dass 41 % der Rutschungen an ungliederten Hängen, 21 % in Muldenzonen und ca. 15 % in unmittelbarer Nähe von Hangkanten auftraten. Weitere Rutschungen konnte man im Bereich von Hangrücken und an Kombinationen wiederfinden.
- Die stratigraphischen Einheiten mit einer kalkig-mergeligen Ausbildung (Diphyoideskalk /-mergel, Kieselkalk, Drusbergsschichten) wiesen die höchsten Rutschungsaktivitäten auf.
- Obwohl sich im Höhenspektrum zwischen 1200 m ü. NN und 1700 m ü. NN überdurchschnittlich viel Rutschungen ereignet haben, kann man aufgrund der Niederschlagsverteilung nicht davon ausgehen, dass dieser Bereich generell gefährdeter ist als andere Höhenbereiche. Es ist demnach fraglich, ob die Höhenlage alleine die Rutschaktivität beeinträchtigt hat, oder nur in Kombination mit einem weiteren Faktor als rutschungsrelevant angesehen werden kann. Die Auswertung der Höhenlagen der Rutschungen über der lokalen Erosionsbasis des S' Melchtales zeigte, dass etwa 90 % der Rutschungen maximal 150 m entfernt lagen.
- Die Schutzwirkung des Waldes konnte bei der Auswertung der Vegetationsverhältnisse bestätigt werden, wobei diese bei anderen Ereignissen viel ausgeprägter war. Während die Rutschaktivität in den Sachsler Bergen im Wald knapp über der im Freiland lag, wurden im Melchtal wesentlich mehr Rutschungen im Wald als im Freiland registriert. Die im Wald betroffenen Rutschhänge wiesen im Durchschnitt rund 2,5° steilere Neigungen auf. Daraus kann gefolgert werden, dass die Schutzwirkung des Waldes mit ansteigender Hangneigung abnimmt. Die Berücksichtigung der Waldzustandsbeurteilung von 1987 zeigt eindeutig auf, dass im gesunden Wald weitaus weniger Rutschungen auftraten als in schwach- und mittelstark geschädigten Beständen.
- Mit einfachen bodenmechanischen Betrachtungen, bei denen die Kohäsion nicht berücksichtigt wird, lassen sich die Stabilitätsverhältnisse, wie sie während der Starkniederschläge geherrscht haben, nicht simulieren. Die Böschungsbruchberechnungen nach Janbu führten zu dem Ergebnis, dass für ein labiles Gleichgewicht ($\eta=1$) die

Sickerwasserlinie an der Geländeoberfläche liegen muss und die Kohäsionen der betroffenen Lockergesteine nicht zu hoch angesetzt werden dürfen. Jedoch ist es mit diesem Verfahren nicht möglich, flächendeckend die Stabilitätsverhältnisse des Niederschlagsereignisses zu modellieren.

Basierend auf diesen Ergebnissen konnte gefolgert werden, dass die Rutschungen vom 15. August 1997 nicht durch einen Parameter, sondern eine Kombination mehrerer Parameter ausgelöst wurden.

Klassifikationssysteme

Die Rutschungserkennung wurde mit drei unterschiedlichen Klassifikatoren – lineare Diskriminanzanalyse, Klassifikationsbäume und Support Vector Machines - durchgeführt, wobei bei allen Methoden versucht wurde, Konstellationen von rutschungsrelevanten Parametern zu finden, die auf Rutschungen hinweisen.

Die **lineare Diskriminanzanalyse (LDA)** ist ein klassisches Verfahren im Bereich der multivariaten Datenanalyse, das bereits seit Jahrzehnten in vielen Bereichen Anwendung findet. Bei der linearen Diskriminanzanalyse werden die optimalen Linearkombinationen der eingehenden Merkmale x_1 bis x_n gesucht, indem die Objekte auf gut separierbare Bereiche in einem eindimensionalen Raum abgebildet werden. Während der Zuweisung werden, im Gegensatz zu anderen Klassifikatoren, alle Merkmale gleichzeitig verwendet. Die bei der Klassifikation erzeugte Diskriminanzfunktion d entsteht durch die Gewichtung der Merkmale und ist folgendermaßen festgelegt:

$$d = \left(\sum_{i=1}^n b_i \cdot x_i \right) + a$$

Dabei sind x_1 bis x_n der input (rutschungsrelevante Parameter) und b_1 bis b_n sowie die Konstante a die von der Analyse abzuschätzenden Koeffizienten. Ziel der Diskriminanzanalyse ist es b_1 bis b_n , die häufig auch als Gewichte bezeichnet werden, und a so zu optimieren, dass die gegebenen Klassen nach der Transformation der Daten maximal getrennt erscheinen.

Klassifikationsbäume wurden erstmals Anfang der sechziger Jahre in den Sozialwissenschaften eingesetzt. Die Terminologie eines Entscheidungsbaumes kann als graphisch beschrieben werden. Der Entscheidungsbaum besitzt als obersten Knoten die Wurzel und setzt sich in seiner graphischen Darstellung nach unten fort. Während die Knoten (Verzweigungspunkte und Blätter) mit den Merkmalen des Datensatzes assoziiert sind, sind die Kanten (Äste) des Baumes mit den möglichen Werten der Merkmale belegt (Abb. 4). Ein Klassifikationsbaum trennt den

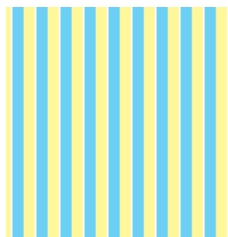


Abb. 4: Arbeitsweise eines vereinfachten Klassifikationsbaumes am Beispiel der Rutschungserkennung.

Fig. 4: Structure of a simplified classification tree for soil slip recognition.

Merkmalsraum in Unterräume, indem jedes Objekt entlang der Kanten bis zu den Blättern klassifiziert wird. Somit verläuft im Gegensatz zur LDA die Zuordnung zu einer Klasse sequentiell. Bei diesen Untersuchungen wurde der C5.0-Algorithmus verwendet, der auf dem 1993 entwickelten ID3-Algorithmus basiert.

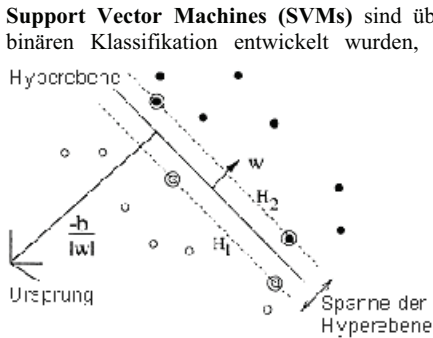


Abb. 5: Lösung eines Zwei-Klassenproblems in einem höherdimensionalen Raum mit Support Vector Machines.

Fig. 5: Solution of a two-classes-problem in a higher dimensional space using support vector machines.

Support Vector Machines (SVMs) sind überwachte maschinelle Lernverfahren, die zur binären Klassifikation entwickelt wurden, mittlerweile aber auch multiclass-Aufgaben übernehmen. Bei der Anwendung von Support Vector Machines zur Lösung eines Klassenproblems wird versucht, Muster effektiv in einem hochdimensionalen Raum abzubilden. Dabei wird eine passende nichtlineare Funktion konstruiert, die mit Hilfe einer optimal trennenden Hyperebene, in einem typischerweise wesentlich höherdimensionalen Raum als der ursprüngliche Merkmalsraum, eine Separation der beiden Klassen vornimmt (Abb. 5). Potenziell besitzt der erzeugte hochdimensionale Merkmalsraum unendlich viele Dimensionen.

RUTSCHUNGSERKENNUNG

Das Ziel der Untersuchungen lag darin, Modelle zu entwickeln, die basierend auf den Aufnahmen der Rutschungsergebnisse in Sachseln und Melchtal zwischen Rutschungen und Nichtrutschungen unterscheiden können. Diese Modellentwicklung bedarf zahlreiche Teilschritte, die den folgenden Arbeitsbereichen zuzuordnen sind (Abb. 6):

- Vorverarbeitung (Preprocessing)
- Analyse (Analyse)
- Nachbearbeitung (Postprocessing)

Preprocessing

Innerhalb des Preprocessings werden all die Schritte vereinigt, die erforderlich sind, um die Analyse im Rahmen der Rutschungserkennung einzuleiten. Hierfür wurden die im Gelände erhobenen Daten und weitere Daten in ein Format gebracht, das von den verwendeten Klassifikationssystemen während der Analyse verarbeitet werden kann.

Ein Großteil der Arbeiten wurde mit dem Geographischen Informationssystem ArcGIS durchgeführt. ArcGIS diente hierbei nicht nur als visuelle Datenbank, in der die Geländedaten digitalisiert und verwaltet wurden, sondern auch als tool, mit dem neue Daten generiert werden konnten. Alle morphologischen Parameter basieren auf einem abflussgerechten digitalen Geländemodell (DGM), das mit der topogrid-Funktion in ArcGIS entwickelt und anhand der aufgenommenen Rutschungsdaten mit Hilfe einer Abweichungsanalyse verifiziert wurde. Die Rasterzellengröße des verwendeten DGM lag nach einer Vergrößerung (anfangs 2 x 2 m) bei 5 x 5 m. Mit der flowaccumulation-Funktion wurde ebenfalls in ArcGIS ein

Parameter erzeugt, der die Anzahl der Zellen angibt, welche eine Zelle durchfließen. Die Parameter, die in GIS verwaltet wurden und bei der Rutschungserkennung zum Einsatz kamen, werden in Tab. 1 aufgeführt. Jede Zelle hat somit immer den gleichen Parametersatz.

Nach dem Export der Daten wurde mit einem Perl-Skript die Rutschungsumgebung einer Auswahl von Parametern miteinbezogen, um deren Einfluss auf die Rutschungsanfälligkeit zu ermitteln. Mit diesem Skript war es nicht nur möglich, die Morphologie in der Umgebung einer Zelle zu berücksichtigen, sondern es konnte zudem die Änderung eines jeden Parameters in der Nachbarschaft berücksichtigt werden. Die Parameter der Nachbarschaft wurden sodann als zusätzliche Parameter in die Rutschungserkennung miteinbezogen.

Tab. 1: Parameter bei der Rutschungserkennung.
Tab. 1: Parameters of soil slip recognition.

Parameter	Einheit	Datentyp	Codierung	Wertebereich
Absolute Höhenlage	[m NN]	metrisch	keine	[469; 1987]
Exposition (2 Parameter)	[°]	metrisch	sin, binär	{0; 1}, {0; 1}
Hangneigung	[°]	metrisch	keine	{0; 90}
Abstand zu Bächen	[m]	metrisch	keine	{0; 2843}
Flowaccumulation	-	metrisch	keine	{0; 577414}
Geologie	-	nominal	binär	{1; 2; ...21}
Lockergesteinsart	-	nominal	binär	{1; 2; ...12}
Lockergesteinsmächtigkeit	[m]	ordinal	binär	{1; 2; 3}
Niederschlagsmenge	[mm]	metrisch	keine	{21; 148}
Verteilung Wald-/Freifläche	-	metrisch	binär	{0; 1}
Waldzustand	-	ordinal	binär	{1; 2; ...6}
Lage gegenüber Bauwerken	-	nominal	binär	{1; 2; 3; 4}
Lage gegenüber Bächen	-	nominal	binär	{1; 2; ...6}

Mit diesem Skript konnte die Nachbarschaft einer Zelle auf zwei Arten erzeugt werden. Zum einen konnten die Zellen, die sich in einer quadratischen Nachbarschaft um eine Ausgangszelle befinden, erfasst werden. Hier war es möglich, einen oder mehrere Parameter der 9er-, 25er- oder 49er-Nachbarschaften (3 x 3, 5 x 5, 9 x 9 Zellen inklusive Ausgangszelle) in die Rutschungserkennung einfließen zu lassen. Zum anderen konnten aber auch die Werte der Zellen entlang des steilsten Gradienten in hangaufwärtiger oder hangabwärtiger Richtung als zusätzliche Parameter erzeugt werden. Die Generierung der Parameter der Nachbarschaft entlang des Gradienten erfolgte unter Verwendung der unterschiedlichen statistischen Funktionen: Maximum, Minimum, Durchschnitt, Varianz. Größere Nachbarschaften als 9er-Nachbarschaften wurden nicht verwendet, da dadurch für jeden Datensatz zu viele Parameter erzeugt worden wären.

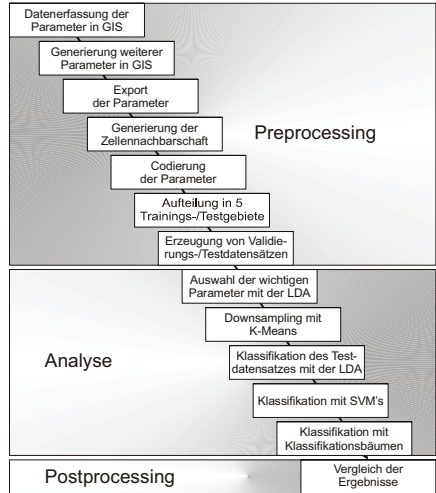


Abb. 6: Arbeitsschritte der Rutschungserkennung mit unterschiedlichen Klassifikationssystemen. Das downsampling mit K-Means wird hier nicht beschrieben.
Fig. 6: Procedure of soil slip recognition using different classification methods (ignoring downsampling with K-Means).

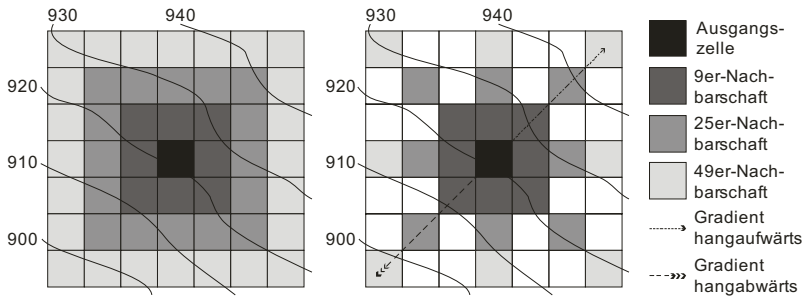


Abb. 7: Erzeugung der Zellennachbarschaft; links: ringförmige Nachbarschaften; rechts: Nachbarschaften entlang des steilsten Gradienten.

Fig. 7: Generation of associated cells within the surrounding neighbourhood. Left: circular neighbourhood; right: neighbourhood along the steepest path.

Am Ende des Preprocessings wurde das gesamte Untersuchungsgebiet in 5 Teilgebiete aufgeteilt sowie die Validierungs- und Testdatensätze erzeugt. Bei der Aufteilung des Modellgebietes wurden drei unterschiedliche Aufteilungsvarianten getestet. Bei der ersten Methode „Einzug“ erfolgte eine Aufteilung nach den Einzugsgebieten der Wildbäche. Bei der zweiten Methode „Alle Höhen“ wurden E-W-ausgerichtete Gebiete erzeugt, die jeweils Flächen aus den Sachsler Bergen sowie den orographisch linken und rechten Hängen des Melchtales enthielten. Bei der dritten Variante „Statist“ wurde das Modellgebiet in ein Raster mit 141 Zellen (500 m²) zerlegt, die anschließend mit einem Zufallsgenerator 5 Klassen zugeordnet wurden (=“Statist“). Dadurch wurden in jedem Trainings-/Testgebiet mehr Informationen aus unterschiedlichen Regionen des gesamten Modellgebietes bei der Analyse herangezogen.

Analyse

Die einzelnen Teilschritte der Analyse werden ebenfalls in Abb. 6 wiedergegeben. Der erste Schritt – die Auswahl der Parameter bzw. die Validierung – wurde mit der LDA durchgeführt und erforderte sehr viel Zeit und Rechenleistung. Vor dem eigentlichen Test mit den unterschiedlichen Klassifikatoren wurde zudem ein downsampling mit K-Means getestet, das keine zufriedenstellenden Ergebnisse lieferte und hier nicht weiter beschrieben wird. Bei der Analyse wurde mit Hilfe der unterschiedlichen Klassifikationsmethoden versucht, ein Zwei- bzw. Drei-Klassenproblem bestmöglich zu lösen. Die folgenden drei Probleme wurden klassifiziert:

- Zwei-Klassenproblem „Kante-Nichtrutschung“:
Nur die oberen Rutschungskanten, die als Auslöser der Rutschung angesehen werden, sowie die Zellen, die nicht von der Rutschung betroffen waren (Nichtrutschungen), wurden klassifiziert.
- Drei-Klassenproblem „Kante-Fläche-Nichtrutschung“:
Neben den oberen Rutschungskanten wurden auch die unterhalb liegenden Rutschungsflächen und die Nichtrutschungen klassifiziert.
- Zwei-Klassenproblem „Rutschung-Nichtrutschung“:
Die Rutschungskanten und Rutschungsflächen sind zusammen in einer Klasse. In der anderen Klasse befinden sich die Nichtrutschungen.

Es wurden zwei Validierungsphasen mit unterschiedlichen Datensätzen durchgeführt, wobei in der Validierungsphase 1 das Verhältnis der verwendeten Klassenstärken (z.B. Kante : Nichtrutschung = 1 : 10) unausgeglichene war. Jeder Validierungsdurchlauf bestand aus 5 Einzeltests, wobei jeweils in vier der fünf Gebieten der Klassifikator trainiert und im fünften getestet wurde. Tab 2. gibt das Ergebnis der ersten Validierungsphase mit unterschiedlichen Parameterkonstellationen wieder, bei der das Zwei-Klassenproblem „Kante-Nichtrutschung“ untersucht wurde.

Tab. 2: Erkennungsraten bei der Parameteridentifikation mit der LDA in der Validierungsphase 1 zur Lösung des Zwei-Klassenproblems *Kante-Nichtrutschung* (Verhältnis Kanten : Nichtrutschungen = 1:10).

Tab. 2: Results of validation of phase 1 for two-cases-problem *upper-edge/non-soil-slip* using LDA (ratio *upper-edge : non-soil-slip* = 1 : 10).

Modell		Erkennungsraten (%)			
Gebietsaufteilung	Anzahl Parameter	gesamt	klassenweise gemittelt	Kanten	Nichtrutschungen
<i>Einzug</i>	58	74,3	65,1	60,7	69,4
<i>Einzug</i>	962	79,9	64,6	50,4	78,7
<i>Einzug</i>	1200	80,6	65,1	50,6	79,6
<i>Einzug</i>	3200	82,7	62,9	42,1	83,6
<i>Alle Höhen</i>	58	71,1	70,9	70,7	71,0
<i>Alle Höhen</i>	962	77,9	70,5	61,7	79,2
<i>Alle Höhen</i>	1200	78,6	70,6	61,1	80,1
<i>Alle Höhen</i>	3200	78,9	70,0	59,2	80,7
<i>Statist</i>	58	69,8	71,1	73,1	69,1
<i>Statist</i>	962	77,9	69,4	63,1	75,9
<i>Statist</i>	1200	78,5	72,8	66,0	79,6
<i>Statist</i>	3200	79,4	72,5	60,6	84,4

Neben den in Tab. 2 dargestellten Ergebnissen der Validierungsphase 1 wurden auch das Drei-Klassenproblem „Kante-Fläche-Nichtrutschung“ und das Zwei-Klassenproblem „Rutschung-Nichtrutschung“ klassifiziert. Zusammenfassend konnten aus diesen Validierungsergebnissen folgende Erkenntnisse gewonnen werden:

- Die Güte des Ergebnisses ist stark von der Gebietsaufteilung abhängig. Bei der Variante *Einzug* findet ein overfitting beim Training statt, was schlechte Ergebnisse während des Tests nach sich zieht. Die mit dem Gebietsaufteilungstyp *Statist* entwickelten Modelle lösen alle drei Klassifikationsprobleme am besten. Der Grund hierfür liegt bei der besseren Generalisierungsfähigkeit der Modelle.
- Die Tests des Drei-Klassenproblems haben gezeigt, dass eine Unterscheidung der Rutschungsklassen in Rutschungsflächen und Rutschungskanten zu keinen besseren Ergebnissen führt. Speziell bei den Flächen der Rutschungen wurden extrem niedrige Erkennungsraten erzielt. Es kann daher angenommen werden, dass die Objekte der Kantenklasse und der Flächenklasse sich sehr ähnlich sind, da Falschzuordnungen zwischen den beiden Klassen häufig auftreten.
- Bei der Lösung eines Zwei-Klassenproblems mit einer übergewichtigen Klasse bevorzugt der Klassifikator bei der Steigerung der Komplexität (mehr Parameter) eine Zuordnung zur übergewichtigen Klasse. Würde man daher das Verhältnis der Objekte der beiden Klassen noch weiter reduzieren - z.B. von 1:10 auf 1:100 - würden noch höhere Gesamterkennungsraten erzielt werden können, die jedoch nicht aussagekräftig wären, da das Zwei-Klassenproblem nicht besser gelöst werden würde, sondern nur die stark übergewichtige Klasse höhere Erkennungsraten aufweisen würde.

Aus diesen Erkenntnissen wurden in der Validierungsphase 2 nur noch die Zwei-Klassenprobleme mit Datensätze klassifiziert, die nach dem Gebietsaufteilungstyp *Statist* aufgeteilt wurden und ausgeglichene Klassenstärken auswiesen. Zudem wurde das Verfahren der Parameterauswahl verändert. Während in der Phase 1 eine ständige Reduktion der Parameter vorgenommen wurde, wurde in der Phase 2 die minimale Parameteranzahl gesucht, mit der eine maximale Erkennungsrate erzielt werden kann.

Das beste Modell (*Val-326*) zur Lösung des Zwei-Klassenproblems *Kante-Nichtrutschung* berücksichtigt folgende Parameter:

- Minimale Hangneigung entlang des Gradienten in hangaufwärtiger Richtung im Bereich von 50 m
- Minimale Hangneigung entlang des Gradienten in hangabwärtiger Richtung im Bereich von 50 m
- Abstand zu Bächen
- Niederschlagsmenge
- Geologie entlang des Gradienten in hangaufwärtiger Richtung im Bereich von 40 m
- Lockergesteinsart in der 9er-Nachbarschaft
- Lockergesteinsmächtigkeit in der 9er-Nachbarschaft
- Waldschaden entlang des Gradienten in hangaufwärtiger Richtung im Bereich von 40 m
- Waldschaden entlang des Gradienten in hangabwärtiger Richtung im Bereich von 20 m

Für die Lösung des Zwei-Klassenproblems *Rutschung-Nichtrutschung* verwendet das beste Modell (*Val-401*) folgende Parameter:

- Minimale Hangneigung entlang des Gradienten in hangaufwärtiger Richtung im Bereich von 45 und 50 m
- Minimale Hangneigung entlang des Gradienten in hangabwärtiger Richtung im Bereich von 50 m
- Abstand zu Bächen
- Niederschlagsmenge in der 9er-Nachbarschaft
- Geologie entlang des Gradienten in hangaufwärtiger Richtung im Bereich von 40 m
- Lockergesteinsart in der 9er-Nachbarschaft
- Lockergesteinsart entlang des Gradienten in hangaufwärtiger Richtung im Bereich von 20 m
- Lockergesteinsmächtigkeit in der 9er-Nachbarschaft
- Waldschaden entlang des Gradienten in hangaufwärtiger Richtung im Bereich von 40 m
- Waldschaden entlang des Gradienten in hangabwärtiger Richtung im Bereich von 20 m

Auf die einzelnen Ergebnisse der Tests mit den unterschiedlichen Klassifikatoren wird hier nicht weiter eingegangen; eine Gegenüberstellung liefert jedoch das Postprocessing.

Postprocessing

In Tab. 3 wird eine Zusammenfassung der bei den Testreihen erzielten Ergebnisse zur Lösung des Zwei-Klassenproblems *Kante-Nichtrutschung* wiedergegeben, bei denen die unterschiedlichen Klassifikatoren mit Modellen getestet wurden, die während der Parameteridentifikation bzw. Validierung mit der LDA entwickelt wurden.

Die dargestellte Tab. 3. verdeutlicht vor allem, dass kein großer Unterschied bei den Gesamterkennungsraten zwischen den einzelnen Modellen, die mit den unterschiedlichen

Klassifikatoren getestet wurden, besteht. Die höchsten Gesamterkennungsraten (74,4 %) wurden mit den Modellen erreicht, die mit den Klassifikationsbäumen getestet wurden und nur eine geringe Parameteranzahl verwenden. Etwas geringere Gesamterkennungsraten wurden mit Modellen erzielt, die wesentlich mehr Parameter berücksichtigen und mit der LDA oder den SVMs getestet wurden.

Tab. 3: Vergleich der Erkennungsraten der Testdurchläufe mit verschiedenen Klassifikatoren zur Lösung des Zwei-Klassenproblems *Kante-Nichtrutschung* (* KA = Kante, NR = Nichtrutschungen).

Tab. 3: Recognition rates of test phase for two-cases-problem *upper-edge/non-soil-slip* using different classification methods (* KA = upper edge, NR = non-soil-slip).

Modell		Erkennungsraten(%)					
Name	Anzahl Param.	LDA		SVM		Klassifikationsbäume	
		gesamt	KA/NR*	gesamt	KA/NR*	gesamt	KA/NR*
<i>Val-34</i>	34	73,0	78,5 / 64,3	-	- / -	74,3	77,9 / 69,0
<i>Val-58</i>	58	73,1	78,2 / 64,1	70,9	78,5 / 61,2	74,3	74,2 / 73,1
<i>Val-326</i>	326	74,0	79,0 / 65,8	71,8	80,0 / 61,0	74,4	74,3 / 72,3
<i>Val-339</i>	339	73,8	78,2 / 65,9	-	- / -	73,7	75,5 / 70,5
<i>Val-391</i>	391	74,0	79,0 / 65,8	-	- / -	73,5	75,3 / 69,9
<i>Val-962</i>	962	73,7	79,8 / 65,8	73,9	77,9 / 66,2	73,7	75,0 / 71,5
<i>Val-1200</i>	1200	71,1	74,0 / 65,4	73,7	80,7 / 63,7	74,1	69,7 / 76,1
<i>Val-3200</i>	3200	70,7	72,5 / 66,8	71,8	79,2 / 61,8	73,1	71,2 / 73,4

Tab. 4: Vergleich der Erkennungsraten der Testdurchläufe mit verschiedenen Klassifikatoren zur Lösung des Zwei-Klassenproblems *Rutschung-Nichtrutschung* (* KA = Kante, NR = Nichtrutschungen).

Tab. 4: Recognition rates of test phase for two-cases-problem *soil-slip/non-soil-slip* using different classification methods (* RU = upper edge, NR = non-soil-slip).

Modell		Erkennungsraten(%)					
Name	Anzahl Param.	LDA		SVM		Klassifikationsbäume	
		gesamt	RU/NR*	gesamt	RU/NR*	gesamt	RU/NR*
<i>Val-58</i>	58	72,9	79,5 / 63,8	71,2	78,1 / 62,8	72,1	58,8 / 83,3
<i>Val-326</i>	326	74,8	80,9 / 67,0	70,7	73,4 / 64,3	67,2	58,2 / 73,7
<i>Val-391</i>	391	75,2	81,1 / 67,7	71,7	80,5 / 60,4	66,3	57,2 / 73,3
<i>Val-401</i>	401	75,4	81,0 / 68,1	72,1	78,8 / 62,6	71,4	59,2 / 81,7
<i>Val-962</i>	962	74,4	77,1 / 69,7	73,5	79,9 / 64,3	72,9	61,8 / 81,6
<i>Val-1200</i>	1200	73,6	76,0 / 69,5	73,0	79,5 / 63,6	71,5	57,9 / 82,5

Ein Vergleich der erzielten Ergebnisse zur Lösung des Zwei-Klassenproblems *Rutschung-Nichtrutschung* wird in Tab. 4 dargestellt. Gegenüber der Lösung des Zwei-Klassenproblems *Kante-Nichtrutschung* wurde z.T. eine Verbesserung bei den Gesamterkennungsraten erreicht. Zudem zeigte sich, dass bei der Lösung des Zwei-Klassenproblems *Rutschung-Nichtrutschung* die Unterschiede zwischen den einzelnen Klassifikatoren viel stärker ausgeprägt sind. Während mit den schlechtesten Modellen nur knapp zwei Drittel der Zellen richtig zugeordnet wurden, erfolgte bei den besten Modellen eine richtige Zuordnung bei über drei Viertel der untersuchten Zellen. Die besten Ergebnisse bei den Gesamterkennungsraten (75,4 %) liefern hier die Testdurchläufe, die mit der LDA vollzogen wurden.

Zusammenfassend belegen die Ergebnisse der Klassifikationen, dass es möglich ist, gute Lösungen für die zwei untersuchten Zwei-Klassenprobleme *Kante-Nichtrutschung* und *Rutschung-Nichtrutschung* zu finden. Es sollte jedoch generell darauf geachtet werden, mit welcher Fragestellung das Ziel der Untersuchungen am besten erreicht werden kann. Hierzu hat es sich empfohlen, Verwechslungsmatrizen (%-Satz richtig und falsch erkannt) bei der Analyse aufzuzeichnen und anschließend auszuwerten. Sind möglichst hohe Erkennungsraten

in der Rutschungsklasse (über 80 %) sowie hohe Gesamterkennungsraten das primäre Ziel, so sollte eher das Zwei-Klassenproblem *Rutschung-Nichtrutschung* mit einem Modell mit der LDA klassifiziert werden, welches auch Parameter aus der Zellenumgebung berücksichtigt.

Will man dagegen nicht nur die Rutschungskanten, sondern auch die Nichtrutschungen gut erkennen, liefern die Klassifikationsbäume die besten Ergebnisse. Hier werden die beiden Klassen, ohne Berücksichtigung der Zellenumgebung, nahezu gleich gut erkannt. Die Gesamterkennungsrate liegt jedoch mit 74,3 % etwa 1 % unter der Erkennungsrate der Linearen Diskriminanzanalyse für die Lösung des Zwei-Klassenproblems *Rutschung-Nichtrutschung*.

Der Vergleich dieser Ergebnisse mit den Modellierungsergebnissen von Liener (2000) unter Verwendung des SLIDISP-Modells zeigt auf, dass nicht hohe Erkennungsraten die Güte eines Modells bestimmen, sondern die Anwendung des Modells zur Naturgefahrenbeurteilung stets ein wichtiger Aspekt sein sollte. Hier sollte bedacht werden, dass eine Falschzuordnung einer Zelle, egal ob Rutschungs- oder Nichtrutschungszelle, zu einem materiellen oder wirtschaftlichen Schaden führen kann, oder im Extremfall einen Personenschaden nach sich zieht.

Die Verwendung der sehr komplexen und zeitaufwendigen SVMs führte zu keinen besseren Ergebnissen. Dies liegt eventuell daran, dass die Auswahl der Parameter mit der LDA erfolgte und die Tests mit diesen Parameterkombinationen durchgeführt wurden.

In vielen Fällen hat sich gezeigt, dass eine hohe Anzahl von verwendeten Parametern nicht automatisch zu besseren Ergebnissen führt; dies hat lediglich eine Überanpassung des Systems im Training und schlechte Erkennungsraten bei den Tests zur Folge.

LITERATUR

- BWW (Bundesamt für Wasserwirtschaft) (1998): Ereignisdokumentation Sachseln - Unwetter vom 15. August 1997. Studienbericht 8, Bundesamt für Wasserwirtschaft, Bern.
- Cruden, D., Varnes, D. (1996): Landslide Types and Processes. In Turner, A. and Schuster, R., editors, Landslides: Investigation and mitigation, number 247, pages 36–75. National Research Council, Transportation Research Board, Washington, D.C.
- Fernandez-Steeger, T. (2002): Erkennung von Hangrutschsystemen mit Neuronalen Netzen als Grundlage für Georisikoanalysen. Dissertation, Universität Karlsruhe, Karlsruhe.
- Hamberger, M. (2000): Geologisch-geotechnische Untersuchungen und Analyse von Hangbewegungen anlässlich von Starkniederschlägen im Melchtal/Schweiz. Unveröffentlichte Diplomarbeit, Lehrstuhl für Angewandte Geologie, Universität Erlangen-Nürnberg, Erlangen.
- Kienholz, H. (1977): Kombinierte geomorphologische Gefahrenkarte 1:10.000 von Grindelwald. Geographica Bernensia, G4, Bern.
- Liener, S. (2000): Zur Feststofflieferung in Wildbächen. Geographica Bernensia, G64, Bern.
- Moser, M. (1980): Zur Analyse von Hangbewegungen in schwachbindigen bis rolligen Lockergesteinen im alpinen Raum anlässlich von Starkniederschlägen. In Internationales Symposium Interpraevent 1980 - Bad Ischl, volume 1, pages 121–148.
- Rickli, C. (2001): Vegetationswirkungen und Rutschungen - Untersuchungen zum Einfluss der Vegetation auf oberflächennahe Rutschprozesse anhand der Unwetterereignisse in Sachseln OW am 15. August 1997. Technical report, Eidgenössische Forschungsanstalt WSL und Bundesamt für Umwelt, Wald und Landschaft, Bern.
- Zanetti, A., Enz, R., Heck, P., Green, J., Suter, S. (2004): Sigma – Natur- und Man-made-Katastrophen im Jahr 2003. Ausgabe 1/2004, Swiss Re, Zürich.