

# Die Verwaltung großvolumiger Geo-Rasterdaten in Standard-Datenbanken

Peter Baumann

Rasdaman GmbH, FORWISS (Bavarian Research Center for Knowledge-based Systems), E-Mail: baumann@rasdaman.com

Rasterdaten gewinnen als Ergänzung zu den klassischen Vektordaten im GIS rapide an Bedeutung. Die Gründe sind sowohl anwendungsseitig als auch technisch: Rasterdaten erhöhen den Nutzen kartographischer Produkte auf Grund ihrer Anschaulichkeit, der zusätzlichen Informationen sowie des oftmals höheren Aktualitätsgrads im Vergleich zu Vektordaten. Zudem wird der Zugang zu aktuellen Nah- und Fernerkundungsdaten immer kostengünstiger, und die verfügbaren Speichertechnologien sind mittlerweile ausreichend, um auch großvolumige Bildarchive im Direktzugriff zu halten.

Ein relativ junger Trend ist, diese Rasterdaten nicht wie bisher als Dateien vorzuhalten, sondern im Zuge einer integrierten Informationsverwaltung mit Meta- und Vektordaten zusammen in einer Datenbank vorzuhalten und via Internet anzubieten. Aus Datenbanksicht ergeben sich daraus neue Anforderungen, welche dazu zwingen, zentrale Systemkomponenten neu zu überdenken.

In diesem Beitrag diskutieren wir die Möglichkeiten der Datenbankunterstützung für Mehrwertdienste auf multidimensionalen Rasterdaten. Wie wir zeigen, lässt sich die Standard-Datenbanksprache SQL in natürlicher Weise um Operationen auf Rasterdaten erweitern, so dass sich ein breites Feld von Analyseaufgaben damit erledigen lässt. Wir veranschaulichen dies an Hand von implementierten Fallbeispielen auf Basis des Systems rasdaman, dem historisch ersten System mit Anfrageunterstützung für multidimensionale Rasterdaten.

---

## 1 Einleitung

Traditionell werden Luft/Satellitenbilder und andere Rasterdaten im Dateisystem gehalten, während die Ablage von Meta- und Vektordaten in Datenbanken heutzutage mehr oder weniger Standard ist. Nun ist flexible Suche und Auswertung auf großen Datenvolumina und unabhängig von Ausgangs- und Zielformaten die traditionelle Domäne von Datenbanksystemen (DBSen).

Das Konzept einer Anfragesprache, welche dem Anwender (besser: dem Anwenderprogramm, denn die Vorstellung von SQL als einer „natürlichen“ Sprache, geeignet für Endanwender, ist längst vom Tisch; SQL ist heute die Kommunikationssprache zwischen Anwendungsprogramm und Datenbanksystem) vielfältige Freiheit im Zugriff erlauben und gleichzeitig server-intern umfangreiche automatisierte Optimierungsmöglichkeiten zulassen, hat sich in vielen Anwendungsfeldern bewährt – von administrativbetriebswirtschaftlichen Applikationen über CAD bis zur Geoinformatik. Für Bilder und, allgemeiner, multidimensionale Rasterdaten sind adäquate Anfragetechniken allerdings erst in jüngerer Zeit entstanden.

Beispiele für solche multidimensionale Rasterdaten oder Multidimensionale Diskrete Daten (MDD) sind skalare Messreihen (1D) wie Temperatur oder Radioaktivität, 2D-Satellitenbilder kombiniert zu großen blattschnittfreien Karten der Erdoberfläche, 3D-Zeitreihen von Satellitenbildern (x/y/t) sowie geophysikalische Daten (x/y/z) und 4D-Klimamodelle (x/y/z/t).

Aus Datenbanksicht wird die Situation verkompliziert dadurch, dass offenbar eine große Vielfalt von MDD-Typen existiert. Herausforderungen ergeben sich aus der strukturellen Vielfalt, komplexen Auswertoperationen, und dem schieren Datenvolumen. In der Tat stellt sich heraus, dass es nicht ausreicht, einfach SQL um MDD-Operationen zu erweitern. Effiziente MDD-Unterstützung erfordert ein Überdenken der Architektur von Datenbanksystemen und die Komponenten auf die neuen Anforderungen auszurichten.

In diesem Beitrag stellen wir unsere Arbeiten auf dem Gebiet anwendungsunabhängiger Datenbankunterstützung für MDD vor. Basierend auf der Middleware rasdaman (“raster data manager”) zeigen wir, wie MDD-Datenbanken so

analysiert werden können, dass dem Anwender semantisch angemessene Information präsentiert wird, welche direkt und ohne komplexe Applikationslogik aus den Basisdaten abgeleitet wird. Den Schwerpunkt legen wir auf die einheitliche Behandlung unterschiedlich-dimensionaler Rasterobjekte.

In Abschnitt 2 geben wir eine kurze Einführung in die *rasdaman*-Anfragesprache, gefolgt von einer Übersicht über die Systemarchitektur. In Abschnitt 3 diskutieren wir geo-relevante Anwendungen unterschiedlicher Dimensionalität. Eine Zusammenfassung mit Ausblick auf weitere Forschungsthemen findet sich in Abschnitt 4.

## 2 Dimensionsübergreifende Manipulation von Rasterdaten

Das konzeptuelle Modell von *rasdaman* basiert auf Arrays (im programmiertechnischen Sinn), welche von beliebiger Dimension und raumzeitlicher Ausdehnung sein können. Als Zellentypen sind alle in C/C++ möglichen Datentypen erlaubt, mit Ausnahme von Zeigern. Dem relationalen Paradigma folgend unterstützt *rasdaman* Mengen von Arrays. Eine *rasdaman*-Datenbank kann damit als eine Menge von zweispaltigen Tabellen aufgefasst werden, deren jede eine array-wertige Spalte (Attribut) enthält sowie einen systemgenerierten Objektidentifikator (OID).

Die Anfragesprache von *rasdaman*, *rasql*, basiert auf einer eigens dafür entwickelten Array-Algebra (BAUMANN 1999) und erweitert das ISO-standardisierte SQL92 um deklarative MDD-Operatoren. Wie bei SQL (vgl. [www.sql.org](http://www.sql.org)) liefern *rasql*-Anfragen mengenwertige Resultate zurück. Die server-basierte Auswertung nutzt algebraische Optimierung (REINER et al. 2002) und eine spezialisierte Speicherverwaltung (FURTADO & BAUMANN 1999).

### 2.1 MDD-Retrieval

Wir stellen im Folgenden die wichtigsten Basisoperationen vor; sie lassen sich zu beliebig komplexen Ausdrücken kombinieren.

#### 2.1.1 Ausschnittsbildung

Dies umfaßt *Trimmen* (rechteckige Ausschnitte) und *Schnitte* (Extraktion von niedrigerdimensionalen Schichten).

#### Beispiel 1

Die folgende Anfrage extrahiert einen 2000x3000 Ausschnitt von jedem Landsat-Bild in der Tabelle („Kollektion“) *LandsatCollection*:

```
SELECT c[ 1001:3000, 1001:4000 ]
FROM   LandsatCollection AS c
```

#### Beispiel 2

Nehmen wir an, *ClimateCollection* enthält einen 4D-Datenwürfel mit den Dimensionen *x*, *y*, *z* und *t*. Dann extrahiert die folgende Anfrage daraus einen 3D-Würfel bei Position *t=1020*:

```
SELECT c[ 1020, ** , ** , ** ]
FROM   ClimateCollection AS c
```

#### 2.1.2 Induzierte Operationen

Für jede Operation, welche auf dem Zellentyp eines MDD-Objekts definiert ist, existiert eine korrespondierende sogenannte induzierte Operation, welche die Basis-Operation simultan auf jede Zelle eines MDD anwendet. Sowohl einstellige Operationen lassen sich induzieren (z.B. Record-Zugriff oder Kontrasterhöhung) als auch binäre Operationen (z.B. Maskierung eines Bilds).

#### Beispiel 3

"Band 3 von allen Landsat TM Bildern, mit Intensität reduziert um 2":

```
SELECT c.band3 / 2
FROM   LandsatCollection AS c
```

#### Beispiel 4

Klassifikation, also die Abbildung von Zellwerten auf andere Werte, ist eine gebräuchliche Operation auf Fernerkundungsdaten, aber auch in anderen Bereichen wie Analyse von Geschäftsdaten. Kann diese Aufgabe vom DBS erledigt werden, so erleichtert dies nicht nur die Anwendungsentwicklung, sondern erhöht potenziell auch die Antwortgeschwindigkeit deutlich. Unser Beispiel (Abbildung 1) klassifiziert 8-bit DGMs in der Kollektion *DEMCollection* in verschiedenen Höhenlevel:

```
SELECT (d>=64 AND d<128)*1
      + (d>=128 AND d<192)*2
      + (d>=192 AND d<256)*3
FROM   DEMCollection AS d
```

### Beispiel 5

Abbildung 2 zeigt die drei sichtbaren Kanäle eines Satellitenbilds der spanischen Südküste. Die folgende Anfrage maskiert – basierend auf einem einfachen Farb-Schwellwert – alle Pixel aus, welche keine Vegetation enthalten.

```
SELECT c * (c.green > 130 AND c.red < 110 AND
c.blue < 140)
FROM LandsatCollection AS c
```

### Beispiel 6

In realen Applikationen lassen sich bessere Aussagen beispielsweise durch den Normalisierten Vegetationsindex (NDVI) gewinnen. Das Ergebnis wird in der folgenden Anfrage auf den Grauwert-Bereich (0..255) skaliert und auf 8-bit-Integer abgebildet:

```
SELECT (char) ( 255c * abs(c.infrared - c.red)/
(c.infrared + c.red))
FROM LandsatCollection AS c
```

Im Allgemeinen können MDD-Ausdrücke im SELECT-Teil einer Anfrage und, wenn der Ergebnistyp des Ausdrucks Boolean ist, auch im WHERE-Teil verwendet werden. Wir benötigen daher noch ein Mittel, um multidimensionale Ausdrücke zu skalaren Werten “kollabieren” zu lassen. Dazu dienen die sogenannten Condenser.

#### 2.1.3 Ableitung von aggregierten (summarischen) Daten

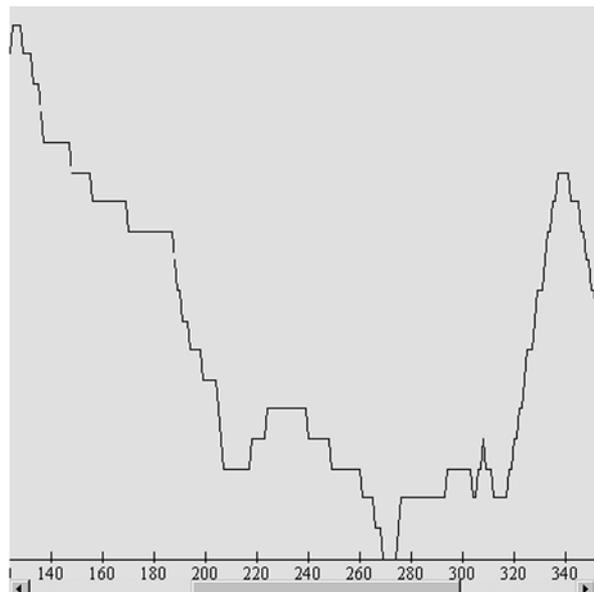
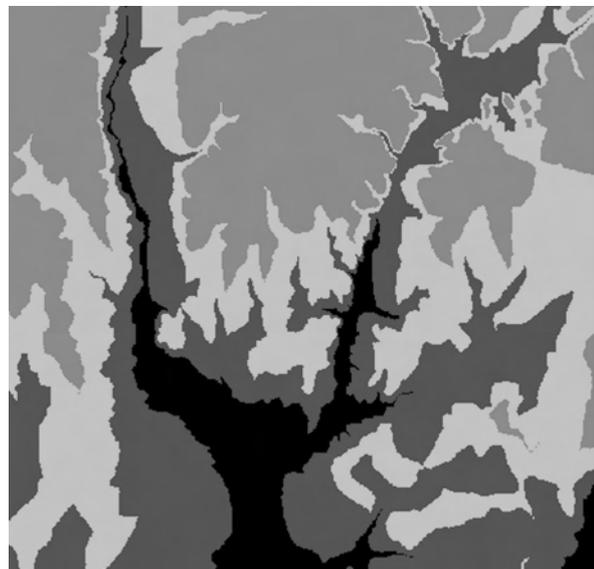
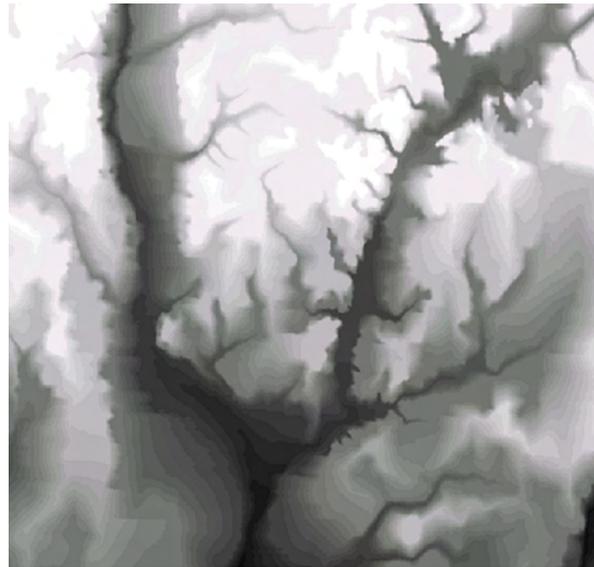
Die *Condenser* – in SQL *Aggregation* genannt – erlauben die Summation über alle Werte eines Gebiets. Beispielsweise liefert der Ausdruck `count_cells(a)` die Anzahl aller Zellen in `a`, welche den Wert `true` enthalten. Der Ausdruck `sum_cells(a)` liefert die Summe aller Zellwerte, `avg_cells(a)` den Durchschnitt.

### Beispiel 7

“Die Anzahl grüner Pixel in einer Landsat-Szene” (naive Form).

```
SELECT count_cells(c.green > 130 AND c.red < 110
AND c.blue < 140 )
FROM LandsatCollection AS c
```

Um solche Anfragen effizient zu beantworten, sind zwei Merkmale besonders wichtig. Zum einen sind die Anfragen so abstrakt formuliert, dass die Reihenfolge der Auswertung nicht festgelegt ist; damit kann intern eine Partitionierung der Daten erfolgen, und die Auswertung kann in der geeignetsten Reihenfolge stattfinden. Zum anderen bewirken interne Optimierungsmecha-



**Abb. 1:** Klassifikation eines 2D-DHM (oben) in 4 Höhenlevel (Mitte) und 1D-Geländeprofil (unten).

nismen, dass unabhängig von der Komplexität der Anfrage jede Zelle höchstens einmal gelesen wird.

#### 2.1.4 Datenformate

In den bisherigen Anfragebeispielen liefert rasmadan die MDD-Resultate im Hauptspeicherformat des Clients zurück, bereit zur unmittelbaren Weiterverarbeitung. Auf Anforderung können MDDs auch in gängigen Datenformaten geladen bzw. geliefert werden (z.B. in einem Web-tauglichen Format).

#### Beispiel 8

“Band 3 aus den Landsat-Bildern, codiert in TIFF”:

```
SELECT tiff( c.band3 )
FROM LandsatCollection AS c
```

## 2.2 MDD-Speicherung

Anfragen werden möglichst nahe am großen Datenarchiv ausgewertet, nämlich im Server. Auch die Speicher- und Zugriffsoptimierung findet hier statt. Das unterliegende Speichermodell basiert auf einer multidimensionalen Partitionierung der Daten, also einer Zerlegung in kleinere MDD-Objekte, sogenannte *Kacheln* („tiles“, Abbildung 3). Mehrere Kachelungs-Strategien sind realisiert worden, von einem äquidistanten Gitter (TAMURA 1980) bis zu systemgenerierter Kachelung basierend auf Zugriffsmustern und „Hot Spots“ (FURTADO & BAUMANN 1999; FURTADO et al. 1997). Die Kacheln bilden die Einheit des Datenzugriffs im Server; jede Kachel wird in einem BLOB (Binary Large Object, also ein Bytestring ohne weitere Semantik; vgl. LORIE 1982) des relationalen DBS abgelegt, optional verlustfrei oder verlustbehaftet komprimiert. Ein Geo-Index hilft, die von einer Anfrage betroffenen Kacheln schnell zu bestimmen.

## 3 Anwendungsszenarien

Wir präsentieren drei ausgewählte Anwendungen, welche praktisch relevante zwei-, drei- und vierdimensionale MDD-Strukturen illustrieren (siehe N.N.) und die dort angegebene Literatur für weitere Details. Alle Applikationen werden durch dieselbe Datenbank-Technologie bedient, so dass die unterschiedliche Handhabung lediglich in den Front-Ends zum Tragen kommt.

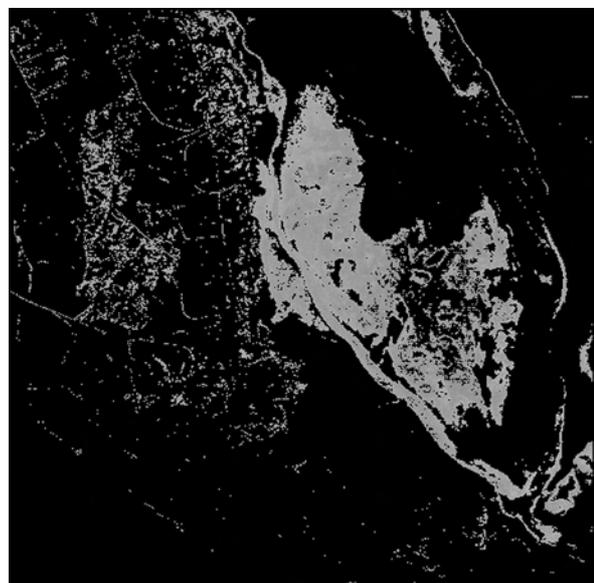
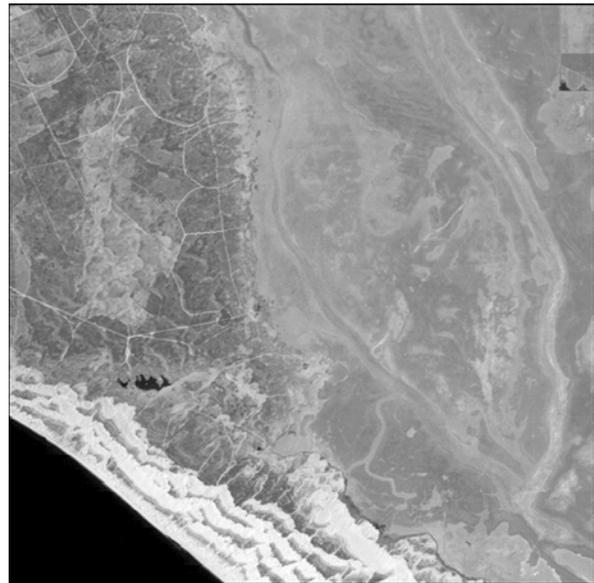


Abb. 2: Landsat-Szene (oben) und Vegetations-suche (unten).

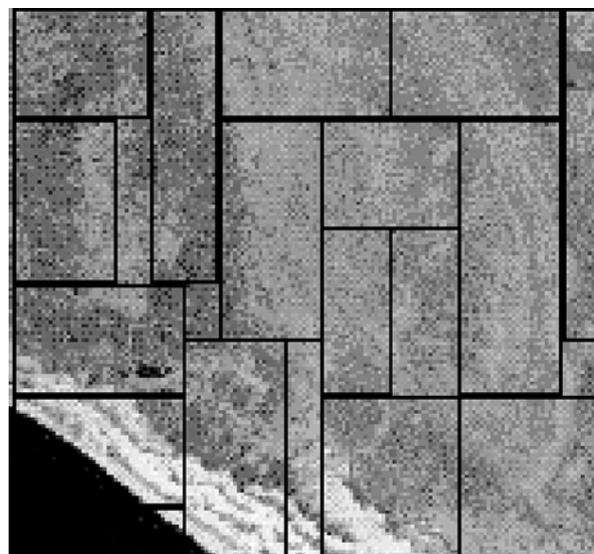


Abb. 3: Exemplarische 2D-Kachelung.

### 3.1 Szenario 1: 2-D Navigation

Größere Satellitenbild-Archive bewegen sich in den Größenordnungen multi-Terabyte bis mehrere Petabyte (1 Petabyte = 1 PB = 1.024 TB) von Mosaiken. Nachdem traditionell eigenentwickelte Nutzerschnittstellen zum Tragen kamen, existiert mit dem *Web Map Service* (WMS) des OpenGIS Consortiums (OGC) seit einiger Zeit ein Standard für den interoperablen Zugriff auf Geodaten ([www.opengis.org](http://www.opengis.org)).

WMS-Anfragen enthalten u.a. die Angabe der Zielkarten („Layer“), welche zu überlagern sind, Styles, nach denen Karten ggf. einzufärben sind, den geographischen Ausschnitt sowie die Zielfenstergröße in Pixeln, etwa wie folgt:

```
http://MeinServer/wmsServlet?
VERSI-
ON=1.1.0&SERVICE=WMS&REQUEST=GetMap&
SRS=EPSG:31464&BBOX=4636000.0,5717000.0,468
7000.0,5768000.0&
LAYERS=ortho_photo&STYLES=standard&
WIDTH=246&HEIGHT=300&
FOR-
MAT=image/jpeg&EXCEPTIONS=application/vnd.ogc.
se_inimage
```

Bei einem datenbank-gestützten Server wird aus diesen per URL übermittelten Parametern intern eine oder mehrere Datenbankabfragen generiert. Das Ergebnis einer Anfrage ist ein einziges Rasterbild, welches im Browser direkt angezeigt werden kann. Abbildung 4 zeigt das rasdaman-Frontend rasgeo (BAUMANN 2001) für Navigation, Konfektionierung, Bestellung und Auslieferung von komplexen Kartenprodukten.

### 3.2 Szenario 2: Satellitenbild-Zeitreihen (3-D)

Im Herbst 2003 wurde vom OpenGIS Consortium der Standard *Web Coverage Service* (WCS) 1.0.0 für den interoperablen Zugriff auf mehrdimensionale raum-zeitliche Rasterdaten verabschiedet und kann von der OGC-Website [www.opengis.org](http://www.opengis.org) heruntergeladen werden. Bereits zuvor war im EU-Projekt ESTEDI ([www.estedi.org](http://www.estedi.org)) auf Basis der Vorversion WCS 0.8 ein experimenteller Dienst auf 3D-Daten entwickelt worden. Dieser von der DLR auf Basis von rasdaman entwickelte Service-Prototyp für multitemporale blattschnittfreie Satellitenkarten realisiert exemplarisch Web-basierten Zugriff auf Satellitenbilder sowie eine nutzergesteuerte Datenaufbereitung. Der frei zugängliche Dienst wurde der OGC im Rahmen eines von ihr durch-

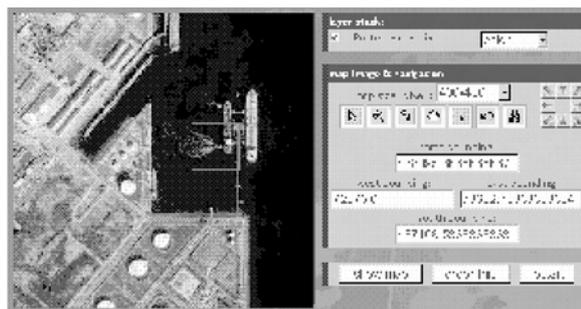


Abb. 4: Web-basierte Kartennavigation.

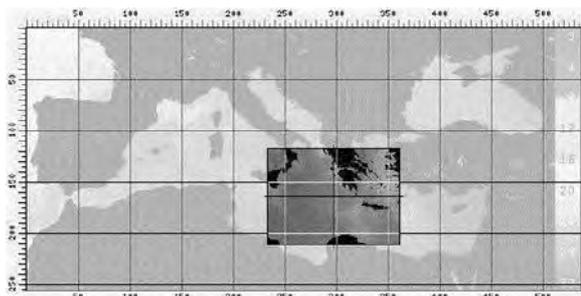


Abb. 5: Gebietsselektion auf dem 3D-Datenwürfel der DLR.

geführten Interoperability Testbeds Anfang 2002 vorgestellt und beeinflusste damit auch die weitere Ausgestaltung des Standards.

Die Datenbank enthält Bilder von AVHRR SST (Sea Surface Temperature) und LST (Land Surface Temperature), ERS-2 SAR GEC und X-SAR Radarbilder sowie vorberechneten NDVI (Normalized Difference Vegetation Index). Für uns von besonderem Interesse sind die SST/LST-Daten. Aus ihnen wurde ein blattschnittfreie Karte von Europa und dem Mittelmeerraum erstellt und diese entlang der Zeitachse in die dritte Dimension erweitert. Das Ergebnis ist ein x/y/t-Würfel, der laut Auskunft der DLR derzeit aus ca. 10.000 Bildern besteht.

Der Zugang geschieht über eine Übersichtskarte (Abbildung 5), aus der durch Aufziehen einer Box eine Region of Interest (RoI) definiert werden kann bzw. durch zweimaliges Klicken ein Zoom-Punkt. Alternativ können zur Auswahl eines Gebiets Geokoordinaten angegeben werden. Das Ergebnis wird spezifiziert durch die Auswahl eines Lagebezugssystems (spatial reference system, SRS), die Auflösung in x, y und t, das Ausgabeformat (z. B. GeoTIFF) und eine optionale Ausgabemosaikierung und -kompression.

Die Anfragesprache erlaubt auf dem Datenwürfel Extraktionen in allen Dimensionen. Ausschnitte in der x/y-Ebene mit Schnitt in der t-Achse zeigt

Abbildung 6 (oberes Bild) als Ergebnis der Anfrage „SST-Ausschnitt um Griechenland mit spezifizierter Boundingbox und vorgegebenem Zeitpunkt“. Abbildung 6 (Mitte) zeigt das Ergebnis der Anfrage „Temperaturkurve über Moskau für das Jahr 1999“, also einen Schnitt in  $x$  und  $y$  mit Trimming in  $t$ . Eine strukturell identische dreijährige NDVI-Kurve über dem Nildelta ist in Abbildung 6 (unteres Bild) dargestellt.

Abbildung 7 zeigt einen mit dem rasdaman-Werkzeug `review` extrahierten raum-zeitlichen Ausschnitt, dem beispielhaft die Kachelstruktur eingezeichnet wurde. Bei dem perspektivisch dargestellten  $x/y/t$ -Würfel um Italien, Korsika, Sardinien und Sizilien erkennt man deutlich links entlang der Zeitachse die jahreszeitlichen Schwankungen der Meeresoberflächentemperatur.

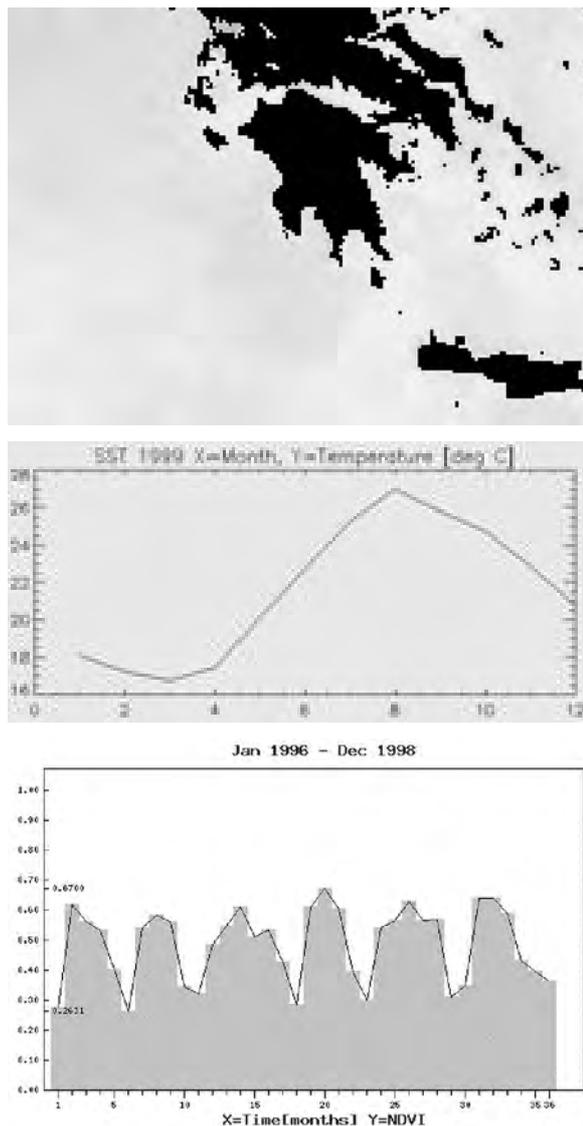
Eine konventionelle Ablage als horizontale Schichten, wie sie sich aus den eingefügten Bildern ergeben, wäre offensichtlich hochgradig ineffizient bei Temperatur- und NDVI-Kurve – eine ganze Schicht bzw. ein ganzes Bild wäre in den Hauptspeicher des Servers zu lesen, nur um jeweils ein Pixel daraus zu bestimmen. Um die verschiedenen Zugriffsmuster zu berücksichtigen, nämlich sowohl horizontale als auch vertikale Schnitte und Ausschnitte wurde, aufbauend auf früheren Untersuchungen (FURTADO & BAUMANN 1999; FURTADO et al. 1997) eine Kachelstruktur gewählt, welche den Gesamtwürfel in kleinere 3D-Teilwürfel zerlegt (Abbildung 7). Beim Import wird jeder Bildinhalt an seine korrekte Position gemäß Georeferenzierung und Zeitstempel einkopiert.

Mit der letzten Anfrage wird besonders augenfällig, wie die serverbasierte Datenanalyse Client und Netzwerk entlasten kann. Das Anfrageergebnis umfasst nur wenige Kilobyte, und nur diese werden vom Server zum Client transportiert. Eine Auswertung der Anfrage auf Clientseite würde sich alleine schon aus Gründen der Netzlast verbieten.

Derartige Techniken lassen sich selbstverständlich auch im Bereich Geophysik anwenden, wenn  $(x/y/z)$ -Daten statt  $(x/y/t)$ -Würfeln zu Grunde liegen.

### 3.3 Szenario 3: Klimamodellierung (4-D)

Supercomputer liefern ihre Simulationsergebnisse als 4D-Datensätze in vorgegebener Auflösung



**Abb. 6:** Anfrageergebnisse aus dem  $x/y/t$ -Würfel: (a) SST-Karte von Griechenland:  $x/y$ -Trim mit  $t$ -Schnitt; (b) Temperaturkurve über Moskau:  $t$ -Trim mit  $x/y$ -Trim; (c) NDVI-Kurve über dem Nildelta:  $t$ -Trim mit  $x/y$ -Trim.

(KLEESE et al. 2000); beispielsweise deckt das vergleichsweise niedrig auflösende Modell ECHAM5 T42 die gesamte Erdoberfläche in  $128 \times 64$  Pixeln und 17 Höhengschichten ab. Bei einer Simulationszeit von 200 Jahren und einer Auflösung von sechs Stunden ergeben sich 2.190.000 Zeitscheiben. Für jeden physikalischen Parameter („Variable“), etwa Temperatur und Luftfeuchtigkeit, entstehen so 2,5 TB. Jeder Simulationslauf kann bis zu 50 und mehr Variablen umfassen. Während die Simulation fortschreitet, werden erzeugte Zeitscheiben in Dateien ausgeschrieben, üblicherweise mehrere Scheiben pro Datei (Abbildung 8).

Die Nutzer interessieren sich für Fragen wie “Klimabedingungen über einer bestimmten Region und einem Zeitintervall“, Trendanalyse und Standardstatistik wie Maxima, Durchschnitt und Minima.

Abbildung 9 zeigt einen Schnitt in der z-Achse für  $z=z_{max}$ , der die horizontale Komponente der Jetstreams in der oberen Atmosphäre als 3D-Würfel visualisiert.

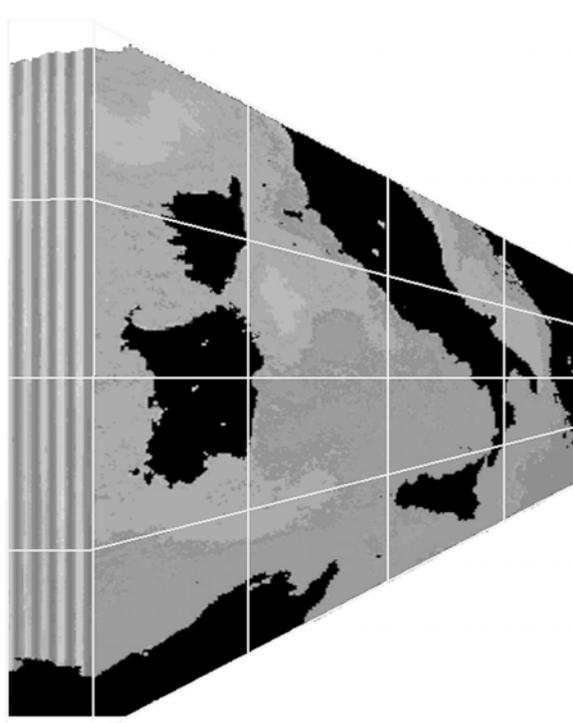
Eine wesentliche Herausforderung an die Datenhaltung ist dabei, das von der Simulation vorgegebene Einfüge-Muster (Zeitscheiben) zu entkoppeln von den Zugriffsmustern der Benutzer-Anfragen (üblicherweise Bereichsanfragen über die Zeit, also orthogonal zur Einfüge-Reihenfolge).

#### 4 Zusammenfassung

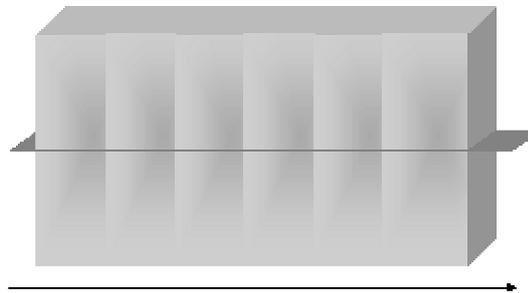
MDD gewinnen mehr und mehr Bedeutung im Bereich Online-Datenanalyse von Mess-, Bild- und Statistikdaten. Mehrere Gründe sprechen dafür, MDD-Unterstützung in den Kanon der Datenbank-Dienste einzubeziehen:

- Volle Informations-Integration: alle Information liegt logisch zentral an einem Platz, das DBS kann einen gewissen Grad an Konsistenz automatisch gewährleisten. Außerdem werden Aufgaben wie z.B. Archivierung deutlich vereinfacht und sicherer gemacht.
- Flexibilität der Dienste: durch die Anfragesprache lassen sich neue Fragestellungen ohne Programmierung direkt auf den Originaldaten beantworten.
- Wohldefinierte, offene Schnittstellen: sie erlauben die herstellerneutrale Kopplung verschiedener Werkzeuge, erlauben so die Auswahl der jeweils besten Werkzeuge und fördern außerdem die Herstellerunabhängigkeit.
- Nutzung vorhandener Dienste: Die über mehrere Jahrzehnte entwickelten Mechanismen der Standard-Datenbanken wie Trigger/Constraints, Multi-User- und Transaktions-Unterstützung, Backup-Mechanismen und viele mehr lassen sich direkt einsetzen und müssen nicht neu erfunden werden.

Das rasdaman-System ist generisch europäische Technologie, entstanden aus mehreren EU-Projekten und seit einiger Zeit im kommerziellen Einsatz z.B. bei Landesvermessungsämtern (www.rasdaman.de). Parallel wird die Forschung



**Abb. 7:** Kachelstruktur der 3D-AVHRR-Datenbank der DLR an Hand eines 3D-Ausschnitts (Kachelgröße nicht maßstabsgetreu).

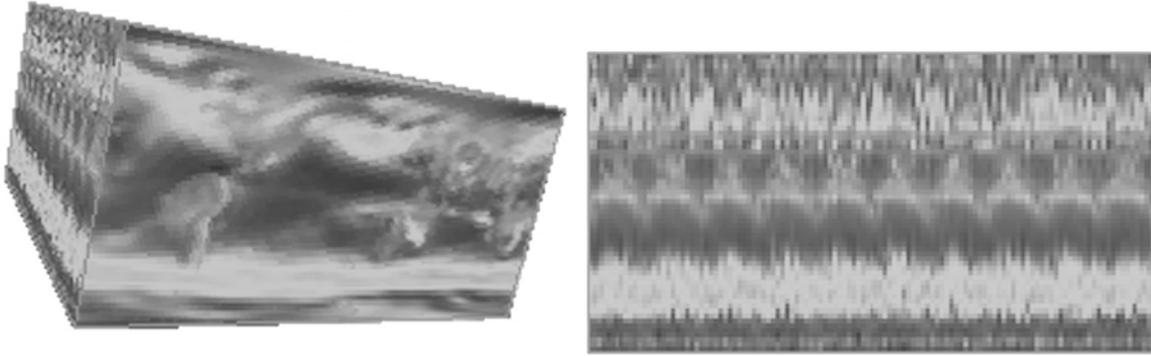


**Abb. 8:** Erzeugungsmuster bei Klimasimulationen (vertikale Zeitscheiben) und zeitorientierte Anfrage (horizontale Ebene).

an rasdaman fortgesetzt; aktuelle Themen sind unter anderem Anfrage-Parallelisierung und optimierte Tertiärspeicher-Anbindung (z.B. große Bandarchive).

Insgesamt ist das Feld der Rasterdatenbanken noch relativ jung, und eine Vielzahl von Fragestellungen erfordert weitere Forschungen. Exemplarisch seien hier einige wichtige Themen aufgezählt:

- Mächtigkeit von MDD-Anfragesprachen (wie wird sie bestimmt? Welche Erfordernisse ergeben sich aus der Praxis?)



**Abb. 9: 3D-Extrakt aus 4D-Klimamodell, Komponente horizontale Windgeschwindigkeit: Einfärbung (links) und Zeitverlauf (rechts).**

- Entwurfs-Methodiken, sinnvollerweise aufbauend auf und eingebettet in Standard-Techniken wie UML
- Standardisierte Benchmarks
- Bestimmung optimaler Tuning-Parameter unter vorgegebenen Applikations- und Hardwareprofilen
- Vereinheitlichte Modellierung von Bild- und Statistikdaten (insbesondere OLAP, vgl. z.B. [www.olapreport.com](http://www.olapreport.com))

## 5 Literatur

BAUMANN, P. (1999): A Database Array Algebra for Spatio-Temporal Data and Beyond. – Proc. NGITS'99, Zikhron-Yaakov, Israel, July 5-8, 1999

BAUMANN, P. (2001): Web-enabled Raster GIS Services for Large Image and Map Databases. "Image-Based Geospatial Databases". – 5th Int'l Workshop on Query Processing and Multimedia Issues in Distributed Systems (QPMIDS'2001), September 3-4, 2001, Munich, Germany.

FURTADO, P. & BAUMANN, P. (1999): Storage of Multidimensional Arrays based on Arbitrary Tiling. – 15<sup>th</sup> Int. Conf. on Data Engineering (ICDE), March 23-26, Sydney, Australia, March 1999.

FURTADO, P. et al. (1997): Object-Oriented Design of a Database Engine for Multidimensional Discrete Data. – Proc. of the OOIS '97 Conference, Brisbane, Australia, November 1997.

HAHN, K. et al. (2002): Parallel Query Support for Multidimensional Data: Inter-object Parallelism. – 13<sup>th</sup> International Conference on Database and Expert Systems Applications (DEXA), September 2-6, 2002, Aix en Provence, France.

KLEESE, K. et al. (2000): Intelligent Support for High I/O Requirements of Leading Edge Scientific Codes on High-End Computing Systems - the ESTEDI Project. – Sixth European SGI/Cray MPP Workshop, 7-8 September 2000, Manchester, UK.

LORIE, R.A. (1982): Issues in Databases for Design Applications. – In: ENCARNACAO, J. & KRAUSE, F.L. (Hrsg.): File Structures and Databases for CAD, North-Holland Publishing, IFIP, 1982.

REINER, B. et al. (2002): Hierarchical Storage Support and Management for Large-Scale Multidimensional Array Database Management Systems. – 13<sup>th</sup> International Conference on Database and Expert Systems Applications (DEXA), September 2-6, 2002, Aix en Provence, France.

RITSCH, R. (1999): Optimization and Evaluation of Array Queries in Database Management Systems. PhD Thesis, Technische Universität München.

ROLAND, P. et al.: A database generator for human brain imaging. – Trends in Neurosciences, **24**(10): 562-564.

TAMURA, H. (1980): Image Database Management for Pattern Information Processing Studies. – In: CHANG, FU (ed.): Pictorial Information Systems. LNCS 80, pp. 198-227 (Springer).

N.N.: [www.estedi.org](http://www.estedi.org), [www.neurogenerator.org](http://www.neurogenerator.org), [foremms.nr.no](http://foremms.nr.no), [euroclim.nr.no](http://euroclim.nr.no).