

Eine Datendrehscheibe für Raster-Massendaten

Markus von Brevern

toposoft GmbH
Kandelfeldstraße 82
52074 Aachen
mvp@toposoft.de

Abstract: Vorgestellt wird eine vollautomatische Datendrehscheibe für Raster-Massendaten, die im praktischen Einsatz ist. Diese importiert automatisch Wetterprognosen (Lufttemperatur und Niederschlag über FTP). Anfragen erfolgen zeilenorientiert für Einzelpunkte oder Flächenmittel oder -summen. Die Ergebnisse können in Modelle einfließen, z.B. zur Hochwasserprognose für Teileinzugsgebiete, oder sind Grafiken (PNG) zum Einbinden in eine Webseite.

1 Übersicht

Rasterdaten sind Massendaten. Sie umfassen oft mehrere 10.000 Punkte und entstehen ca. 9.000 Mal im Jahr mit teilweise über 100 Größen (bei Ensembledaten). Im Beispiel wird ein Raster mit 51.000 Punkten betrachtet, das jede Stunde erzeugt wird. Die zu importierenden Daten enthalten jeweils Prognosen zu den kommenden 48 Stunden. Die Größe der Datenbank beträgt ca. 2GB pro Messgröße und Jahr. Gespeichert werden Lufttemperatur und Niederschlag.

Gezeigt wird der automatische Datenfluss und die Datenspeicherung, die gleichzeitig das Langzeitarchiv darstellt. Datenlieferanten sind ein FTP-Server der Zentralanstalt für Meteorologie und Geodynamik in Österreich und ein Netz von Niederschlagsmessstellen. Datennutzer sind ein Hochwassermodell, das jede Stunde eine Prognose erstellt, und ein Dienst im Web, der Zeitreihen-Grafiken bereitstellt.

An der Drehscheibe beteiligt sind die Datenbanken für Raster-Zeitreihen (toporast-DB) und die für Zeitreihen (ZR-DB). Die Datenbanken werden jeweils durch einen Importdienst mit Daten beliefert (Raster-Import bzw. Callisto). Auf die Datenbanken greift ein Zeitreihen-Server zu (tstpd). Dieser zieht Zeitreihen aus toporast und aus der ZR-DB. Er schreibt auch Zeitreihen in die ZR-DB zurück. Über eine weitere Schnittstelle stellt tstpd Grafiken im PNG-Format bereit. In Abb. 1 ist dies veranschaulicht.

Das Beispiel Hochwasserprognose ist aus der Wasserwirtschaft, lässt sich aber auch auf die Agrarwirtschaft übertragen. Ob die Shapes Einzugsgebiete oder Feldblöcke oder Windparkumrisse enthalten, spielt für die Datendrehscheibe keine Rolle.

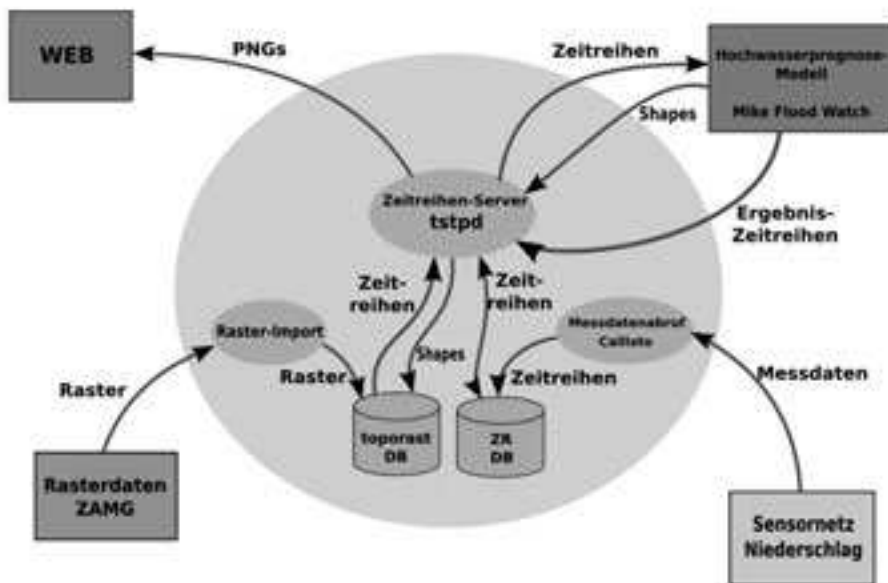


Abb. 1: Datendrehscheibe für Raster und Zeitreihen

2 Import

Klimatologische Rasterdaten sind überwiegend Produkte von nationalen oder multinationalen Wetterdiensten, wie dem DWD, der ZAMG oder dem ECMWF (European Centre for Medium-Range Weather Forecasts). Diese entwickeln und betreiben numerische Modelle und benutzen verschiedene Datenquellen (z.B. Radardaten und Sensornetze), um Prognosen über das Klima zu erstellen. Die Rasterdaten werden auf FTP-Servern zur automatischen Auswertung bereitgestellt und laufend aktualisiert.

Oft werden weit verbreitete binäre Formate wie NetCDF, Grib1 oder Grib2 verwendet. Manchmal wird ein ASCII-Format benutzt, das in komprimierter Form (z.B. gzip) vorliegt. Letzteres setzt die ZAMG ein.

Raster-Import holt nun die Rasterdateien vom FTP-Server ab, packt die Raster aus und integriert sie in der toporast-DB. Dies geschieht per automatischem Job einmal pro Stunde. Die Datenbanken wachsen pro Stunde etwa um 200k, jährlich also um weniger als 2GB.

3 Datenstruktur

toporast speichert keine Raster, sondern zu jeder Kachel ein Zeitreihenstück. Die Zeitreihenstücke aller Kacheln bilden ein Cluster. Die Cluster in ihrem zeitlichen Verlauf bilden die Grundstruktur der toporast-DB.

Eine toporast-DB kann mehrere Parameter (Messgrößen) nebeneinander vorhalten. In den zu importierenden Rasterdateien befinden sich fallweise dutzende Parameter, von denen aber nur einer oder wenige importiert werden sollen. Der Satz an Parametern ist daher konfigurierbar. Der gewünschte Ausschnitt aus dem Gesamtraster ist ebenfalls konfigurierbar. Der Benutzer kann also an zentraler Stelle dafür sorgen, dass die Datenmenge erträglich bleibt.

Modelle liefern gegebenenfalls nicht nur einen Wert pro Parameter, sondern zu verschiedenen Modellrandannahmen jeweils einen. Diese Werte heißen Ensembles. Sie werden parallel vorgehalten und können einzeln oder gemeinsam abgerufen werden.

Neue Daten überschreiben Daten des selben Zeitpunkts. Die Prognosen enthalten Daten für die jeweils zukünftigen 48 Stunden (im Beispiel) und werden in der zeitlich richtigen Reihenfolge importiert. Daher überschreiben die ersten 47 Werte die schon bestehenden und hängen lediglich den Wert für die letzte Stunde hinten an. Von einer Prognose bleibt also immer der erste Wert übrig. Da die ersten Werte (der sog. „now cast“) besonders zuverlässig sind, entsteht das Langzeitarchiv also automatisch in sehr guter Qualität.

4 Datenanfragen

Anfragen an die toporast-DB erfolgen durch den Zeitreihen-Server tstpd. Dieser erhält die Aufträge wiederum vom Modell oder von Skripts, die Webseiten aufbauen. Das Modell schickt die Umrisse der gewünschten Flächen als Shape-Datei an den Zeitreihen-Server. Dieser leitet sie weiter an die toporast-DB. Die Umrisse (Polygone) werden mit den Rasterkacheln verschnitten. So wird das Gewicht jeder Kachel bestimmt, mit dem sie ins Ergebnis eingeht (siehe Abb. 2). Die Zeitreihenstücke der Cluster werden passend ausgeschnitten und die gewichteten Werte dann wahlweise gemittelt oder summiert. Die so gewonnenen Zeitreihen werden über den Zeitreihen-Server als XML-Daten an das HW-Modell geschickt.

Das HW-Modell erstellt mit den Zeitreihen aus der toporast-DB und weiteren Zeitreihen, die aus dem Sensornetz abgerufen wurden, eine Hochwasserprognose. Deren Ergebnis besteht wiederum aus Zeitreihen (Abflüsse in m^3/s und Wasserstände). Diese werden über den Zeitreihenserver zurück in die ZR-DB gespielt. Auch diese Zeitreihen stehen dann bereit, um von Skripts zur Webseiten-Produktion aufgenommen zu werden.



Abb. 2: Gewichtermittlung beim Verschneiden der Kacheln mit den Flächenumrissen

5 Dynamische Aspekte

Im Beispiel liegen Klimaprognosen in einem Raster mit Kacheln zu 1 km² für die nächsten 48 Stunden vor. Das numerische Modell reicht in dieser räumlichen Auflösung nicht weiter in die Zukunft. Ein anderes Modell mit einer Kachelgröße von 92 km² erlaubt jedoch Prognosen, die die zukünftigen 168 Stunden umfassen. Diese Daten werden ebenfalls importiert und in separaten toporast-DB abgelegt. Das Hochwasserprognose-Modell stelle nur eine Anfrage über 168 Stunden und übermittelt eine Shape-Datei. Der Zeitreihen-Server fragt die ersten 48 Stunden bei der 1km²-toporast-DB an und die restlichen 120 Stunden bei der 92 km²-toporast-DB. Als Ergebnis übermittelt er durchgehende Zeitreihen.

Die Verschneidung der Flächenpolygone mit den Rasterkacheln erfolgt dynamisch. Beim Aufbau der toporast-DB muss also nicht darauf geachtet werden, zu welchen Flächen Anfragen erfolgen. Denkbar wäre also auch, die toporast-DB mit einer Web-Schnittstelle zu versehen, an die die Benutzer die Shapes frei übergeben könnten (Shape hochladen, Zeitreihen als Grafik oder im CSV-Format erhalten).

Die Umrechnung der Koordinatensysteme (UTM, WGS84 usw.) erfolgt automatisch.