

Der Einsatz unterschiedlicher Klassifikationssysteme zur Rutschungserkennung Implementations of different classification methods for soil slip recognition

M. Hamberger, M. Moser, E. Nöth, M. Levit, R. Deventer¹

Zusammenfassung

Bei der Unwetterkatastrophe von Sachseln/Schweiz entstanden am 15. August 1997 mehr als 700 flachgründige Rutschungen, die in den Folgejahren evaluiert und analysiert wurden. Die Ergebnisse der untersuchten Rutschungen zeigen, dass verschiedene Parameter wie z.B. morphologische, geotechnische oder hydrologische Faktoren einen entscheidenden Einfluss auf die Rutschungsanfälligkeit eines Hanges besitzen. Neben den Parametern der Rutschungsfläche wurden auch Parameter der Anbruchsumgebung untersucht, die sich negativ auf die Hangstabilität auswirkten. Mit drei verschiedenen Klassifikationsverfahren – Lineare Diskriminanzanalyse (LDA), Klassifikationsbäume und Support Vector Machines (SVMs) – wurde versucht, Parameterkonstellationen herauszufinden, die auf das Auftreten von Rutschungen hinweisen. Die Ergebnisse der Mustererkennung belegen, dass es möglich ist, einen Großteil der Rutschungen (82 %) zu erkennen, wobei mit herkömmlichen Klassifikatoren (LDA, Klassifikationsbäume) bessere Ergebnisse erzielt werden als mit den sehr komplexen und in vielen Bereichen verwendeten SVMs. Neben den Chancen der automatischen Rutschungserkennung mit Klassifikationssystemen werden auch deren Grenzen aufgezeigt.

Keywords: Hangbewegungen, Rutschungserkennung, GIS, Klassifikationssysteme

Abstract

A thunderstorm in Sachseln/Switzerland that had a rainfall total of 150 mm over a 2 hours period was recorded on August 15th 1997 and resulted in more than 700 soil slips. These soil slips were evaluated and analysed in the following years by the Swiss Federal Institute for Forest, Snow and Landscape Research and the Department of Geology at the University of Erlangen in Germany. The results of the analysed soil slips highlighted that different factors, e.g. morphological, geotechnical or hydrological factors, have a crucial influence on the slope stability. Besides the parameters describing soil slip scars, characteristics of the surrounding area were analysed and used to identify instable areas. GIS was not only used for digitising but also for managing data and creating new parameters that may have an influence on slope stability. Using three different classification methods – Linear Discriminant analysis (LDA), classification trees and Support Vector Machines (SVMs) – the set of parameters that best describe the occurrence of soil slips were investigated, while at the same time attempting to find the simplest combination thereof. Analyses were conducted for the case where only cells of the upper edge of the soil slip and non-soil-slip cells were considered, as well as the case where the entire soil slip scars were incorporated. The results of the soil slip recognition demonstrated that it is possible to recognize most of the soil slips (82 %), but the overall classification rate of both classes (soil slips vs. non-soil-slips) was low (75 %). After extensive research with different classifiers, it was concluded that the ratio of soil-slip cells to non-soil-slip cells during the classification must be balanced. The inclusion of adjacent cells of soil slip scars improved the results by 3 %. The use of the modern and time-consuming SVMs or classification trees did not result in better recognition rates.

Keywords: Landslides, soil slips recognition, GIS, classification methods

1 Einleitung

Statistiken über das weltweite Auftreten von Naturkatastrophen zeigen deutlich, dass deren Anzahl in den letzten 50 Jahren stark angestiegen ist. Betrachtet man die zehn schwersten Naturkatastrophen der Schweiz, so wurden bei sieben Ereignissen 273 Menschen durch Massenbewegungen und Lawinen getötet. Die Zahl der durch diese Katastrophen betroffenen Personen lag bei über 3500. (EM-DAT 2003)

Bei der Unwetterkatastrophe von Sachseln/Schweiz haben sich infolge sehr intensiver Niederschläge mehr als 700 flachgründige Rutschungen entwickelt, deren

Material nach dem Transport in den Gerinnen zu großen Übermürungen in den Siedlungsbereichen führte. Der Gesamtschaden wurde mit ca. 100 Mio. CHF beziffert. Da der Siedlungsdruck in den Alpen in den letzten Jahrzehnten immer größer wurde, wird es zunehmend wichtiger, Bereiche auszuweisen, die von Naturgefahren bedroht werden. Seit 1972 werden in der Schweiz Gefahrenbereiche ausgeschieden, die nicht zur Besiedelung freigegeben werden. Die Erstellung von Gefahrenzonenplänen bzw. Gefahrenhinweiskarten, als Produkt der Gefahrenbeurteilung, ist jedoch mit einem hohen Zeitaufwand verbunden.

¹Dipl.-Geol. Maik Hamberger, Prof. Dr. Michael Moser, Lehrstuhl für Angewandte Geologie, Universität Erlangen-Nürnberg, Schlossgarten 5, 91054 Erlangen, mhamberg@geol.uni-erlangen.de;
Dr.-Ing. Elmar Nöth, Dipl.-Ing. Michael Levit, Dr.-Ing. Rainer Deventer, Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg, Martensstr. 3, 91058 Erlangen, noeth@informatik.uni-erlangen.de

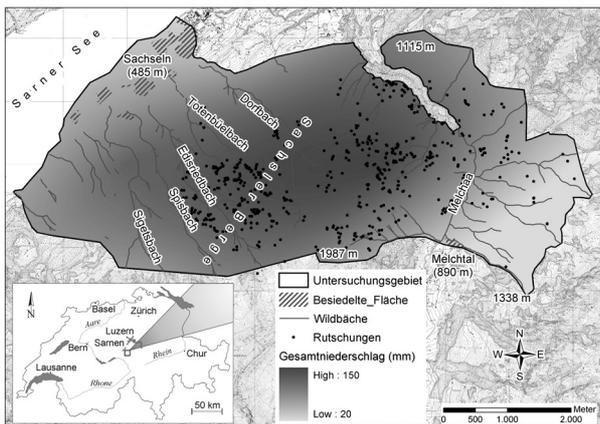


Fig. 1: Überblick über die geografische Lage, Lage der Rutschungen und die Gesamtniederschlagsmengen.
Fig. 1: Overview of the geographical position, position of soil slips and the amount of precipitation.



Fig. 2: Flachgründige Rutschungen vom 15.08.1997 im Melchtal.
Fig. 2: Soil slips of Melchtal from August 15th, 1997.

Die Recherche der Literatur der letzten Jahre zeigt, dass das Interesse immer größer wird, ein Verfahren zu entwickeln, welches effizient und mit einem möglichst geringen Aufwand auf die einzelnen Typen von Naturgefahren hinweist. Dazu werden vermehrt Methoden aus dem Bereich der Mustererkennung und der Künstlichen Intelligenz herangezogen, die sich bereits auf anderen Gebieten, wie z.B. der Spracherkennung, Texterkennung oder aber in der Medizin bewährt haben.

2 Evaluerte Daten

Nach dem Unwetterereignis vom 15. August 1997 wurden umfangreiche Analysen innerhalb unterschiedlicher Projekten durchgeführt. Die Auswertung der Radardaten der Starkniederschläge erfolgte durch das Bundesamt für Wasserwirtschaft (BWW 1998). Der Starkregen vom 15.08.1997 kann folgendermaßen charakterisiert werden:

- Es gab insgesamt zwei extrem intensive Niederschlagsperioden – 17 bis 18 Uhr und 19.20 bis 21.20 Uhr.
- Die Spitzenintensitäten der Niederschläge lag bei 90 mm/h, der Durchschnittswert bei 70 mm/h.
- Die Gesamtniederschlagsmengen lagen bei maximal 150 mm in den oberen Einzugsgebieten der Wildbäche der Sachslar Berge; in Sachseln fielen insgesamt nur 50 mm (Fig. 1).

Im Rahmen des Projektes *Vegetationswirkungen und Rutschungen* evaluierte die Eidgenössische Forschungsanstalt WSL (Birmensdorf/Schweiz) die Rutschungen der Sachslar Berge (RICKLI 2001). Der Lehrstuhl für Angewandte Geologie der Universität Erlangen erfasste die Rutschungsdaten im Bereich des unteren Melchtales (Lage und Art der Lockergesteinsrutschungen siehe Fig. 1, Fig. 2, HAMBERGER 2000).

Bei diesen Untersuchungen wurden folgende Merkmale der Rutschungen bzw. Rutschungsbereiche evaluiert:

- Lage der Rutschungen
- Rutschungsmechanismus
- Geometrie der Rutschungen
- Geomorphologie des Anbruchhanges
- Geologische und geotechnische Situation
- Hydrogeologie im Rutschungsbereich
- Art und Zustand der Vegetation

2.1 Ergebnisse der Evaluation

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 (nach KIENHOLZ 1996) 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 stratigrafischen Einheiten mit einer kalkig-mergeligen Ausbildung (Diphyoideskalk/-mergel, Kieselkalk, Drusbergschichten) weisen die höchsten Rutschungsaktivitäten auf.

Als Schlussfolgerung der Evaluation konnte festgehalten werden, dass die Rutschungen vom 15.08.1997 nicht durch einen Parameter, sondern eine Kombination vieler Faktoren ausgelöst wurden.

3 Rutschungserkennung

In den Arbeiten von MAYORAZ et al. (1996), FERNANDEZ-STEEGER (2002) oder NEAUPANE & ACHET (2004) wurden Künstliche Neuronale Netze herangezogen, um Hangbewegungen zu erkennen bzw. vorherzusagen. CHUNG (1995) hat statistische Regressionsmethoden bei der Gefahrenbeurteilung von rutschgefährdeten Hängen eingesetzt.

3.1 Klassifikatoren

Die Rutschungserkennung wurde mit drei unterschiedlichen Klassifikationsmethoden durchgeführt. Bei allen Methoden wurde versucht, Parameterkonstellationen zu finden, die auf Rutschungen hinweisen. Jede Zelle des Untersuchungsgebietes wurde aufgrund seiner Parameter als Rutschung oder „Nicht-Rutschung“ klassifiziert.

Lineare Diskriminanzanalyse (LDA)

Bei der linearen Diskriminanzanalyse wird eine lineare Transformation der Elemente eines Merkmalvektors (rutschungsrelevante Parameter) gesucht, welche die Merkmale auf gut separierbare Bereiche in einem ein-dimensionalen Raum abbildet. Bei der Zuweisung einer Zelle zu der Klasse Rutschung (oder Kante und Fläche, s.u.) oder Nicht-Rutschung wurden alle Merkmale gleichzeitig benutzt. Die bei der Klassifikation erzeugte Diskriminanzfunktion kann folgendermaßen dargestellt werden:

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

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 war es, die Koeffizienten so zu ermitteln, dass der input beide Klassen – Rutschung und Nicht-Rutschung – möglichst gut trennt.

Klassifikationsbäume

Bei der Klassifikation von Rutschungen mit Klassifikationsbäumen wird ein Baum generiert, dessen Knoten mit den Attributen des Datensatzes assoziiert sind und dessen Kanten mit den möglichen Werten der Attribute belegt sind. Die Blattknoten des Baumes geben die Zuordnung zu der Klasse Rutschung oder Nicht-Rutschung an. Im Gegensatz zu der LDA verläuft die Zuordnung zu einer Klasse somit sequentiell. Ein vereinfachtes Schema eines Klassifikationsbaumes, der bei der Rutschungserkennung eingesetzt werden kann, kommt in Fig. 3 zur Darstellung.

Support Vector Machines (SVMs)

Support Vector Machines sind überwachte maschinelle Lernverfahren, die zur binären Klassifikation entwickelt wurden, mittlerweile aber auch multiclass-Aufgaben übernehmen. Innerhalb des Trainingsdatensatzes wird eine Hyperebene so im Raum platziert, so dass die Klassen Rutschung und Nicht-Rutschung optimal voneinander getrennt werden. Dabei wird die Hyperebene so gelegt, dass die Zone um die Hyperebene, die keinen

Trainingsdatenpunkt enthält, maximal wird. Da eine lineare Trennbarkeit durch Ausreißer meist unmöglich ist, wird versucht, mit einer Kernel Funktion in einem höher- (oder sogar unendlich-)dimensionalen Raum die Klassen Rutschung und Nicht-Rutschung mit der Hyperebene voneinander zu trennen. Eine ausführliche Beschreibung der Arbeitsweise von SVMs geben BURGES (1998) oder SMOLA et al. (2000).

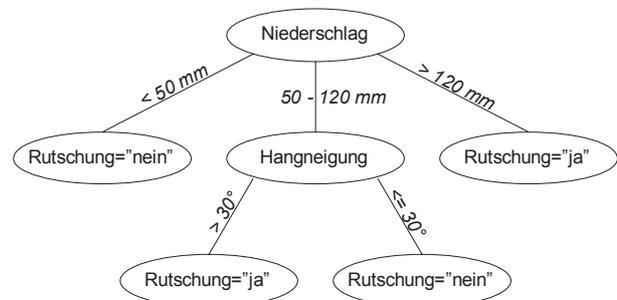


Fig. 3: Vereinfachtes Schema eines Klassifikationsbaumes zur Rutschungserkennung.

Fig. 3: Simplified scheme of a classification tree used for soil slip recognition.

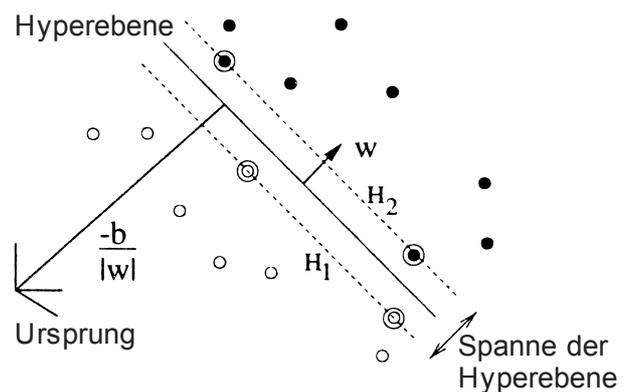


Fig. 4: Lineare trennende Hyperebene eines separierbaren Falles einer SVM (w = Normale zur Hyperebene, $-b/|w|$ = rechtwinkliger Abstand der Hyperebene zum Ursprung, H_1 , H_2 = Begrenzungen der Spanne um Hyperebene; modifiziert nach BURGES 1998).

Fig. 4: Linear separating hyperplane for the separable case (w = normal to hyperplane, $-b/|w|$ = perpendicular distance from hyperplane to origin, H_1 , H_2 = boundary of margin at hyperplane; modified after BURGES 1998).

3.2 Methodik

Die Rutschungserkennung kann in drei Prozesse unterteilt werden – das Preprocessing, die Analyse und das Postprocessing (Fig. 5).

Preprocessing

Das Preprocessing wurde vorwiegend innerhalb eines Geografischen Informationssystems (GIS) durchgeführt. Bei der Datenerfassung konnte zum einen auf die Daten der o.g. Evaluation zurückgegriffen werden, zum anderen wurden zusätzliche Daten wie z.B. Art und Zustand der Vegetation oder die Ergebnisse der geotechnischen Aufnahmen miteinbezogen.

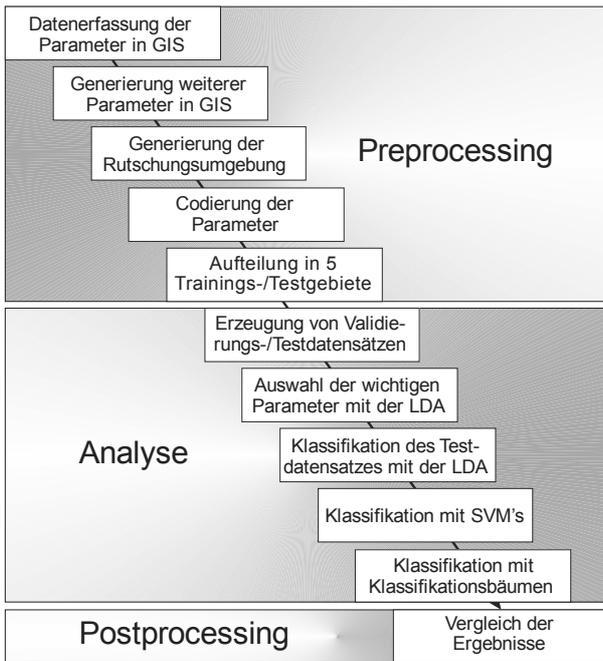


Fig. 5: Vorgehensweise bei der Rutschungserkennung mit unterschiedlichen Klassifikatoren der 2. Testphase. Fig. 5: Method of soil slip recognition of second test phase, using different classifiers.

Zusätzlich wurden in GIS Daten generiert, z.B. Abstand zu Bächen, die bei der Rutschungserkennung berücksichtigt wurden.

Bei der Rutschungserkennung ist es notwendig, jeden Parameter nahezu flächendeckend für das Untersuchungsgebiet zur Verfügung zu haben. Einige rutschungsrelevante Parameter, wie z.B. die spezielle Morphologie im Rutschungsumfeld, konnten nicht berücksichtigt werden, da diese bei der Evaluation zwar erhoben wurden, aber in GIS nicht flächendeckend generiert werden konnten.

Alle verwendeten Parameter wurden in GIS in Rasterdaten mit einer Zellgröße von 5 x 5 m umgewandelt. Das abflussgerechte Höhenmodell wurde in ArcGIS mit *topogrid* aus Höhendaten (Linien- und Punktdaten) entwickelt. Die Kalibrierung der erzeugten Höhendaten sowie der Expositionen erfolgte mittels der 553 Rutschungen. Die besten Übereinstimmungen wurden mit einem Höhenmodell erreicht, das eine anfängliche Rasterweite von 2 x 2 m aufwies und in einem Folgeschritt auf 5 x 5 m vergrößert wurde. Auf diesem Höhenmodell basierend wurden die Exposition, der Abstand zum Bach sowie die *flowaccumulation* - Anzahl der umliegenden Zellen, die eine Zelle durchfließen - erzeugt. Die Daten, die in GIS verwaltet wurden und bei der Rutschungserkennung zum Einsatz kamen, werden in Tab. 1 aufgeführt.

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. Dabei konnten zum einen alle Zellen, die eine Zelle umgeben, zum anderen all die Zellen entlang des steilsten Gradienten in hang-

Tab. 1: Parameter bei der Rutschungserkennung (Skalierung: m=metrisch, o=ordinal, n=nominal). Tab. 1: Parameter of soil slip recognition (scale: m=metric, o=ordinal, n=nominal).

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

aufwärtiger oder hangabwärtiger Richtung berücksichtigt werden. Z.B. wurden bei einer 9er-Nachbarschaft alle Zellen, welche die Ausgangszelle umgeben, als neue Parameter generiert. Die Zellen entlang des Gradienten konnten unter Berücksichtigung verschiedener Funktionen – Maximum, Minimum, Durchschnitt, Varianz – ebenfalls als neue Parameter hinzugezogen werden. Größere Nachbarschaften als 9er-Nachbarschaften wurden nicht verwendet, da dadurch für jeden Datensatz zu viele Parameter erzeugt worden wären (Fig. 4).

In SPSS wurden anschließend alle Daten eingelesen und codiert. Generell konnten für die Klassifikation drei unterschiedliche Datentypen herangezogen werden – metrische (kontinuierliche), ordinal skalierte (diskontinuierliche mit Reihenfolge) und nominal skalierte (diskontinuierliche). Da sowohl die LDA als auch SVMs nicht mit nominal skalierten Variablen umgehen können, wurden diese Variablen binär codiert (Tab. 1).

Zum Abschluss des Preprocessings musste das Arbeitsgebiet in jeweils fünf Trainings- und Testgebiete eingeteilt werden. Durch die Wahl zweier unterschiedlicher Einteilungsmethoden war es möglich, zu testen, wie robust der jeweils verwendete Klassifikator ist, bzw. wie stark dieser generalisiert. Eine Möglichkeit bestand darin, die Einzugsgebiete der Wildbäche so zusammenzufassen, dass das Modellgebiet in insgesamt fünf Ge-

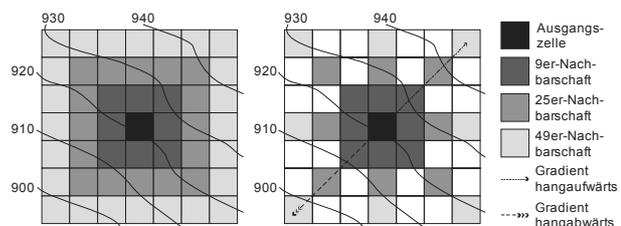


Fig. 6: Erzeugung der Nachbarschaft einer Zelle. Links: Ringförmige Nachbarschaften; rechts: Nachbarschaften entlang des steilsten Gradienten. Fig. 6: Generation of associated cells within the surrounding neighbourhood. Left: circular neighbourhood; right: neighbourhood along the steepest path.

biere unterteilt wird (=“Einzug“). Bei der zweiten Methode 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

Bei der durchgeführten Rutschungserkennung wurde mit Hilfe der unterschiedlichen Klassifikationsmethoden versucht, ein Zwei- bzw. Drei-Klassenproblem bestmöglichst zu lösen. Generell wurde die Gesamtfläche des Untersuchungsgebietes (25 km²), wie oben beschrieben, in ca. 1 Mio. Zellen (Datensätze) zerlegt. 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 Nicht-rutschungen klassifiziert (hier nicht berücksichtigt).

Zwei-Klassenproblem „Rutschung-Nichtrutschung“:

Die Rutschungskanten und Rutschungsflächen sind zusammen in einer Klasse. In der anderen Klasse befinden sich die Nichtrutschungen.

Jeder Testdurchlauf besteht aus 5 Einzeltests, wobei jeweils in vier der fünf Gebiete der Klassifikator trainiert und im fünften getestet wurde.

Testphase 1:

Die ersten Tests des Zwei-Klassenproblems *Kante-Nichtrutschung* wurden mit einer Stichprobe von ca. 20.000 Datensätzen durchgeführt, wobei das Verhältnis von Kanten-Datensätzen zu Nichtrutschungs-Datensätzen bei 10 zu 1 lag. Die Tab. 2 zeigt die Ergebnisse der Rutschungserkennung mit insgesamt 1200 Parametern (nominal skalierte Parameter sind binär codiert → hohe Anzahl) an, die für die Gebietsaufteilungen „Einzug“ und „Statist“ durchgeführt wurden.

Tab. 2: Ergebnisse Zwei-Klassenproblems *Kante-Nichtrutschung* (Verhältnis Kante(1)/Nichtrutschungen (0) = 10/1).

Tab. 2: Results of soil slip recognition (ratio soil-slip/non-soil-slip = 10/1).

Klass. (Anz., Parameter)	Gesamtrate (%)	klassenweise gemittelt (%)	1 auf 1 (%)	0 auf 0 (%)
LDA (1200, „Einzug“)	80,6	66,5	50,6	82,4
SVM (1200, „Einzug“)	82,3	-	-	-
LDA (1200, „Statist“)	78,5	72,8	66,0	79,6
SVM (1200, „Statist“)	80,2	-	-	-

Die Tab. 2 zeigt folgende Probleme der Rutschungserkennung, obwohl Gesamterkennungsraten von bis zu 82,3 % erzielt werden konnten:

- Die Gesamterkennungsraten sind hoch, aber es werden besser Nichtrutschungen (0 auf 0) erkannt, als Rutschungen (1 auf 1).
- Die Erkennungsraten für Rutschungen sind nur geringfügig höher als bei einer zufälligen Auswahl.
- Bei der Gebietsaufteilung „Einzug“ generalisiert der Klassifikator schlechter als bei „Statist“ → Klassifikator lernt Trainingsdaten auswendig und wendet die Parameterauswahl im Testgebiet schlecht an (*overfitting*).

Aufgrund dieser Ergebnisse wurde die Analytik der Rutschungserkennung verändert. Bei der Rutschungserkennung muss das Verhältnis der Rutschungen zu den Nichtrutschungen beim Training in etwa ausgeglichen sein. Des Weiteren wurde nur noch mit der Gebietsaufteilung „Statist“ klassifiziert, da nur diese vernünftig klassifiziert. Außerdem wurde für die Tests der Auswahl der rutschungsrelevanten Parameter ein Validierungsdatensatz verwendet, der bei späteren Tests nicht inkorporiert wurde, um ein *overfitting* zu vermeiden.

Testphase 2:

Die Validierung für das Zwei-Klassenproblems *Kante-Nichtrutschung* wurde mit der LDA durchgeführt, da diese eine Klassifikation am schnellsten bewerkstelligt und es ermöglicht, die Gewichtung der Parameter zu analysieren.

Die Validierung des Zwei-Klassenproblems unter Einbeziehung der Rutschungsumgebung führte zu folgender Parameterauswahl (326 Parameter codiert):

- Hangneigung
- Minimale Hangneigung in hangaufwärtiger Richtung (50 m)
- Minimale Hangneigung in hangabwärtiger Richtung (50 m)
- Abstand zu Bächen
- Geologie
- Geologie in hangaufwärtiger Richtung (40 m)
- Lockergesteinsart
- Lockergesteinsmächtigkeit
- Lockergesteinsmächtigkeit in 9er-Nachbarschaft
- Niederschlagsmenge
- Waldzustand
- Waldzustand in hangaufwärtiger Richtung (40 m)
- Waldzustand in hangabwärtiger Richtung (20 m)

Nach der Auswahl der Parameter durch die LDA erfolgte die Klassifikation auf der Teststichprobe mit allen drei Klassifikationsmethoden.

Die besten Ergebnisse wurden für das Zwei-Klassenproblems *Kante-Nichtrutschung* mit der LDA erzielt (Gesamterkennungsrate 74,0 %). In Tab. 3 werden die Ergebnisse dieser Testreihe wiedergegeben.

Tab. 3: Ergebnisse Zwei-Klassenproblem *Kante-Nichtrutschung* (Verhältnis Kante(1)/Nichtrutschungen (0) = 1/1).

Tab. 3: Results of soil slip recognition (ratio edge-soil-slip/non-soil-slip = 1/1).

Klass./Anz., Parameter)	Gesamt- rate (%)	klassenweise gemittelt (%)	1 auf 1 (%)	0 auf 0 (%)
LDA / 58	73,1	64,1	78,2	64,1
SVM / 58	71,0	70,0	78,5	61,2
Bäume / 58	73,8	69,9	-	-
LDA / 326	74,0	72,4	79,0	65,8
SVM / 326	71,8	70,6	80,0	61,0
Bäume / 326	73,7	73,0	-	-
LDA / 962	73,7	72,8	79,8	65,8
SVM / 962	73,9	72,0	77,9	66,2
Bäume / 962	73,9	73	-	-

Anschließend wurden die Parameter des Zwei-Klassenproblem *Rutschung-Nichtrutschung* in der LDA ausgewählt. Bei dieser Konstellation wurden die nachfolgenden Parameter als rutschungsfördernd angesehen (336 Parameter codiert):

- Hangneigung
- Minimale Hangneigung in hangaufwärtiger Richtung (Kanten, 40 m, 45 m, 50 m)
- Minimale Hangneigung in hangabwärtiger Richtung (Mulden, 50 m)
- Abstand zu Bächen
- Geologie
- Geologie in hangaufwärtiger Richtung (40 m)
- Lockergesteinsart
- Lockergesteinsart in hangaufwärtiger Richtung (20 m)
- Lockergesteinsmächtigkeit
- Lockergesteinsmächtigkeit der 9er-Nachbarschaft
- Niederschlagsmenge
- Niederschlagsmenge der 9er-Nachbarschaft
- Waldzustand
- Waldzustand in hangaufwärtiger Richtung (40 m)
- Waldzustand in hangabwärtiger Richtung (20 m)

Wiederum wurden die besten Erkennungsraten mit der LDA erreicht. Die Erkennungsrate lag bei 75,4 %, bei den SVMs bei 72,2 %. Die höheren Erkennungsraten lassen folgern, dass eine Trennung zwischen Rutschungskante und Rutschungsflächen zu keinen besseren Ergebnissen führt. Dies unterstützt auch die Analyse der Verwechslungsmatrizen. Diese zeigen, dass generell beide Klassen besser erkannt werden.

Die Ergebnisse der Klassifikation des Zwei-Klassenproblem *Rutschung-Nichtrutschung* der ausgewählten Parameter kommen in Tab. 4 zur Darstellung.

Tab. 4: Ergebnisse Zwei-Klassenproblem *Rutschung-Nichtrutschung* (Verhältnis Rutschung(1)/Nichtrutschungen(0) = 1/1).

Tab. 4: Results of soil slip recognition (ratio soil-slip/non-soil-slip = 1/1).

Klass./Anz., Parameter)	Gesamt- rate (%)	klassenweise gemittelt (%)	1 auf 1 (%)	0 auf 0 (%)
LDA / 58	72,8	72,1	81,6	62,5
SVM / 58	71,1	71,0	78,1	62,8
LDA / 336	75,4	68,0	81,1	68,0
SVM / 326	70,7	68,8	74,3	64,3
LDA / 962	74,4	73,4	77,1	69,7
SVM / 962	73,5	72,2	79,9	64,2

4 Schlussfolgerungen

Die dargestellten Ergebnisse zeigen die Möglichkeiten auf, mit Klassifikationssystemen eine Unterscheidung zwischen Flächen, die von Rutschungen betroffen waren, und Nicht-Rutschungsflächen durchzuführen.

Die Einbeziehung der Rutschungsumgebung führte zu einer Verbesserung von maximal 3 %. Die Validierung innerhalb der LDA ließ erkennen, dass z.B. die Hangneigung in hangauf- und hangabwärtiger Richtung ein starkes Gewicht bei der Klassifikation besitzt.

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 dieser Parameterauswahl durchgeführt wurden. Auch die Verwendung der Klassifikationsbäume zeigen bis jetzt keine besseren Ergebnisse, lassen aber eine gute Interpretation der Ergebnisse zu.

Des Weiteren kann festgehalten werden, dass eine hohe Anzahl von verwendeten Parametern nicht automatisch zu besseren Ergebnissen führt.

Bei der Rutschungserkennung sollte generell Folgendes beachtet werden, um verwertbare Ergebnisse zu erhalten:

- Das Verhältnis zwischen Rutschungs- und Nichtrutschungszellen sollte in etwa ausgeglichen sein, da sonst der Klassifikator zu Gunsten der übergewichteten Klasse klassifiziert, womit dieser ein besseres Ergebnis erzielt. Ansonsten ist nur die klassenweise gemittelte Erkennungsrate aussagekräftig.
- Die Daten, die zur Validierung (Auswahl der Parameter) herangezogen werden, dürfen nicht beim Test verwendet werden.
- Bei der Bewertung der Erkennungsraten sollten immer Verwechslungsmatrizen (%-Satz richtig und falsch erkannt) erstellt und ausgewertet werden.
- Eventuell sollte man vorab entscheiden, ob es wichtiger ist, Rutschungen, Nichtrutschungen oder beides zu erkennen. Diese Entscheidung sollte vorab anhand der Fragestellung gefällt werden. Eventuell wird dabei der Einsatz einer Kostenfunktion sinnvoll.

5 Ausblick

Es wird noch untersucht, in wie weit es möglich ist mit einem Clusteringverfahren (K-Means Verfahren) Regeln aufzustellen, die eine erste Reduktion der Nichtrutschungszellen vornimmt. Diese Möglichkeit würde die verwendete stochastische Auswahl der Nichtrutschungsfälle ersetzen und das Problem klarer definieren.

Weitere Analysen wurden bereits in die Wege geleitet, um die rutschungsrelevanten Parameter neu auszuwählen, da die Ergebnisse der Testphase 1 höhere Erkennungsraten bei den Nichtrutschungen zeigten, während in Testphase 2 generell die Rutschungen besser erkannt wurden. Ferner ist es angedacht, die Auswahl der Parameter mit einem anderen Klassifikator (Klassifikationsbäume) vorzunehmen, um zu sehen, ob beide Verfahren die Parameter ähnlich gewichten.

Weiterhin wäre es interessant, die Rutschungserkennung in einem anderen Gebiet durchzuführen um die Übertragbarkeit des Verfahrens zu testen.

6 Dank

Die hier dargestellte Rutschungserkennung basiert auf den Erhebungen und Ergebnissen des Projektes „Vegetationswirkungen und Rutschungen“, das von Herrn Dipl.-Ing. Christian Rickli (Eidgenössische Forschungsanstalt WSL, Birmensdorf/Schweiz) geleitet wurde. Für die Nutzung der Daten möchten wir seinem Team danken. Für die Zusammenarbeit mit dem Amt für Wald und Landschaft, Abteilung Naturgefahren danken wir Herrn Dipl.-Ing. Josef Hess.

Dank geht auch an die DFG für die finanzielle Unterstützung innerhalb des Projektes „Hangbewegungsanalyse“.

Literatur

BURGES, C.J.C. (1998): A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2/2, 121-167.

BWW (Bundesamt für Wasserwirtschaft, 1998): Ereignisdokumentation Sachseln – Unwetter vom 15. August 1997. Studienbericht Nr.8 /1998, 48 S.; Biel.

CHUNG, C.F. (1995): Statistical regression models for predicting landslide hazard. *Atlantic Geology*, Vol. 31, 2, 121 S..

CRUDEN, D.M., VARNES, D.J. (1996): Landslide Types and Processes. In Turner, A.K., Schuster R.L.: *Landslides Investigation and Mitigation*, Special Report, 247, 36-75; Nat. Acad. Press (Washington).

EM-DAT (2003): Country Profiles for Natural Disasters. CRED, URL <<http://www.em-dat.net>>, Data [2004].

FERNÁNDEZ-STEGER, T. (2002): Erkennung von Hangrutschsystemen mit Neuronalen Netzen als Grundlage für Georisikoanalysen. Diss. Univ. Karlsruhe, 136 S.; Karlsruhe.

HAMBERGER, M. (2000): Geologisch-geotechnische Untersuchungen und Analyse von Hangbewegungen anlässlich von Starkniederschlägen im Melchtal/Schweiz. Unveröff. Dipl.-Arbeit Univ. Erlangen, 122 S; Erlangen.

KIENHOLZ, H. (1996): Gefahrenkarten: Massgebliche Parameter und Kriterien zur Festlegung von Intensitätsklassen. Int. Symp. Interpretation 1996, Garmisch-Patenkirchen, Vol. 3, 47-58; Garmisch-Patenkirchen.

MAYORAZ, F., CORNU, T., VULLIT, L. (1996): Using neural networks to predict slope movements. Proc. of 17th Int. Symp. on Landslides 1996, Trondheim, Vol. 1, 295-300; Balkema (Rotterdam).

NEAUPANE, K.M., ACHET, S.H. (2004): Use of backpropagation neural network for landslide monitoring: a case study in the higher Himalaya. *Engineering Geology*, 74, 213-226; Elsevier (Amsterdam).

RICKLI, C. (2001): Vegetationswirkungen und Rutschungen – Untersuchung zum Einfluss der Vegetation auf oberflächennahe Rutschprozesse anhand der Unwetterereignisse Sachseln OW am 15. August 1997. Eidg. Forschungsanstalt WSL, Bundesamt für Umwelt, Wald und Landschaft, 97 S.; Birmensdorf, Bern.

SMOLA, A.J., BARTLETT, P.L., SCHÖLKOPF, B., SCHUURMANS, D. (2000). Introduction to large margin classifiers. In: Smola, A.J., Bartlett, P.L., Schölkopf, B., Schuurmans, D.: *Advances in Large Margin Classifiers*, Massachusetts Inst. of Tech. (MIT), 1-28; MIT Press (Cambridge, Massachusetts).