International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report IR-07-018

The Basin of Mexico Hydrogeological Database: Implementation, queries and interaction with open source software

Jaime J. Carrera-Hernández (jaime.carrera@mail.mcgill.ca)

Approved by

Marek Makowski (marek@iiasa.ac.at)
Leader, Integrated Modeling Environment Project

November 2007

Foreword

This report describes a part of the research the author made as a part of his one-year Colosio scholarship for which he joined the team of the Integrated Modeling Environment (IME) project. The long-term aim of the research is to provide science-based support for groundwater management in the Mexico City Basin.

Groundwater represents the main water source in the Basin of Mexico, where the Mexico City Metropolitan Zone (MCMZ) and its population of nearly 20 million are located. The Basin's aquifer system provides nearly 60% of the Basin's total water supply. This situation has caused a regional drawdown of the groundwater potentiometric level. Despite the importance of the role the aquifer plays on the Basin's water supply, to date no regional studies have been developed in this area and the regional dynamics of aquifer recharge has not been analyzed. A regional groundwater flow model is needed in order to improve the understanding of the regional groundwater flow patterns and to analyze the impacts of different extraction policies, and of urban growth on aquifer recharge.

The first step that needs to be taken for developing such regional groundwater flow model is to design and implement a framework for adequate handling the required hydrogeological data. This step is documented in this interim report. A paper reporting an application of this database to the groundwater management in the Mexico City Basin is available from the Hydrogeology Journal.[1]

The main problems that needed to be solved during the reported research are:

- collecting data from different sources in diversified formats,
- designing a consistent and efficient data structure,
- implementing the data base in PostgreSQL.[2]

The report not only provides the technical details needed for effective use of the developed database, but also shares experience that is valuable for developers of databases for actual decision-support systems that deal with spatial problems.

More information about the author's research is available from the IME project Website.[3]

[1]The paper DOI: 10.1007/s10040-007-0194-9, available from http://www.springerlink.com/content/3g0302175223pk7k/.

[2]PostgreSQL was selected because it is one of the best DBMS available for free for research, and for non-commercial applications.

[3]http://www.iiasa.ac.at/Research/IME.

Abstract

Integrated Water Management at the Basin level concept was introduced in the 1990s, and is a goal in every national and local water management plan. Unfortunately this goal has not been achieved mainly due to a lack of both tools and data management, as data must be gathered from different sources, and converted from diverse formats into a consistent database. Compounding this problem is the fact that in some regions different water agencies are in charge of water supply as is the case in the Basin of Mexico, in which Mexico City and its Metropolitan Zone are located. The inhabitants of the Basin of Mexico, which comprises five different political entities and in which different agencies are in charge of water supply rely on the Basin's aquifer system as its main water supply source.

No regional hydrogeological database in this area however exists therefore a Relational Database Management System was developed, and its use with a Geographic Information System is proposed in order to improve regional data management in the study area. Data stored in this new database (called the Basin of Mexico Hydrogeological Database) comprises data on climatological, borehole and runoff variables, readily providing information for the development of hydrogeological models. A simple example is used to show how geostatistical analysis can be done using the data directly from this database. The structure of the database supports easy maintenance and updating, representing a valuable tool for the development of regional studies.

Acknowledgments

This report describes part of the authors research at IIASA during 2006-2007 as a Colosio Fellow, which was a fruitful experience both on an academic and personal level. The author thanks Dr. Marek Makowski, Leader of the Integrated Modeling Environment Project for his support both before and during the author's research at IIASA.

About the author

Jaime Carrera was a Colosio Fellow at the IIASA Integrated Modeling Environment Project from September 2006 until August 2007, and was a participant in IIASA's Young Scientists Summer Program in 2004. He is a currently a postdoctoral fellow at the Earth and Atmospheric Sciences department at the University of Alberta.

He received his Ph.D. from the Department of Civil Engineering and Applied Mechanics from McGill University in November 2007, and holds a Masters Degree in Hydraulic Engineering from the *Universidad Nacional Autónoma de México* (UNAM), from where he also graduated as a Civil Engineer.

His research interests are water management, groundwater flow modeling and the interaction of land cover change with surface and subsurface hydrology. He is currently developing ground/surface water models to understand wetland dynamics in the Boreal Forest of Northern Alberta.

Contents

List of Tables

List of Figures

The Basin of Mexico Hydrogeological Database: Implementation, queries and interaction with open source software

Jaime J. Carrera-Hernández (jaime.carrera@mail.mcgill.ca)

1 Introduction

The concept of water management evolved from a piece meal approach to that of a Basin level management after the International Conference on Water and Environment in 1992, which resulted in the Dublin statement (ICWE, 1992). This concept shifted to integrated water resources development and management in Rio, later in that same year as expressed in chapter 18 of Agenda 21 (UNCED, 1992). Behind this idea was the introduction of both land and water related aspects of water management at the Basin level as well as stakeholder participation. However, this seldom occurs in practice, mainly due to a lack of both adequate data management and proper tools to achieve an integrated river basin approach.

In an integrated water management approach, the development of regional hydrogeologic models are required in order to predict the impact of different land and water management policies in the future and all data required should be easily accessible to decision makers and modelers. Data accessibility to several users can be provided by a Relational Database Management System (RDMBS) and it can incorporate data such as location of wells, pumping rates, groundwater table elevation, lithology records, concentration of trace metals as well as chemical and physical parameters of the groundwater. The same database should include climatological variables (e.g. precipitation, temperature, evaporation) thus making it possible to undertake spatial and time series analysis of these variables.

The use of a Geographic Information System (GIS) can help to visualize and update the existing database as superposition of different thematic layers can be accomplished in order to verify existing data and existing modules can be used on the database (e.g. geostatistics). The database structured in this way can be queried with the use of Structured Query Language (SQL) statements and new tables can be formed from existing data. Although the use of a GIS can help to visualize existing data through the simultaneous display of different layers or 3-D views, statistical analysis can be done without the use of a GIS.

The relational database PostgreSQL (`http://www.postgresql.org`) was selected in the present work as it can be linked to the statistical language R (R Development Core Team, 2005) which is an open source project similar to the S language. R provides libraries for statistical analysis. Another advantage of PostgreSQL is that it can be linked to the Open Source GIS GRASS (`http://grass.itc.it`) which provides tools for raster, vector and point analysis as well as tools for image processing. In addition, PostgreSQL can handle spatial attributes such as points, polygons or lines by using the postGIS extension (http://www/refractions/postgis) which makes it possible to undertake spatial queries to the database.

The objective of this paper is to present a database management system for the Basin of Mexico, and also provide some technical details that are beyond the scope of journal articles. The database comprises data previously available in different locations and formats in order to make it available to all interested users. This database can be accessed using Open Source software, freely available from the internet in order to make data accessible to those who can not acquire proprietary software due to its cost. The database can be easily updated and it has been used to determine the daily spatial variation of rainfall and both minimum and maximum temperature for 12 years in the Basin of Mexico (Carrera-Hernández and Gaskin, 2007b) and to develop the first regional analysis of the groundwater potentiometric level in the same area (Carrera-Hernández and Gaskin, 2007a).

2 The Basin of Mexico

The Basin of Mexico (referred to as *the Basin* in the remaining part of this paper) with an approximate area of 9,600 km^2 encloses one of the largest cities in the world: Mexico City and its Metropolitan Zone (MCMZ). The Basin is located in the central part of Mexico and is enclosed by mountains as high as 5500 masl (Fig. 1) while the valley's mean elevation, where Mexico City is located is near 2240 masl.

The Basin's aquifer system is the most important part of the water supply system to its inhabitants. Its exploitation started in 1847, when the first well was drilled (Ortega and Farvolden, 1989) a number which by 1990 had increased to 3537 officially registered wells in the MCMZ (NRC, 1995). This heavy dependence on the aquifers has had its toll and a decline in the potentiometric level of up to 80 meters was recorded by 2002 in some areas (Edmunds et al., 2002). Compounding this problem, the Basin comprises five different political entities (Fig. 1); accordingly, different governmental agencies are in charge of water supply, the most important being the *Comisión Nacional del Agua* (CNA) and the *Dirección General de Construcción y Operación Hidráulica* (DGCOH). The CNA has under its charge the *Gerencia Regional de Aguas del Valle de México* (GRAVAMEX) which in conjunction with the DGCOH operates the water supply infrastructure for the MCMZ. However, water management at the basin level is not fulfilled as these agencies operate on their own, making it difficult to share information between them.

The need for accessible and up-to-date data at the Basin level is shown in recent studies undertaken within the Basin of Mexico as they have considered only subareas of the Basin or rely on short term records such as Birkle et al. (1998) who used rainfall data for the 1980-1985 period to develop a "long-term" water balance.

3 Improving data management

Data required for any type of surface or groundwater study in the Basin are currently spread throughout different agencies in charge of water supply and even within these agencies data are found in different reports and diverse formats. Furthermore, the existing databases are limited to particular data such as climatological or run-off data. In order to improve water management in the Basin and to foster an Integrated Water Management approach in the study area, the Basin of Mexico Hydrogeological Database (BMHDB) has been developed using both a Relational Database Management System (RDBMS) and a Geographic Information System (GIS). Regarding well related data, the

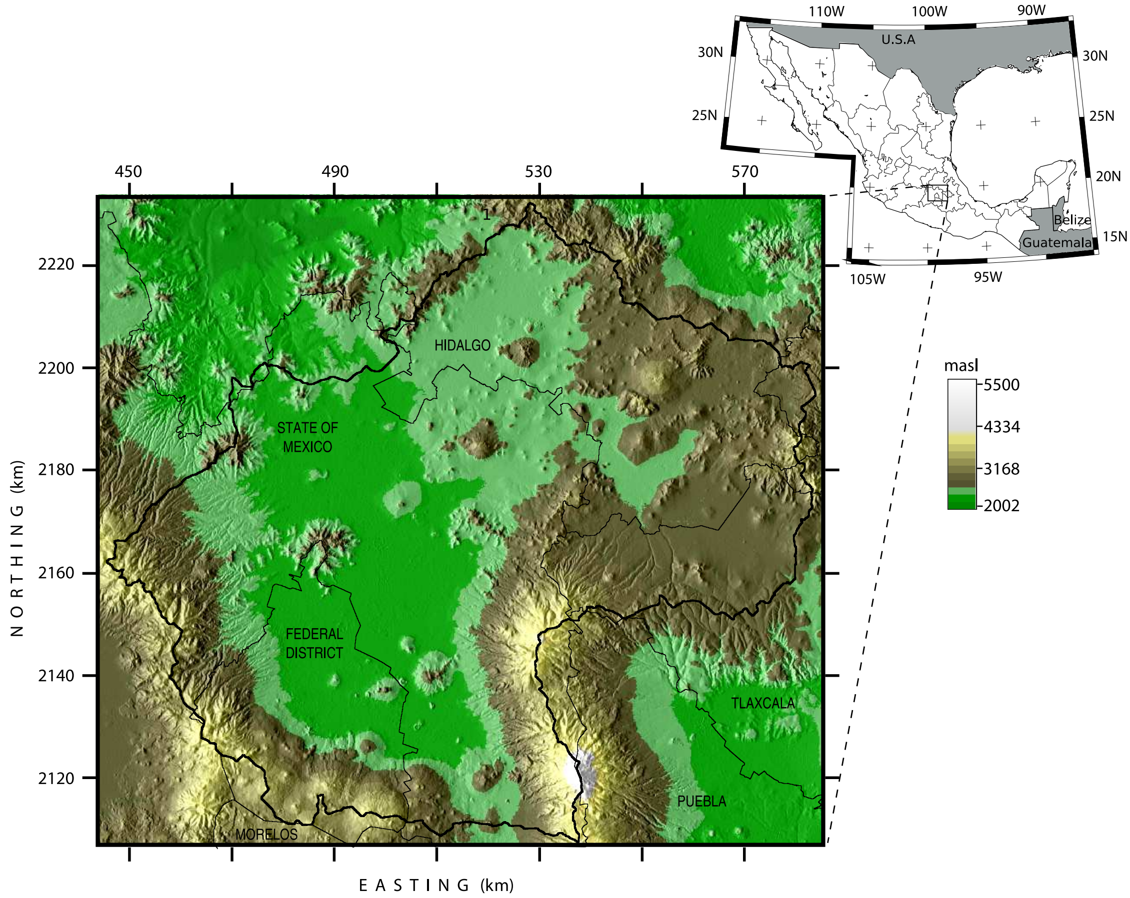


Figure 1: Location, Digital Elevation Model and political boundaries for the Basin of Mexico. Coordinates are in km for UTM–14; elevations are in meters above sea level.

BMHDB comprises monitoring wells from both CNA and DGCOH, allowed extraction volumes from those wells registered at the *Registro Público de Derechos de Agua* (REPDA) and annual extraction rates for those wells registered at DGCOH. Additionally, the database contains lithology records, pumping tests and chemical data for some of the wells. Climatological data (i.e. rainfall, minimum and maximum temperature) are available on a daily, monthly and annual basis, as are run-off date (volumes and flows). This information, which is currently distributed in the water supply agencies (DGCOH, CNA) and in previous studies realized in other areas of the Basin was gathered in order to develop this new database. The information gathered so far was obtained in different formats, such as spreadsheet files, shape files, hard-copy maps (e.g. soils and land-use), hand written tables and reports, which had to be processed and georefenced in order to provide readily accessible data.

3.1 Existing databases

There are currently two databases in Mexico which contain data required in hydrogeological studies. The databases that are available to any user are the *Extractor Rápido de Información Meteorológica* (ERIC), (IMTA, 1990) and *Banco Nacional de Datos de Aguas Superficiales* (BANDAS), (IMTA, 1995) which are briefly explained below.

1. ERIC: This database is distributed on one CD which includes nation-wide daily meteorological data: Rainfall, pan evaporation, minimum temperature, maximum temperature, average temperature, indicators of storm (0=no storm 1=storm), overcast conditions (0,1,2) and hail. The data stored on the CD have to be copied to the user's hard disk and accessed through the DOS command line. To query this database the user's input is required; the user needs to type the desired query in a specific order on the command-line: 1) variable selection (e.g. rainfall, evaporation), 2) station selection (one station, all stations, rectangular, polygon or state-wide selection) 3) time interval (one day, one period or one period over several years). The output of this query is an ASCII file with text and data which needs to be formatted and cleaned in order to undertake any type of statistical analysis.

2. BANDAS: This database was developed by the *Instituto Mexicano de Tecnología del Agua* and as ERIC, it comprises nation-wide data for Mexico. It is distributed as six CDs which are available from IMTA; the first of these CDs provides the installation program and is required in order to access the data. The information stored in this database is organized in 13 hydrological regions and in order to query it the user has to make a predetermined number of selections which can only be made through a scroll-menu. First the user has to select the hydrometric station of interest and then click on an icon to query the selected station which brings up another window. In this window the user is presented with different options through selection boxes which can not be selected simultaneously (and which in some cases are repetitive): monthly data, mean daily flows, flow records greater than a user-defined threshold, average and extreme annual flows, daily hydrometric data (flow and volume), monthly and annual hydrometric data. The way in which this database is structured makes it tedious to gather data for more than one year. As ERIC, the output file contains text and relevant data.

3.2 Drawbacks of the existing databases

In order to undertake any type of statistical analysis with the output data from either ERIC or BANDAS, the data have to be processed in order to clean from them additional information printed by these databases. In summary, the existing databases for climatological and stream-flow data have the following drawbacks:

- They need to be installed on computers running proprietary software, which means that they are not platform independent thus hindering their access
- The output of these databases has to be processed in order to be analyzed as it contains text (e.g. NA or sentences) within the data.
- The data stored on these databases comprise only a fraction of the required input in any type of hydrogeological study.

4 The Basin of Mexico Hydrogeological Database (BMHDB)

As previously explained, hydrogeological information is spread throughout different agencies in the Basin of Mexico. In order to improve Water Management in the Basin it is first suggested to improve data management through a comprehensive database system which provides remote access in order to facilitate its updating and maintenance.

The development of the Basin of Mexico Hydrogeological Database (BMHDB) comprised three main procedures as illustrated in Fig. 2 and which consisted of:

1. Data transcription: This stage consisted of transcribing the data acquired as hard-copy reports such as location of wells, lithology records and groundwater table elevations.

2. Data processing: Data from the existing databases or data provided in spread sheet formats were extracted and reformatted in a format usable by PostgresQL. Spatial properties were reformatted according to the requirements of PostGIS.

3. Map processing: Hard-copy maps (e.g. geology, land cover and edaphology) had to be digitized and georeferenced before being processed. The processed maps are stored as both vector and raster maps in the GRASS database.

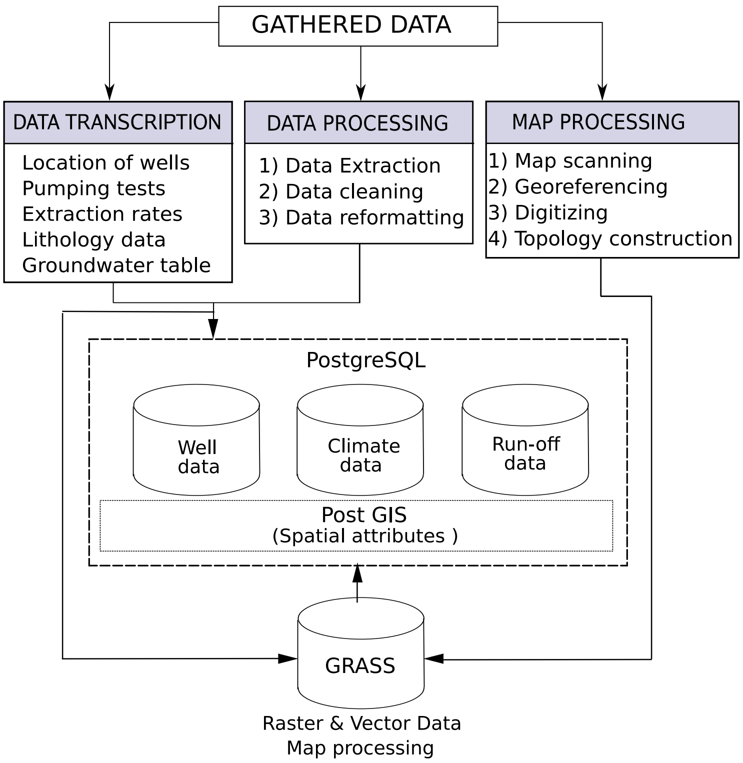
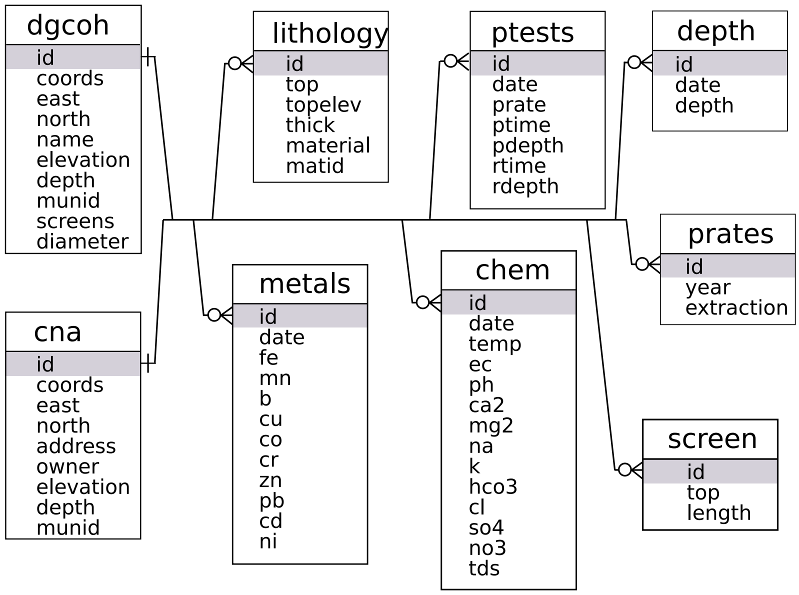


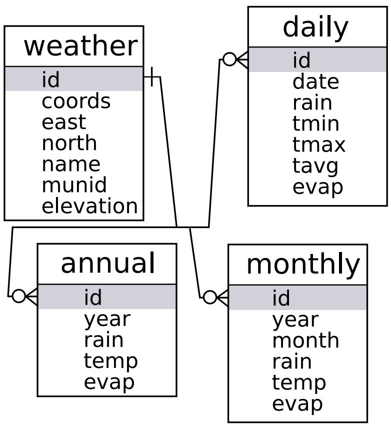
Figure 2: Development of the Basin of Mexico Hydrogeological Database and interaction of its different components

4.1 Data description

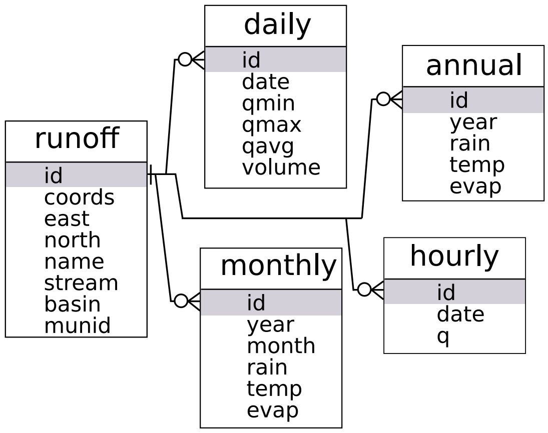
The structure of the BMHDB currently comprises thirteen tables as illustrated in Fig. 3 which can be divided in three subdatabases: climatological records, well records and run-off data. Some tables are organized by the agency which has the data (e.g. CNA or DGCOH) in order to facilitate the task of updating the database as it avoids duplication. The relevant fields of each table are shown in Fig. 3 and explained in Tables 2, 3 and 4. As the BMHDB is a relational database, all tables are related by the `id` fields of each well (Fig.3(a)), climatological station (Fig 3(b)) or gauging station (Fig. 3(c)).



(a) Well tables



(b) Climatological tables



(c) Run-off tables

Figure 3: Structure of the Basin of Mexico Hydrogeological Database

The BMHDB comprises data at the Basin scale; this was accomplished by gathering data from different governmental agencies, as illustrated by Fig. 4 which shows the areal coverage of the BMHDB well related data and the agency which holds relevant data for each well; it should be stressed that no attempt has been done to date in order to integrate all these data. This new database also integrates climatological and hydrometric information which were retrieved from BANDAS and ERIC (Fig. 5), improving the way in which this information can be retrieved and visualized. The BMHDB also comprises spatially variable information in both vector and raster format which can be used in distributed hydrogeological modeling such as to analyze the impact of urban growth on aquifer recharge. The development of such analysis requires land cover map for different years, as illustrated in Fig. 6, which shows such a map for both 1978 and 1985, the spatial distribution of geological and soil units in the Basin, as illustrated in Fig. 7 and the spatial distribution of rainfall. The spatial distribution of climatological variables in the Basin was developed by Carrera-Hernández and Gaskin (2007b) using their correlation with topography, through the use of local Kriging with External Drift to develop

Table 1: Data stored in the GIS database as raster and vector maps

Data	Scale	Type	Source
Land Cover	1:250 000	Paper map	INEGI F14-11 (Pachuca)
	1:250 000	Paper map	INEGI E14-2 (Mexico City)
Topography	1:250 000	Digital Elevation Model	Shuttle Radar Topography Mission
Surface Geology	1:100 000	Paper map	Mooser et al. (1996)
Edaphology	1:250 000	Paper map	INEGI F14-11 (Pachuca)
	1:250 000	Paper map	INEGI E14-2 (Mexico City)

rainfall maps and Kriging with External Drift for both minimum and maximum temperature in the study area. The overall goal behind the development of the BMHDB was to provide the basis for a regional groundwater flow model, which requires time-series data (e.g. groundwater levels), and a proper aquifer characterization, for which well lithology and surface geology are needed.

When developing the BMHDB, some data were missing from the original sources such as the elevation of each well; to complete the `wellsdgcoh` or `wellscna` tables with the `elev` field, the DEM was queried for those wells which did not have this information as explained in a later section.

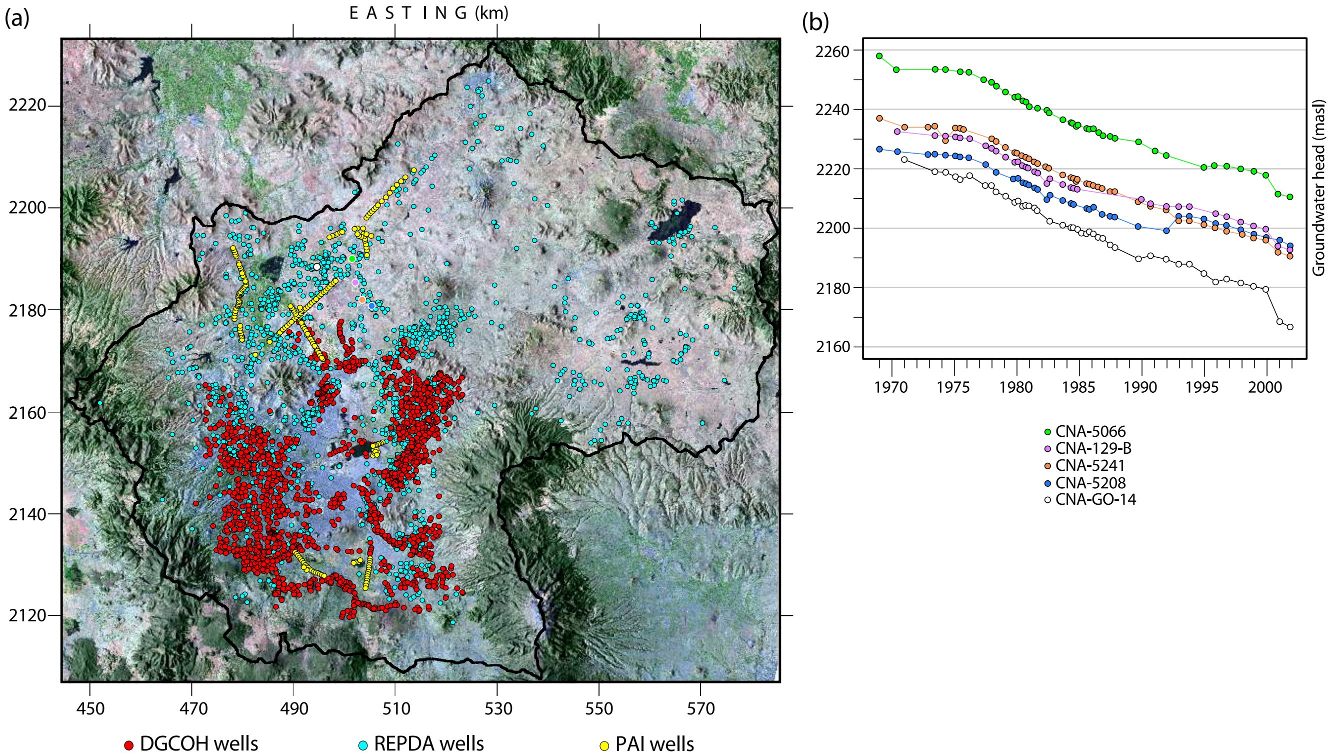


Figure 4: Well related data available in the Basin of Mexico Hydrogeological Database:(a) Spatial coverage of well data, color-coded according to the agency from which data were gathered,(b) evolution of the groundwater table elevation for five wells located in the *Tizayuca* region. The wells are shown on a false color composite derived from LANDSAT-ETM+ imagery for March, 2000.

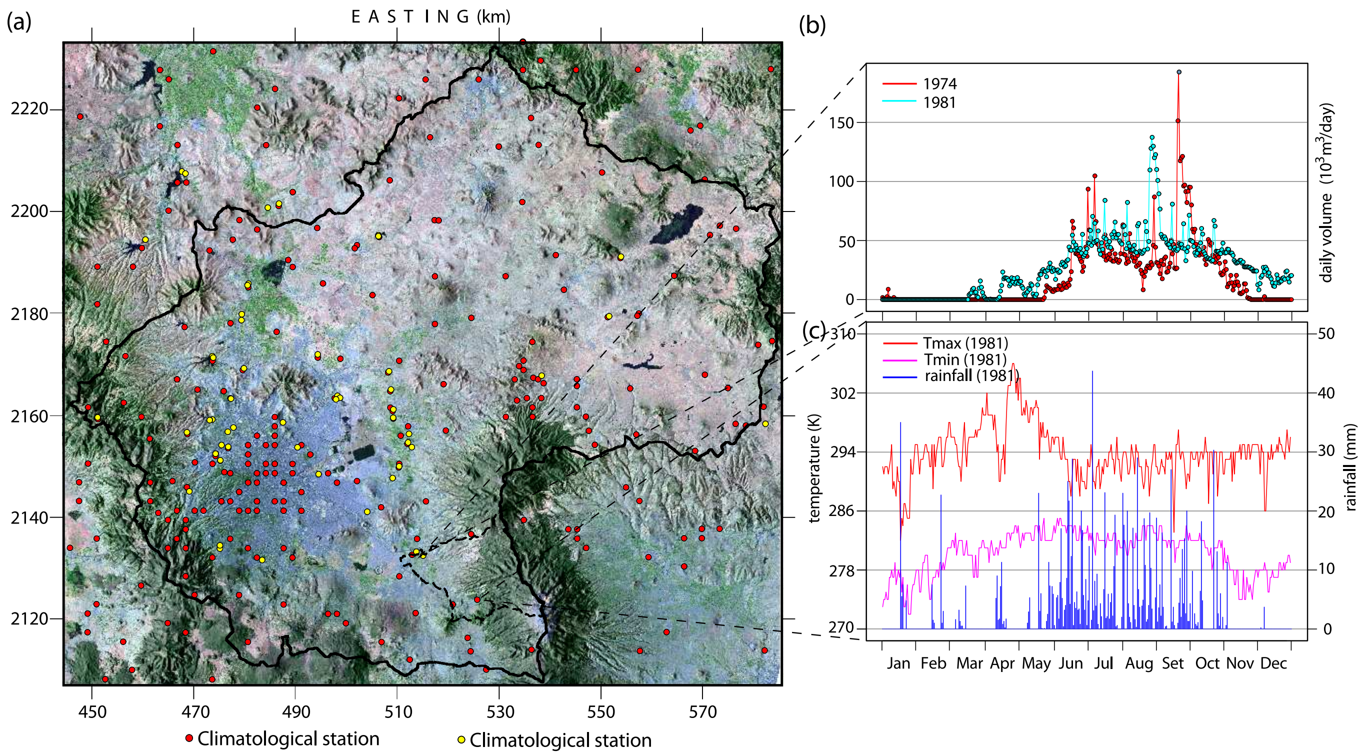


Figure 5: Surface water related data: (a) Spatial distribution of climatological and hydrometric stations, (b) daily river flow volume and (c) daily climatological data. Coordinates are in UTM, zone 14.

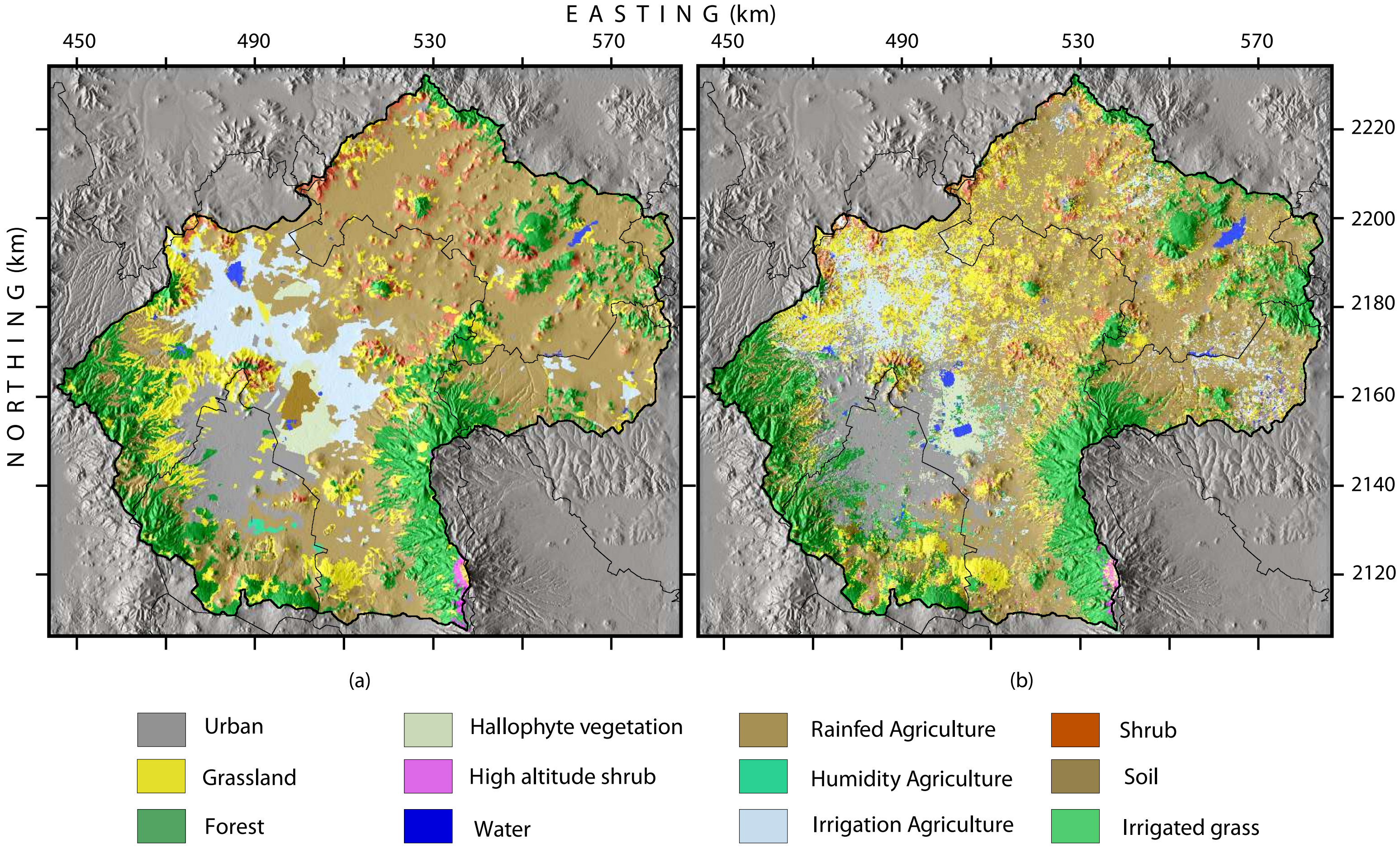


Figure 6: Land Cover map for the Basin of Mexico for two different years: (a) 1978 and (b) 1985. Coordinates are in UTM zone 14

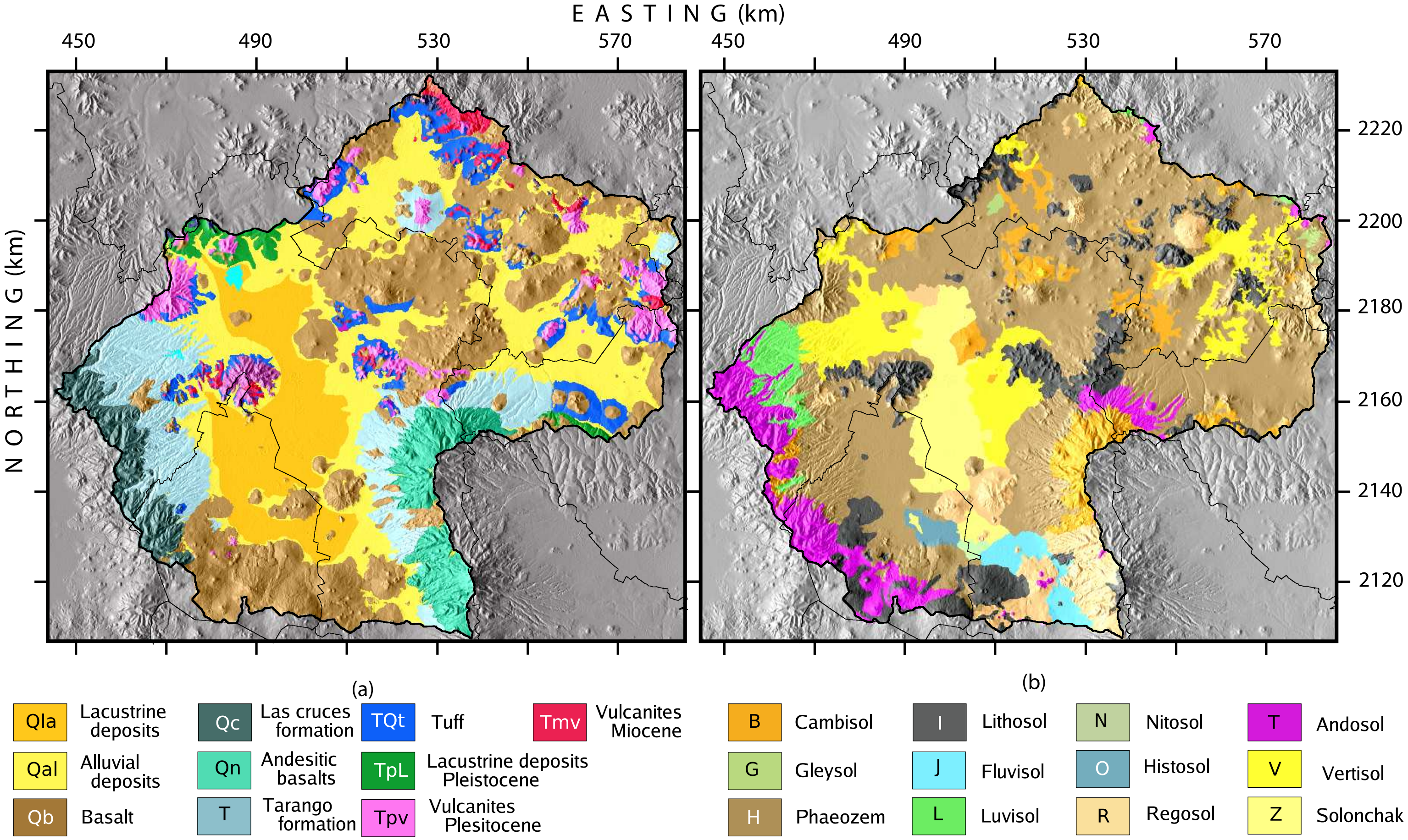


Figure 7: Spatial data stored in the GRASS database as raster maps, originally available as hard copy maps: (a) Surface geology adapted from Mooser et al. (1996) and (b) Soil units in the Basin.

4.2 Socioeconomic data

Sustainable water management also comprises both social and economic aspects, therefore the database can also be extended to include these types of data. Socioeconomic data are available on a municipal basis, thus a subdatabase with a main table called *municipalities* can be linked to the tables described in the previous section through the *munid* field. The data currently stored in the municipal database is shown in Fig. 8, which can be used when analyzing water demand. For the sake of brevity, a detailed description of each field is omitted.

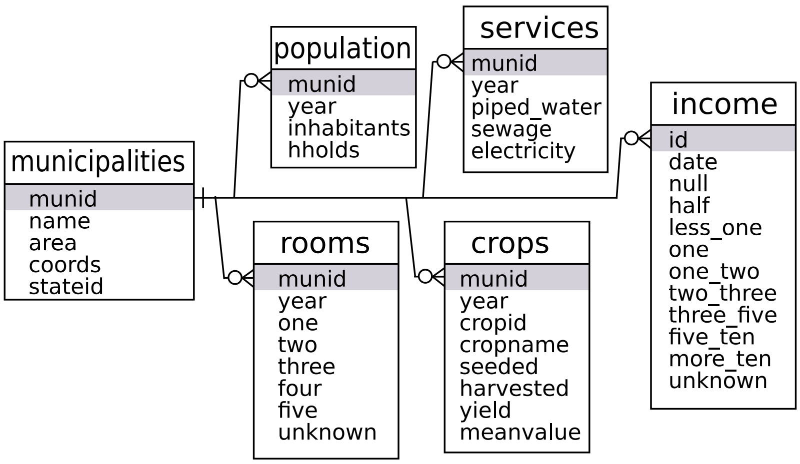


Figure 8: Structure of the municipal-socioeconomic database

5 Querying the database

The BMHDB can be queried by using SQL statements for which knowledge of the database structure is required (Fig. 3 and Tables 2, 3 and 4). The information provided in this section aims to illustrating how the database can be queried and by no means aims to provide a review of SQL statements; interested readers are encouraged to read the PostgreSQL and postGIS documentation in order to undertake more complex queries. Generally speaking, the procedure to analyze the data stored in the database can be summarized in three steps:

1. the database must be queried and a new table is written with the data of interest (this can be done either within GRASS, R or the psql command line) using SQL statements,

2. the new table is used as input for the interpolation procedure, and

3. the resulting map is saved as a raster file.

An example is developed in the next section in order to illustrate the procedure.

In order to build the BMHDB, data stored in it was queried to obtain the elevation of wells or hydrometric stations. Borehole information compiled from the different water agencies in charge of water supply in the Basin comprised data on depth to the water table while each geologic stratum had an associated depth and not its elevation which is required in order to characterize the aquifer. To obtain the elevation of each borehole the DEM was queried using each borehole's coordinates in order to get their elevation which was stored as a table with two fields: id and elevation. The elevation column was added to the well's main table as column elevation (Fig. 3(a) and Table 2). A new table was created in order to account for the elevation of each stratum by using the following SQL command:

```
BMHDB=# CREATE TABLE lithodgcoh AS SELECT l.id, l.top
(d.elev-l.top) AS topelev, l.thick, l.material, l.matid
FROM lithology AS l, dgcoh AS d
WHERE l.id=d.id;
```

The above command creates a table named lithodgcoh with fields id, top, topelev, thick, material and matid. The field topelev is computed by subtracting the top depth of each stratum to the DEM elevation at the corresponding well (accomplished by the WHERE condition of the SQL statement).

5.1 Spatial queries

The BMHDB is a spatial database, i.e., the data have spatial attributes such as coordinates (e.g. x-y, lat-lon), spatial reference (e.g. UTM zone) and datum (e.g. NAD–27). This information is stored as geometry by postGIS and allows queries involving spatial information. In order to include the munid field in the dgcoh table, the following spatial query was used:

```
BMHDB=# SELECT w.id, m.id as munid
FROM pai as w, muni_poly as m
WHERE w.coords && m.municipality
AND contains (m.municipality,w.coords) and m.id<30000
ORDER by w.id;
```

This query selects the well and municipality id fields for those wells located inside a municipality whose id is less than 30000 and evidently, other spatial queries can easily be done. Let us assume that a user wants to analyze the lithology records of those wells located in the quaternary lacustrine deposits (Qla; Fig. 7(a)); this is accomplished by using the postGIS function contains, the surface geology vector map and the well database as follows:

```
BMHDB=# SELECT w.id FROM dgcoh as w, geology as g
WHERE w.coords && g.vector
AND contains (g.vector,w.coords) AND g.cat=1
ORDER by w.id;
```

Through the previous SQL command, those wells that are enclosed by polygons of category 1 (where category 1 = Qal) from table dgcoh are selected. The output is ordered by the well's field id. In this manner only those wells located inside one or more polygons can be found.

6 Interaction with other open source software

The BMHDB was developed for providing readily accessible data for hydrogeological studies, which can be done by using different open source pieces of software. In order to show how these analyses can be undertaken, the current section provides simple examples on how data stored on the BMHDB can be readily accessed and analyzed using annual rainfall for 1979.

6.1 Geostatistical analysis of data

The R statistical software (R Development Core Team, 2005) provides tools for classical statistical tests, time series analysis and spatial interpolations as different libraries can be used to undertake these tasks. One of these libraries is the GSTAT library (Pebesma, 2004) which can be used for spatial interpolation through different Kriging methods (i.e. Ordinary, Universal or local Kriging). Once the spatial distribution of a variable is obtained (i.e. temperature or rainfall), it can be stored as a GRASS raster map through the use of R's grass library (Bivand, 2000). This approach was used by Carrera-Hernández and Gaskin (2007b) to undertake a daily analysis of rainfall and both minimum and maximum temperature in the study area.

The methodology described is exemplified by (1) analyzing the correlation between annual accumulated rainfall and elevation for years 1972–1985 by developing a scattergram showing the correlation value between these two variables and (2) by developing a spatial map of accumulated rainfall for the Basin of Mexico in 1979 through the use of Kriging with External Drift (KED), with elevation as a secondary variable. Although different GUIs are available for R, its main advantage is that it can be used from the command line, providing flexibility and the capability of using scripts and accessing it in batch mode, allowing to undertake large amounts of statistical analysis. The commands required to access the database and analyze the correlation for the previously mentioned period are shown on Fig. 9 for which a brief description is given in order to illustrate the capabilities of the BMHDB. The goal is not to develop a brief tutorial and interested users are referred to the R project web page, which provides a listing of all the available packages and their documentation.

```
1      library(RPgSQL)
2      db.connect(host=NULL,port=NULL,dbname='BMHDB')
3      db.execute("SELECT␣a.id,a.year,a.rain,w.elevation,w.east,w.
           north␣FROM
4  annual␣as␣a,␣weather␣as␣w␣WHERE␣a.year>=1972␣AND␣a.year<=1983␣AND
5  a.id=w.id␣AND␣w.elevation␣is␣not␣null␣AND␣a.rain␣is␣not␣null",clear
      =F)
6      id<-db.read.column("id",as.is=F)
7      rain<-db.read.column("rain",as.is=F)
8      year<-db.read.column("year",as.is=F)
9      s1<-db.read.column("east",as.is=F)
10     s2<-db.read.column("north",as.is=F)
11     dem200<-db.read.column("elevation",as.is=F)
12     anualrain<-data.frame(id,year,east,north,dem200,rain)
13     xyplot(rain~dem200|year,data=anualrain,ylab="accumulated␣
           rainfall
14 (mm)",xlab="elevation␣(masl)",panel =
15 function(x,y){panel.xyplot(x,y,pch="+")+panel.abline(lm(y~x));grid.
      text(round(cor(x,y),2),x=unit(1,"mm"),y=unit(1,"npc")-unit(1,"mm
      "),just=c("left","top"))},layout=c(4,3),ylim=0:2000)
```

Figure 9: Analysis of annual accumulated rainfall in the Basin of Mexico using the R statistical language.

In order to be able to read and write to the GRASS database, R needs to be called from within GRASS from the command line, after which the required libraries must be loaded (Fig 9, line 1): spgrass6 is used to write/read data from GRASS, RPGSQL is used to access

PostgresQL and GSTAT is used for the spatial interpolation. In addition, R automatically loads other libraries such as `grid` and `lattice` which were used to plot the different correlation values for 1972–1983 in Fig. 10. Once the libraries have been loaded, the database is accessed from R (Fig. 9, line 2) which in this case is being accessed on a local computer and so both the host and port options are set to null values. The database is queried using standard SQL commands (Fig. 9, line 3) and a dataframe is created in order to ease the statistical analysis which can be done with R (Fig. 9, lines 4–10). The scattergram plot showing both the correlation line and value (Fig. 10) were computed with line 11 of Fig. 9.

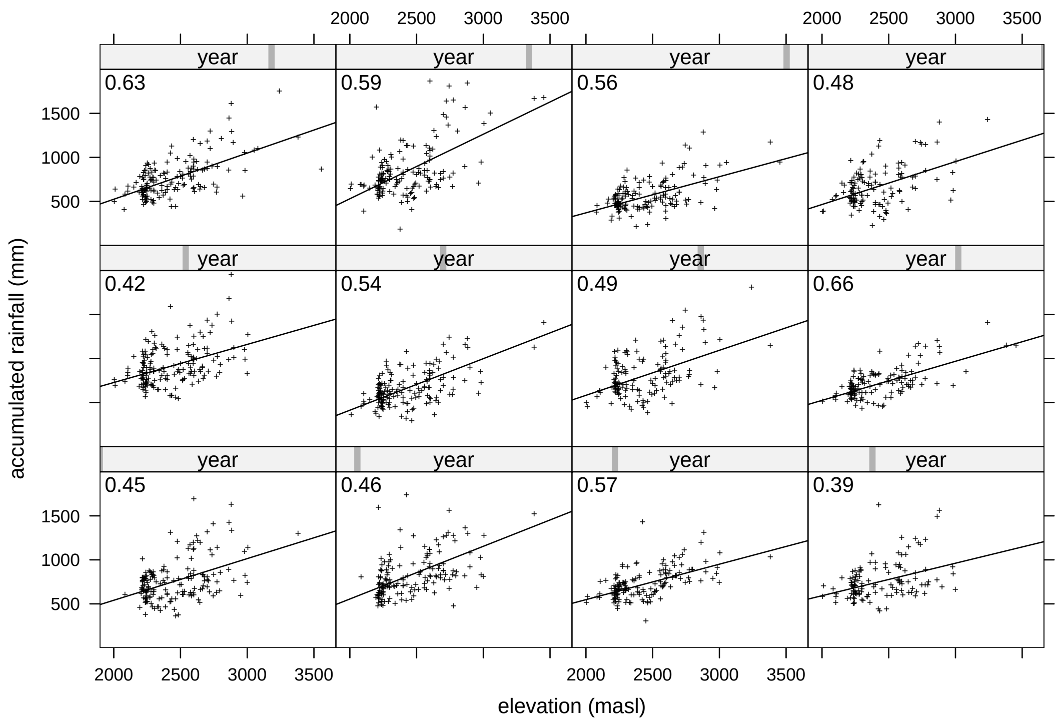


Figure 10: Correlation between annual accumulated rainfall and elevation for years 1972–1983.

As previously mentioned, Kriging with External Drift is used in order to develop the spatial pattern of rainfall for 1979 which was chosen as it is the year that exhibits the largest correlation value between rainfall and elevation (Fig. 10). The spatial interpolation is undertaken through the commands shown in Fig. 11 in which lines 1–3 are used to load the required libraries. The Digital Elevation Model (Fig. 1), which is stored in GRASS is used as an auxiliary variable in the use of KED and is read in line 4. Using the annual data retrieved in the previous step (Fig. 10), a new dataframe for the selected year is created to facilitate the example (Fig. 11, line 8). A semivariogram is computed, visualized and used to undertake the spatial interpolation (Fig. 11, lines 9–14) and plotted using R's `sp` library (Fig. 11, lines 15–17) as shown in Fig. 12. The interpolated surface can be written to the GRASS database (Fig. 11, line 18) in order to be used in further studies such as a spatially distributed water balance.

```
    library(spgrass6)
    library(gstat)
    library(RColorBrewer)
    elevation<-readCELL6sp("dem200")
    ccacoords<-coordinates(elevation)
    dem200<-elevation$dem200
    ccagrid<-data.frame(ccacoords,dem200)
    rain79<-data.frame(anualrain[anualrain$year==1979,])
    vgm79<-variogram(rain~dem200,~s1+s2,rain79,cutoff=40000)
    sil<-max(vgm79$gamma)
    nug<-min(vgm79$gamma)
    vgmfit79<-fit.variogram(vgm79,vgm(nug,"Exp",15000,sil))
    plot(vgm79,main="1979",model=vgmfit79)
rainked79<-krige(rain~dem200,locations=~s1+s2,data=rain79,model=vgmfit79,newdata=ccagrid)
    coordinates(rainked79)=~s1+s2
    gridded(rainked79)=TRUE
spplot(rainked79["var1.pred"],sp.layout=list("sp.points",stations,pch=19,cex=0.45,col="black"),pretty=TRUE,cuts=9,col.regions=brewer.pal(9,"Blues"),xlab="EASTING",
ylab="NORTHING",scales=list(draw=TRUE))
    writeRast6(rainked79,"kedrain79")
```

Figure 11: Geostatistical analysis of annual rainfall in the Basin of Mexico for 1979 using elevation as a secondary variable.

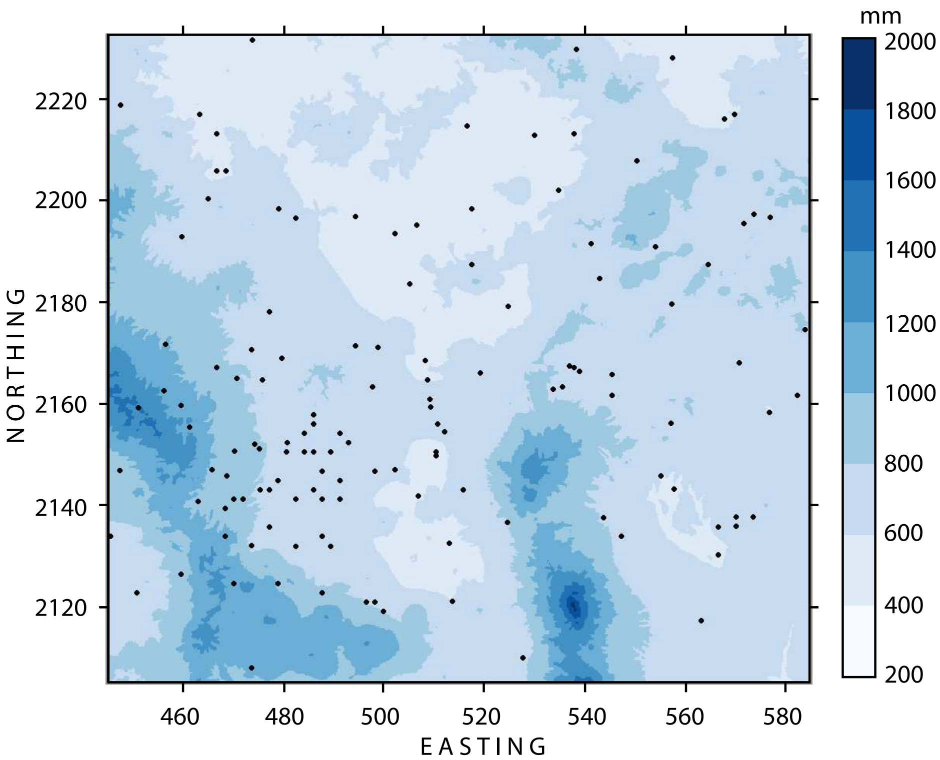


Figure 12: Spatial distribution of rainfall for 1979; black dots represent the climatological stations used to develop the interpolation. Coordinates are in km, UTM–14.

7 Discussion

New tools are required to improve water management at the Basin level, and to support a participatory water management approach. The BMHDB provides both local and remote access as it uses the RDBMS PostgresQL and can be queried with or without the use of a GIS as shown in this work. In addition, the database has been designed in a way that supports efficient statistical analysis without time-consuming data processing.

In order to undertake analysis based on the data stored in the BMHDB its structure needs to be transparent to the end users. The data tables are therefore related to each other as illustrated in Fig. 3, and all data have a spatial reference stored in the `coords` field which supports easy dealing with different coordinate systems.

The developed database facilitates the task of compiling information as it can be easily updated and accessed due to the RDBMS client-server capabilities. It is worth mentioning that data stored in the BMHDB comprises officially registered wells; however, non registered wells exist throughout the Basin. This adds another uncertainty factor to be considered when groundwater modeling is undertaken. Some information such as the wells' coordinates was verified with existing maps when available or by locating them in a map and checking if their location corresponded to that stored in the database.

The BMHDB provides data in spatial format thus allowing queries of spatial nature such as distance or polygon inclusion. Furthermore, the database can be processed with GRASS' raster and vector modules which allows the development of new maps by using map algebra for raster maps or unions and intersections for vector data. These modules can be used when developing groundwater flow models for calculating e.g., the width and length of a river reach within a finite difference model cell. Another advantage of the BMHDB is that its data can be used "as-is" to develop groundwater flow models through the use of the `r.gmtg` tool (Carrera-Hernández and Gaskin, 2006) which links the finite difference groundwater flow model MODFLOW with the GRASS GIS.

The development of the BMHDB represents a step towards Integrated Water Management in the Basin, as its data can be used to analyze the relationship between land cover change and groundwater; in addition, it also provides a framework to analyze water demand, as socioeconomic data can be also added to the database as shown in Fig. 8. Improving water management involves proper data management, a task that can be achieved using the framework presented in this paper as: (1) all tools used in the development of the BMHDB are open-source software, which can be downloaded from the Internet and (2) these tools are cross platform, meaning that they can be used on different operating systems without restrictions.

8 Conclusions

The Basin of Mexico Hydrogeological Database has been developed as a first step towards achieving an Integrated Water Resources Management approach in the Basin of Mexico. The BMHDB provides spatial data on soils, land cover, geology and climatological variables which can be used in regional hydrogeological studies without further processing. As exemplified in this work, the BMHDB can be accessed and analyzed with the use of open source software, freely available from the Internet and completely cross-platform: the database uses PostgreSQL (`www.postgresql.org`), the R statistical software (`www.r-project.org`) and the GRASS GIS (`grass.itc.it`).

The GIS visualization capabilities provide a means to verify the spatial location of point data and to communicate the result of future water management decisions obtained through modeling; however, a GIS is not required to analyze and visualize the data stored in the RDBMS as exemplified in this work.

Although this database was developed for the Basin of Mexico, this work provides a general framework on the development of spatial databases for hydrogeological modeling applicable to any watershed. The way in which the BMHDB is structured facilitates its application in both the development and calibration of distributed hydrological and groundwater models. The structure of the database allows it to be easily maintained and kept up to date allowing the inclusion of additional data as they become available.

The BMHDB represents the first effort in the study area to integrate a regional database, with the goal to develop a regional groundwater flow model. This work can be extended to include nation-wide data, which will help to improve water management in Mexico.

References

Birkle, P., Torres-Rodriguez, V., and González-Partida, E. (1998). The water balance for the Basin of the Valley of Mexico and implications for future water consumption. *Hydrogeology Journal*, 6:500–517.

Bivand, R. S. (2000). Using the R statistical data analysis language on GRASS 5.0 GIS database files. *Computers & Geosciences*, 26:1043–1052.

Carrera-Hernández, J. J. and Gaskin, S. J. (2006). The Groundwater Modeling Tool for GRASS (GMTG): Open Source Groundwater Flow Modeling. *Computers & Geosciences*, 32:339–351.

Carrera-Hernández, J. J. and Gaskin, S. J. (2007a). The Basin of Mexico aquifer system: regional groundwater level dynamics and database development. *Hydrogeology Journal*, 15(8):1577–1590. DOI: 10.1007/s10040-007-0194-9.

Carrera-Hernández, J. J. and Gaskin, S. J. (2007b). Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico. *Journal of Hydrology*, 336(3–4):231–249. DOI: 10.1016/j.jhydrol.2006.12.021.

CNA (2002). Registro público de derechos de uso de agua (public register of water rights). GRAVAMEX, internal data.

CNA (2003a). Groundwater levels gauged at the monitoring network in the basin of mexico. GRAVAMEX, internal data.

CNA (2003b). Lithology records from wells operated by cna. Several files, internal data.

DGCOH (1994). Diágnostico del estado presente de las aguas subterráneas de la ciudad de méxico y determinación de sus condiciones futuras (diagnosis of the present state of groundwater in mexico city and its future condition). Technical report, Instituto de Geofísica, UNAM.

DGCOH (1996). Actualización de parámetros hidrogeológicos y elaboración de pruebas de bombeo. Technical report, Consultores en Geología.

DGCOH (1997). Digitalización en ambiente cad y captura de datos para integrar la base de información de pozos a cargo de la dgcoh. Technical report, ORVI impulsora de Ingeniería.

DGCOH (1999). Estudio para actualizar la base de datos de la modelación del acuífero y simulación mediante modelos matemáticos aplicando diferentes políticas de explotación en el distrito federal. Technical report, DITAPSA Consultores, S.A. de C.V.

DGCOH (2000). Piezometría del valle de méxico. Technical report, Lesser y Asociados S.A. de C.V.

DGCOH (2002). Estudio geológico de la zona poniente del valle de méxico e interpretación estratigráfica de los pozos de agua a cargo de la dgcoh localizados en las delegciones de azcapotzalco, miguel hidalgo, álvaro obregón, magdalena contreras y cuajimalpa, méxico, d.f. Technical report, Ing. Luis Chavelas-Peña.

Edmunds, W. M., Carrillo-Rivera, J. J., and Cardona, A. (2002). Geochemical evolution of groundwater beneath mexico city. *J. of Hydrology*, 254:1–24.

Huizar-Álvarez, R. (1993). Simulacion matemática del sistema acuífero de chalco-amecameca, méxico (mathematical modelling of the chalco-amecameca aquifer system). *Geofísica Internacional*, 32(1):57–79.

ICWE (1992). International conference of water and environment. www.wmo.ch/web/homs/documents/english/icwedece.html. Accessed on August, 2005.

IMTA (1990). *Extractor Rápido de Información Climatológica (Fast Extractor of Climatological Information) (ERIC)*. Instituto Mexicano de Tecnología del Agua.

IMTA (1995). *Banco Nacional de Datos de Aguas Superficiales (National data of Surface waters) (BANDAS)*. Instituto Mexicano de Tecnología del Agua.

Mooser, R., Montiel, A., and Zúñiga, A. (1996). *Nuevo mapa geológico de las cuencas de México, Toluca y Puebla. Estratigrafía tectónica regional y aspectos geotérmicos (new gelogical map for the Basins of Mexico, Toluca and Puebla)*. Comisión Federal de Electricidad.

NRC (1995). *Mexico City's Water Supply: Improving the Outlook for sustainability*. National Academy of Sciences.

Ortega, A. and Farvolden, R. N. (1989). Computer analysis of regional groundwater flow and boundary conditions in the basin of mexico. *J. of Hydrology*, 110:271–294.

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30:683–691.

R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

UNCED (1992). United nations conference on environment and development. www.un.org/esa/sustdev/documents/agenda21/english/agenda21toc.htm. Accessed on August, 2005.

A Description of the database tables

Table 2: Description of the fields of the BMHDB borehole subdatabase

Table	Field	Data Type	Description	Source
cna and dgcoh	id	varchar	Borehole ID (prefix cna)	CNA (2003a),DGCOH (2000),DGCOH (2002), this paper
	east	long int	Easting of well	
	north	long int	Northing of well	
	munid	int	Municipality on which the well is located	
	owner	char	Owner of well	
	elevation	float	Well top elevation (masl)	
	coords	geometry	coordinates and elevation of climatological station with Spatial Reference ID number for UTM-14	
depth	id	varchar	Borehole ID	DGCOH (2000), CNA (2003a)
	date	timestamp	acquisition date	
	depth	float	depth to groundwater table	
metals	id	varchar	Borehole ID	DGCOH (1994), Huizar-Álvarez (1993),Edmunds et al. (2002)
	fe	float	Iron (mg/L)	
	mn	float	Manganese (mg/L)	
	b	float	Boron (mg/L)	
	cu	float	Copper (mg/L)	
	co	float	Cobalt (mg/L)	
	cr	float	Chromium (mg/L)	
	zn	float	Zinc (mg/L)	
	pb	float	Lead (mg/L)	
	cd	float	Cadmium (mg/L)	
	ni	float	Niquel (mg/L)	
chem	id	varchar	Borehole ID	Same as above
	date	timestamp	Acquisition date	
	temp	float	Temperature (° C)	
	ec	float	Electric Conductivity ($\mu\Omega/cm$)	
	ph	float	Phosphorous (mg/L)	

Table 2: Description of the fields of the BMHDB borehole subdatabase

Table	Field	Data Type	Description	Source
	ca2	float	Calcium (mg/L)	
	mn2	float	Manganese (mg/L)	
	na	float	Sodium (mg/L)	
	k	float	Potassium (mg/L)	
	HCO3	float	Bicarbonate (mg/L)	
	cl	float	Chlorine (mg/L)	
	so4	float	Sulphate (mg/L)	
	no3	float	Nitrate (mg/L)	
	tds	float	Total Dissolved Solids (mg/L)	
lithology	id	varchar	Borehole ID	DGCOH (2002),DGCOH (1997),DGCOH (1996), CNA (2003b)
	top	int	top depth of stratum	
	topelev	int	top elevation of stratum	
	thick	int	stratum thickness	
	material	varchar	material's letter code	
	matid	int	material's id number	
prates	id	varchar	Borehole ID	CNA (2002),DGCOH (1999)
	year	int	year	
	month	int	month	
	prate	float	pumping rate	
ptests	id	varchar	well id	DGCOH (1996)
	date	timestamp	acquisition date	
	prate	float	extraction rate (m^3/s)	
	ddown	float	observed drawdown (m)	
	recov	float	observed recovery (m)	
screen	id	varchar	well id	DGCOH (1997),DGCOH (1996)
	top	float	top elevation of screen	
	length	float	screen length	

Table 3: Structure of the fields of the BMHDB weather subdatabase.

Table	Field	Data Type	Description	Source
weather	id	int	station id number	IMTA (1990)

Table 3: Structure of the fields of the BMHDB weather sub-database.

Table	Field	Data Type	Description	Source
	east	longint	easting of climatological station (meters, UTM-14)	
	north	longint	northing of climatological station (meters, UTM-14)	
	name	varchar	name of climatological station	
	munid	int	id of municipality on which the station is located	
	elevation	int	elevation of climatological station (masl)	
	coords	geometry	coordinates and elevation of climatological station with Spatial Reference ID number for UTM-14.	
daily	id	int	station id number	IMTA (1990)
	date	timestamp	date of measurement	
	rain	float	daily rainfall (mm)	
	tmin	float	daily minimum temperature (°C)	
	tmax	float	daily maximum temperature (°C)	
	tavg	float	daily average temperature (°C)	
	evap	float	daily pan evaporation (mm)	
monthly	id	int	station id number	This paper
	year	int	year of measurement	
	month	int	month of measurement	
	rain	float	monthly rainfall (mm)	
	tavg	float	monthly average temperature (°C)	
	evap	float	monthly pan evaporation (mm)	
annual	id	int	station id number	This paper
	year	int	year of measurement	
	rain	float	annual rainfall (mm)	
	temp	float	annual temperature (°C)	
	evap	float	annual pan evaporation (mm)	

Table 4: Structure of the fields of the BMHDB run-off sub-database

Table	Field	Data Type	Description	Source
runoff	id	int	hydrometric station id number	IMTA (1995)
	east	longint	easting of station (UTM-14)	
	north	longint	northing of station (UTM	
	coords	geometry	coordinates and elevation of climatological station with Spatial Reference ID number for UTM-14	
	name	string	station name	
	stream	string	stream on which the station is located	
	basin	string	basin on which the station is located	
	munid	int	id of municipality on which the station is located	
hourly	id	int	hydrometric station id number	IMTA (1995)
	date	timestamp	date and time of acquisition date	
	flow	float	flow measurement (10^3 m^3)	
daily	id	int	hydrometric station id number	IMTA (1995)
	date	timestamp	acquisition date	
	qmin	float	daily minimum flow (m^3/s)	
	qmax	float	daily maximum flow (m^3/s)	
	qavg	float	daily average flow (m^3/s)	
	vol	float	daily volume (10^3 m^3)	
monthly	id	int	hydrometric station id number	IMTA (1995)
	year	int	year of flow	
	month	int	month	
	qmin	float	monthly minimum flow (m^3/s)	
	qmax	float	monthly maximum flow (m^3/s)	
	qavg	float	monthly average flow (m^3/s)	

Table 4: Structure of the fields of the BMHDB run-off sub-database

Table	Field	Data Type	Description	Source
	vol	float	monthly volume (10^3 m^3)	
annual	id	int	hydrometric station id number	IMTA (1995)
	year	int	year of flow	
	qmin	float	annual minimum flow (m^3/s)	
	qmax	float	annual maximum flow (m^3/s)	
	qavg	float	annual average flow (m^3/s)	
	vol	float	annual volume (10^3 m^3)	