

Das Informationssystem Einzugsgebietsmanagement: Ein generisches Informationssystem zur Verwaltung komplexer Geodaten mit Funktionen zur wissensbasierten Prognose

Dipl.-Ing. Manja Sieste, Dr. Andreas Barth, Dipl.-Ing. Sven Etzold

Beak Consultants GmbH

Am St. Niclas Schacht 13

09599 Freiberg

Im Rahmen des internationalen Verbundprojekts EMTAL (Einzugsgebietsmanagement für Talsperren in Mittelgebirgslandschaften) , gefördert vom Bundesministerium für Bildung und Forschung und bearbeitet von der TU Bergakademie Freiberg und der Beak Consultants GmbH, wurde ein komplexes Informationssystem zur Verwaltung der Projektergebnisse entwickelt. Dieses System sollte geeignet sein, eine große Anzahl komplexer Fachdatenmodelle für geographische Daten schnell, einfach und benutzerfreundlich aufzunehmen. Als Lösung wurde ein Mechanismus zur automatischen Generierung einzelner Softwaremodule entwickelt, der beschreibende Daten (Metadaten) über die Fachinformationen in der Datenbank und deren Modulstruktur auswertet. Aufbauend auf diesem flexiblen Datenmanagementinstrument wurde in das System eine Wissenskomponente integriert, die in der Lage ist, mit Hilfe von geostatistischen Verfahren, statistisch determinierte Regeln für das Auftreten von bestimmten Prozessen (z.B. Georisiken) zu ermitteln.

Within the scope of the interdisciplinary project EMTAL (Management of Catchment Areas with Reservoirs in Mountainous Regions), carried out for the BMBF (Federal Ministry of Education and Research) in cooperation with the University of Technology and Mining Freiberg, a complex information management system was developed. This system shall be applicable to store a great number of complex geographical data models in a fast and user-friendly way. As solution, a mechanism for automatical generation of single software modules was implemented, which interprets meta data about these datamodels and their software module structure. On the basis of this flexible tool for data management a knowledge based module was integrated into the system, which uses multivariate statistical methods to determine rules for the occurrence of certain processes (e.g. georisks).

1 Motivation

1.1 Ziele des Forschungsprojekts EMTAL

Ziel des interdisziplinären Projektverbundes ist die exemplarische Entwicklung eines Risiko-Managements für europäische Mittelgebirgseinzugsgebiete. Integrative Strategien des Umweltmanagements, sozioökonomische Szenarien und Monitoring-Richtlinien für Talsperren-Einzugsgebiete in Mittelgebirgsregionen unter Beachtung der gültigen EU-Wasserrahmenrichtlinie (EU-WRRL) sind wesentliche Bestandteile dieses Vorgehens. Die Arbeitsergebnisse und Vorgaben der Ländergruppe Wasser (LAWA) sollen durch eine enge Kooperation beachtet und integriert werden. Dieser Ansatz hat nicht allein den unmittelbaren Hochwasserschutz im Blick, sondern sieht das Zielgebiet als System, das nur in einer integrierten Herangehensweise robuste Ergebnisse und Handlungsempfehlungen erarbeiten kann.

Basierend auf einer umfassenden Daten- und Informationsbasis sollen die Ergebnisse zum Talsperren-Management im Zielgebiet in dem geplanten Bearbeitungszeitraum von drei Jahren eine Übertragbarkeit sowohl auf das untersuchte Flussgebiet als auch für Gebiete vergleichbarer Fragestellungen auf bundesdeutscher bzw. europäischer Ebene ermöglichen.

1.2 Anforderungen an das Informationssystem

Während der Bearbeitung des Forschungsprojekts EMTAL wurden Fachdaten mit einem Volumen von ca. 50 Gigabyte akquiriert bzw. von den Arbeitsgruppen neu erstellt. Um diese wichtigen Ergebnisse bzw. deren Grundlagen dauerhaft recherchierbar zu machen, ist die strukturierte Verwaltung dieser Daten in einer zentralen Projektdatenbank unabdingbar. Nur mit Hilfe eines solchen Instruments kann die weitere Verwertbarkeit der Projektergebnisse für zukünftige Projekte bzw. Maßnahmen gesichert werden.

Das wichtigste Ziel bei der Entwicklung des Informationssystems war die Realisierung einer übertragbaren Softwarelösung. Unter Übertragbarkeit sind in diesem Zusammenhang insbesondere folgende Eigenschaften einer Software zusammenzufassen:

- Möglichkeit der Verwaltung, Recherche und Visualisierung einer nahezu unbegrenzten Zahl von Fachdatenbeständen,
- Verwaltung von Fachdatenbeständen unterschiedlichen Komplexitätsgrades,
- keine hart vorgegebene Definition der Fachdatenmodelle der Datenbestände im System (beispielsweise wird eine Bodenkarte in Sachsen durch andere Attribute und Schlüssel Listen beschrieben als in Schleswig-Holstein), d.h.
- aufwändige Anpassungsprogrammierung für die Integration neuer Fachdatenbestände in das System ist unbedingt zu vermeiden.

Des Weiteren sollte ein Softwaremodul integriert werden, das in der Lage ist, auf der Grundlage der vorhandenen beschreibenden Daten über das Gebiet (räumliche Informationen, Gebietscharakteristika) und vorhandener Beobachtungen Regeln für das Auftreten von bestimmten Prozessen zu ermitteln:

- Das verwendete Verfahren muss so flexibel sein, dass es sich zur Analyse unterschiedlichster Prozesse und Phänomene eignet. Dabei müssen auch Prozesse, die weniger erforscht sind, untersucht werden können.
- Das Verfahren soll vorhandene räumliche Informationen über das Projektgebiet für die Analyse nutzen, weiterhin sind beobachtete Phänomene für die Analyse der Prozesse zu verwenden.
- Das in einer Analyse ermittelte Wissen, unter welchen Bedingungen ein bestimmter Prozess / ein bestimmtes Phänomen auftritt, ist in der Datenbank recherchierbar abzulegen und in ein Szenarienmanagement zu integrieren.
- Dieses Wissen muss für eine Übertragung in ähnliche Bearbeitungsgebiete geeignet sein.
- Gefährdungskarten mit Eintrittswahrscheinlichkeiten für einen bestimmten Prozess sind zu generieren.

1.3 Softwarekomponenten

Das entwickelte „Informationssystem Einzugsgebietsmanagement“ ist eine klassische Client-Server-Lösung und besteht im Wesentlichen aus drei Komponenten (siehe Abbildung 1):

- Die **zentrale Datenbank** verwaltet die strukturierten Sachdaten und übernimmt ebenfalls das Management der Geodaten und Dokumente. Weiterhin werden hier semantische Metadaten zur automatischen Generierung einzelner Softwareteile gespeichert.
- Die Softwarekomponente „**Geodatenmanagement**“ (Bestandteil der Client-Applikation) besteht aus verschiedenen Fachmodulen, die mit der integrierten GIS-Komponente gekoppelt sind.
- Das Softwaremodul „**Statistische Prognose**“ (Bestandteil der Client-Applikation) ist ein separates Modul und verfügt über Funktionen zum Management verschiedener Szenarien sowie über GIS-Funktionen zum Ausführen des Auswertalgorithmus.

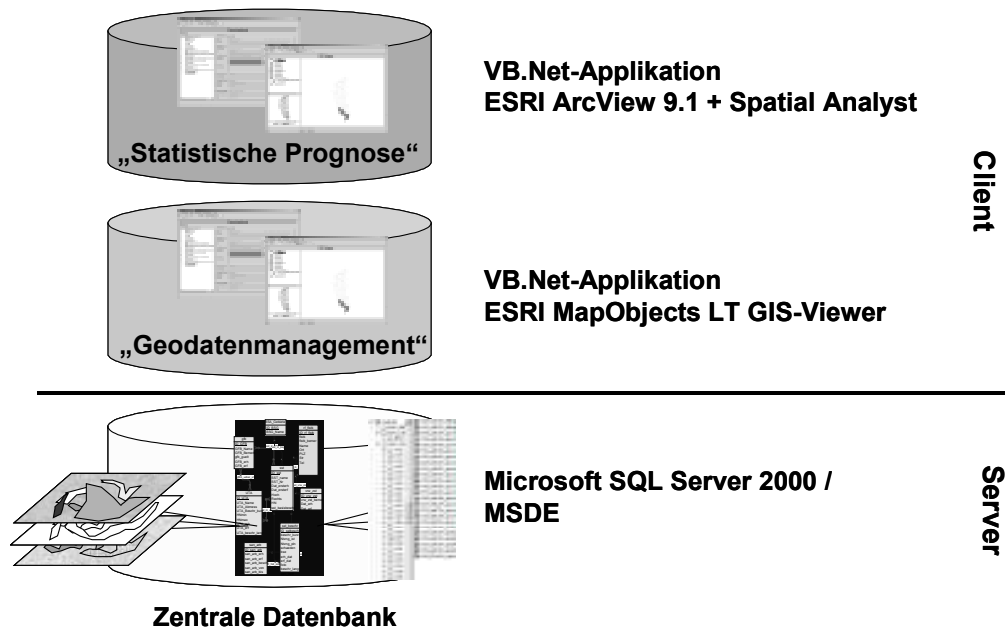


Abbildung 1: Softwarekomponenten des Informationssystems

Besonderer Wert wurde auf die Skalierbarkeit des Systems gelegt: Das Informationssystem Einzugsgebietsmanagement ist mit ausschließlich lizenzkostenfreier oder mit lizenzkostenpflichtiger Systemsoftware verfügbar:

- Datenbankmanagementsystem: Microsoft SQL Server 2000 (lizenzkostenpflichtig) / Microsoft SQL Server Desktop Engine MSDE 2000 (lizenzkostenfrei)
- Sachdatenanzwendung: VisualBasic.Net 2003-Applikation (lizenzkostenfrei)
- GIS: ESRI MapObjects LT (lizenzkostenfrei) / ESRI ArcView 9.1 und ESRI Spatial Analyst 9.1 (lizenzkostenpflichtig)

Damit kann das Informationssystem Einzugsgebietsmanagement in zwei unterschiedlichen Konfigurationen, mit unterschiedlichem Funktionsumfang, jedoch mit weitestgehend einheitlichem Layout der Programmmodule installiert werden:

Die lizenzkostenfreie „Thin Version“ auf der Basis des Datenbankmanagementsystems MSDE 2000 ist eine gute Basisvariante mit komplexen Funktionen zur Datenvisualisierung für gut dokumentierte mittelgroße Einzugsgebiete, deren Datenvolumen in der strukturierten Datenbank 2 GB nicht übersteigen sollte, da ab dieser Größe laut Auskunft des Herstellers erhebliche datenbankseitige Leistungseinbußen zu erwarten sind. Die lizenzkostenpflichtige „Thick Version“ implementiert einen Datenbankserver mit Speichermöglichkeiten von nahezu unbegrenzt hohem Datenvolumen. Darüber hinaus können die wissensbasierten Funktionen des Systems mit dieser Version genutzt werden.

2 Die Softwarekomponente „Geodatenmanagement“

Die Softwarekomponente „Geodatenmanagement“ besteht im Wesentlichen aus drei Modulen, die auf eine bestimmte Art und Weise miteinander interagieren:

- Im **Modul „Datenbestände“** werden die Metadaten zu einzelnen Fachdatenbeständen mit geographischem Bezug verwaltet. Die Fachdatenbestände können dabei nach Wissensgebieten geordnet in einer Baumstruktur dargestellt werden.
- Die Sachdaten eines Fachdatenbestandes können im Informationssystem in einem separaten relationalen Fachdatenmodell verwaltet werden. Dabei kann der Komplexitätsgrad des Modells je nach Erfordernis unterschiedlich hoch sein. Die Sachdaten werden im Modul **„Legenden / Sach-**

daten“ dem Nutzer präsentiert. Jedes Fachdatenmodell wird in der Datenbank durch semantische Metadaten beschrieben, aus denen zu Laufzeit eine neue Instanz des Moduls „Legenden / Sachdaten“ generiert wird.

- Im **GIS-Modul** können die geographischen Datenbestände visualisiert werden.

2.1 Generischer Ansatz für die Verwaltung komplexer Fachdatenbanken

Werden in der Datenbank neue Fachdatenmodelle für geographische Datenbestände definiert, muss die Clientapplikation diese mit zugehörigen Benutzeroberflächen präsentieren können. Da aus Übertragbarkeitsgründen für diese Benutzeroberflächen keine Neuprogrammierung in Frage kam, wurde ein modellgesteuertes Verfahren entworfen, das auf der Basis von Metainformationen die entsprechenden Softwareteile zur Programmlaufzeit generiert. Bei dem entwickelten Verfahren handelt es sich jedoch nicht um einen klassischen Ansatz der modellgetriebenen Softwareentwicklung (Model Driven Software Development – MDSD) im Sinne von STAHL & VÖLTER 2005, die eine Auswertung dieser Metainformation zur Compilezeit der Software voraussetzt, sondern um eine modellgesteuerte Metadatenverarbeitung mit interpretativen Methoden zur Systemlaufzeit. Der Vorteil dieses Verfahrens liegt insbesondere in der leichteren Wartbarkeit des Systems.

Die Generierung neuer Fachmodule im Informationssystem Einzugsgebietsmanagement über Metadaten basiert auf den folgenden Überlegungen: Die einzelnen Fachmodule des Informationssystems sollten über ein einheitliches Layout verfügen. Damit ist ein hoher Wiedererkennungswert einzelner Funktionen und somit eine hohe Benutzerfreundlichkeit gewährleistet. Über semantische Metadaten in der Datenbank wird der Aufbau der einzelnen generischen Fachmodule beschrieben, so u.a.:

- Welche physischen Datenbanktabellen sind mit der Übersichtsanzeige im unteren Modulbereich verknüpft? Welche Spalten werden in welchem Modulbereich dargestellt?
- Welche Tabellenspalten sind in welcher fachlichen Reihenfolge und mit welcher Beschriftung in der Detailanzeige des Moduls zu sehen?
- Welche Informationen in der Detailanzeige sind mit fest vorgegebenen Schlüssellisten verknüpft, welche Spalten sind Pflichtfelder?
- Welche Verknüpfungen zu nachgeordneten Tabellen bestehen, und wie werden sie im Modul dargestellt?

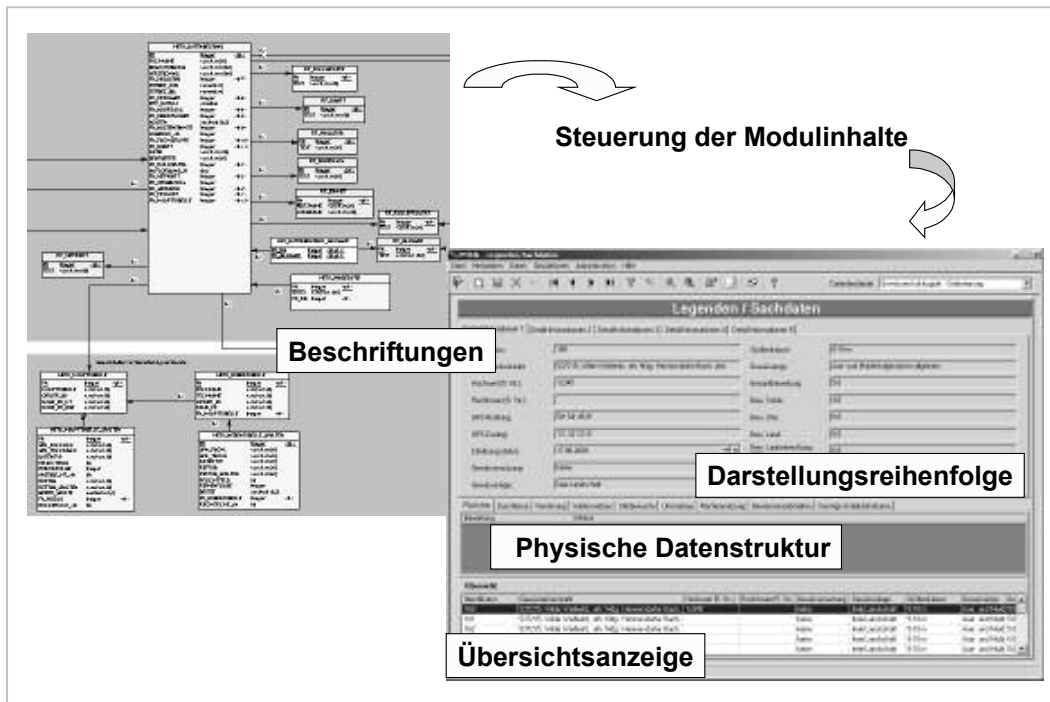


Abbildung 2: Prinzipskizze des generischen Ansatzes für Fachmodule

Das Informationssystem Einzugsgebietsmanagement verfügt über insgesamt drei dieser generischen Module, die in der Lage sind, aus in der Datenbank vorhandenen Metadaten zur Laufzeit neue Fachmodule zu generieren:

- Modul „**Sachdaten / Legenden**“: Verwaltung von Informationen zu wesentlichen Entitäten der Datenbank und komplexen Kartenlegenden
- Modul „**Kontinuierliche Daten**“: Management von Messstellen und den zugehörigen Loggerdaten
- Modul „**Proben und Analysen**“: Verwaltung der Beprobungspunkte, entnommener Proben und Analysen

2.2 Funktionalitäten

Die generierten Module des Informationssystems interagieren mit dem Modul „Datenbestände“ und dem jeweiligen GIS-Modul in der folgenden Art und Weise (siehe auch Abbildung 3):

Das Modul „Datenbestände“ dient als Einstieg in das System. Hier können die Metadaten zu den Projektergebnissen verwaltet und recherchiert werden. Neben der allgemeinen Datenbeschreibung besteht die Möglichkeit zu definieren, wie der Datenbestand im Informationssystem verwaltet wird:

- ausschließlich in Form von Metadaten (Name, Aktualität, Maßstab usw.) oder
- über eine strukturierte Verwaltung der Sachdaten in der Datenbank und automatischer Generierung eines Fachmoduls zur Präsentation der Daten.

Für einen durch den Benutzer ausgewählten Datensatz im Modul „Datenbestände“ kann im zugehörigen Sachdatenmodul nach konkreten Objekten mit bestimmten Attributen recherchiert werden. Liegt ein Rechercheergebnis vor (z.B. alle Gewässerabschnitte mit der Strukturgüteklasse 1), können die Ergebnisse mit der GIS-Komponente dargestellt und weiterverarbeitet werden. Weiterhin besteht die Möglichkeit, dass der Nutzer für einen ausgewählten Datensatz direkt aus dem GIS- bzw. Sachdatenmodul heraus die entsprechenden Metadaten anzeigt.

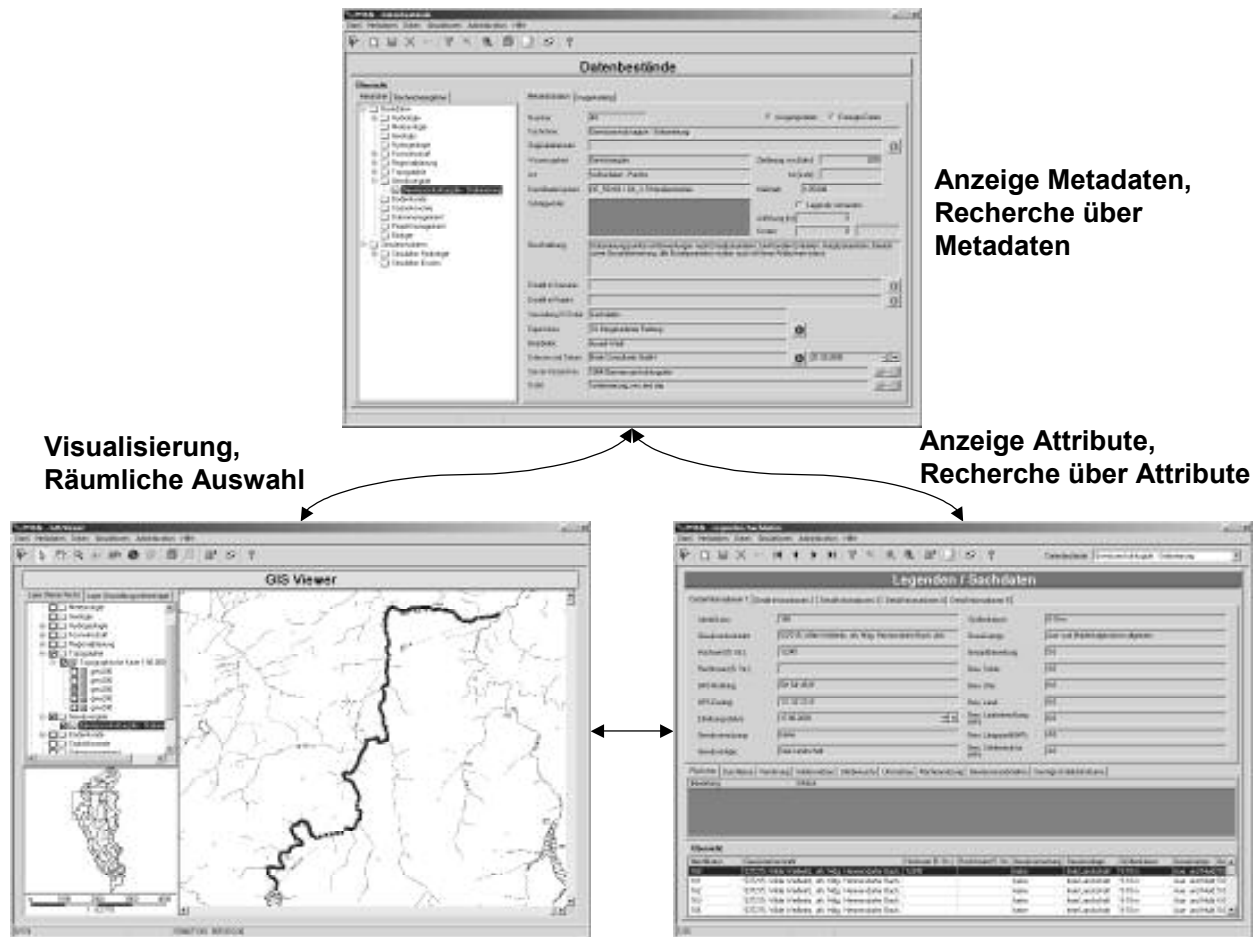


Abbildung 3: Interaktion zwischen den Modulen der Softwarekomponente "Geodatenmanagement"

3 Softwaremodul „Statistische Prognose“

Für das Management eines Einzugsgebietes ist die Kenntnis bzw. die Prognose von bestimmten Prozessen und Situationen (z.B. Prognose von Georisiken, wie Hangrutschungen, Schlammströme, Steinerschlag, Bodenerosion, Hohlraum- und Höhlenbildung, Bergschäden und Erdbrüche) von großer Bedeutung. Das in das Informationssystem integrierte Modul „Statistische Prognose“ kann auf der Basis bekannter Regeln statistisch begründete Vorhersagen für das Auftreten von Ereignissen treffen:

- Erstellung von **Regeln über das Auftreten von Georisiken** anhand dokumentierter Ereignisse
- **Anwendung dieser Regeln** zur Prognose von Ereignissen in anderen Gebieten
- **Erstellung von Gefährdungskarten** mit Eintrittswahrscheinlichkeiten für einen bestimmten Prozess

3.1 Methodik

Die umgesetzte probabilistische Methodik der Evidenzgewichtung wurde ursprünglich für den Bereich der explorativen Mineralogie von BONHAM-CARTER et al. 1989 entwickelt. Sie basiert auf den folgenden Überlegungen:

Die Charakteristik von Prozessen wird durch bestimmte gebietsbezogene Eigenschaften bzw. Steuerungsfaktoren bestimmt. Mit Hilfe der Methode der gewichteten Evidenzen kann eine indirekte Bewertung der Anfälligkeit eines Gebietes für einen bestimmten Prozess auf der Basis solcher Steuerungsfaktoren (Evidenzen) erfolgen. Im methodischen Ansatz wird davon ausgegangen, dass die Entstehung von Georisiken von gleichen bzw. ähnlichen Bedingungen bewirkt wird, wie in der Vergangenheit auftretende Ereignisse des gleichen Typs.

Das Prinzip der Evidenzgewichtung geht von der A-Priori-Wahrscheinlichkeit aus, die mit Hilfe der Erfahrung über das Eintreten eines bestimmten Ereignisses unter gleichen Bedingungen berechnet wird. Diese wird mit flächenhaften Informationen über mögliche Steuerungsfaktoren modifiziert: Daraus resultiert die bedingte A-Posteriori-Wahrscheinlichkeit für das Eintreten des Ereignisses.

Die praktische Umsetzung dieser Methodik erfolgt durch paarweise statistische Analyse der dokumentierten Ereignisse eines bestimmten Typs (z.B. Bodenerosion) mit jeweils einem Kartenthema der Steuerungsfaktoren (z.B. Hangneigung, Bewuchs, Vernässungsflächen, Bodenart,...). Daraus wird der statistische Zusammenhang zwischen beiden Ebenen ermittelt und damit beurteilt, wie signifikant eine Evidenz für das Auftreten eines Prozesses ist. Dies erfolgt durch ein Paar von Gewichten und verschiedene andere statistische Kennzahlen. Nach der Berechnung dieser Kennzahlen für jede Evidenzklasse, können diese anschließend verwendet werden, um eine flächendeckende Vorhersage (Eintrittswahrscheinlichkeit) eines Prozesses unter Integration aller Evidenzen zu treffen.

Mit dieser Methode wird ein quantitatives Modell zu Ursachen eines Prozesses erzeugt, das einerseits Aussagen erlaubt, welche der vorhandenen Faktoren die größte relative Bedeutung für ein Auftreten haben, andererseits die Übertragung der vorhandenen Evidenzgewichte auf andere bzw. ähnliche Gebiete und damit eine Erstellung von flächendeckenden Gefährdungskarten ermöglicht: **Der Algorithmus besitzt die Fähigkeit, in bestimmten repräsentativen Testgebieten anhand vorhandener Ereignisse zu „lernen“ und das so ermittelte „Wissen“ über Steuerungsfaktoren auf andere (ähnliche) Gebiete anzuwenden.**

Ein wesentlicher Vorteil der Methode ist, dass Aspekte der Unsicherheit im Modell behandelt werden können. Dies betrifft insbesondere Unsicherheiten, die durch starke Varianz der Daten und durch fehlende Daten entstehen. Die Wahrscheinlichkeiten können somit auf ihre Plausibilität überprüft und gegebenenfalls die Unsicherheiten bei der Interpretation des Endergebnisses berücksichtigt werden.

3.2 Einbindung des Algorithmus in das Informationssystem

Der Algorithmus zur statistischen Prognose wurde als Funktion einer ESRI ArcGIS-9.1 Extension auf der Basis des ESRI Spatial Analyst für ArcGIS 9.1 umgesetzt (GIS-Modul der „Thick Version“) und direkt an die Programmkomponente „Geodatenmanagement“ angebunden. Damit ist die dauerhafte Nachvollziehbarkeit der Berechnung von den Ausgangsdaten, die in ein Szenario einfließen, bis hin zu den Zwischenergebnissen und der Ergebniskarte gewährleistet. Es wurde ein Fachmodul „Statistische Analysen“ mit folgenden Funktionen in das System integriert:

- **Szenarienmanagement:** Hier besteht unter anderem die Möglichkeit, die Fachthemen der Steuerungsfaktoren und eine Gebietsgrenze anzugeben. Weiterhin kann der Nutzer entweder ein Fachthema über bekannte Ereignisse (im Falle eines Trainingsdurchlaufs) oder ein Basisszenario (im Falle der Anwendung erarbeiteten Wissens) zur Erstellung einer flächenhaften Gefahrenkarte auswählen.
- **Ausführung des Algorithmus** direkt aus dem Fachmodul heraus. Der Algorithmus generiert ein Grid mit einer flächenhaft verteilten Eintrittswahrscheinlichkeit für das jeweils betrachtete Georisiko im bearbeiteten Gebiet. Dieses wird nach Ablauf automatisch im GIS dargestellt.
- **Übernahme der generierten Statistiken und der Ergebniskarte in die zentrale Datenbank**

3.3 Anwendungsbeispiel

In einem Testgebiet wurden auf der Grundlageder Befliegungsdaten nach dem Hochwasser 2002 Stellen mit Bodenerosion identifiziert. Weiterhin wurden mögliche Themen der Steuerungsfaktoren aus den vorhandenen Daten ausgewählt bzw. neu erstellt. Im einfachsten Fall waren das die Kartenthemen Landnutzung, Bodenarten und Hangneigung. Im Testgebiet konnte gezeigt werden, dass ca. 90 % der beobachteten Flächen, an denen Erosion nach den Starkregenereignissen im August 2002 auftrat, mit der Gefährdungsklasse „sehr hoch“ bzw. „hoch“ eingestuft wurden.

Die im Testgebiet erstellten Regeln wurden auf das Einzugsgebiet des Höckenbaches übertragen. Hier lag die Trefferquote bei 80 %. Für dieses, mit relativ wenigen Eingangsdaten konstruierte Szenario (die Einbeziehung weiterer Kartenthemen als Steuerungsfaktoren wäre denkbar, so u.a. Wölbungsra-

dien oder orographisch bedingte Luv-/ Lee-Effekte) kann sie bereits als hoch eingeschätzt werden. Wichtig ist vor allem die Anwendung des Algorithmus auf ein gut dokumentiertes Testgebiet, damit das System zunächst „Wissen aufbauen“ kann. Die Anwendung dieses ermittelten Wissens kann prinzipiell auf ähnliche Gebiete erfolgen, für die die gleichen Kartenthemen als Steuerungsfaktoren einbezogen werden können wie im Testgebiet.

4 Zusammenfassung und Ausblick

Der generische Ansatz des entwickelten Informationssystems macht diese zu einem universell einsetzbaren Instrument zur Verwaltung und Visualisierung von komplexen Geodaten. Darüber hinaus eröffnen Konzept und Aufbau der Datenkomponente und insbesondere der Wissenskomponente breite Möglichkeiten für die Anwendung der Untersuchungsergebnisse auf vergleichbare Regionen. Das in der Wissenskomponente zum Einsatz kommende Verfahren ist in vielen Bereichen der Umweltwissenschaften und der Analyse von Georisiken einsetzbar.

5 Literaturverzeichnis

K. BACKHAUS, B. ERICHSON, W. PLINKE, R. WEIBER (2003): *Multivariate Analysemethoden – eine anwendungsorientierte Einführung*; Springer-Verlag, Berlin; ISBN 3-540 00491

G.F. BONHAM-CARTER, F.P. AGTERBERG, D.F. WRIGHT (1989): *Weights of evidence modelling: a new approach to mapping mineral potential*; Statistical Applications in the Earth Sciences; Geological Survey of Canada, S. 171-183

T. STAHL, M. VÖLTER (2005): *Modellgetriebene Softwareentwicklung – Techniken, Engineering, Management*; dpunkt.verlag, Heidelberg; ISBN 3-89864 310 7;

6 Ansprechpartner

Dipl.-Ing. Manja Sieste

Tel.: ++49 (0)3731 781371

Email: Manja.Sieste@beak.de