CTMSOLAP: A Regional geo-decisional Support System Based on the SOLAP Approach and a Chemistry Transport Model

Youness OUBENAALLA (*), **Omar EL BEQQALI** (*)
oubenaalla@yahoo.fr, omarelbeqqali@gmail.com

(*) Department of Computer Sciences, Dhar Mahraz Faculty of Sciences, BP 17.96, Fez, Morocco

Mots clefs :
Decision-support system, SOLAP, regional air pollution, forecast atmospheric model, in situ captor, satellite observations
Keywords:
Decision-support system, SOLAP, regional air pollution, forecast atmospheric model, in situ captor, satellite observations

Abstract

Atmospheric pollution has become a serious growing and persisting problem. The problem is assuming alarming dimensions in many cities around the globe, causing effects on human health and environment [12,13]. This study describes the SOLAP approach we have adopted to build a spatial air pollution decision-support system for Morocco. As they are based on the multidimensional paradigm, SOLAP facilitates the interactive spatio-temporal exploration of data [30,31,34] The conceptual data model we have used has the originality of encompassing data provided by three completely different kinds of sources: a numerical forecast air pollution model, in-situ captors and satellites[35]. The load of data generated by the numerical model into a multidimensional database set the predictive nature of the target data warehouse. Indeed, the corresponding published factual measures can relate to past, present and future times series, which in turn can be obtained under realistic or hypothetical conditions. This provides a powerful tool for decision makers to study and anticipate potential environmental crisis and thus proactively prepare and plan for accurate and optimal actions. Cartographic representation of the results is then achieved by using the SOLAP approach. Thematic maps can be generated as a response to a particular query that the end user may submit. A range of options are offered to explore these charts in both space and time. Qualities of the design and implementation are finally assessed according to recognized criteria established by the OLAP/SOLAP community as scalability, time response, data exploration and intuitiveness.

1 Introduction

Multidimensional structures are powerful in supporting efficiently the decision-making process as they provide pertinent and in-time responses to specific users' queries. They provide aggregate information witch is re-arranged in the most optimal database schema. OLAP (One Line Analytical Processing) structures are easier to establish and to query than transactional structures called On Line Transactional Processing (OLTP). Deficiencies in these approaches are filled in by Spatial OLAP (SOLAP) technology that offers cartographic options that are provided neither by the traditional GIS nor by OLAP tools [30,31,34].

At the present time, the use of SOLAP technologies covers a large range of applications, for example, the projects conducted at "Center for Research in Geomatics at Laval University", related to forestry management, lorry driver networks and environmental health [34,31]. Other applications using SOLAP are developed at LIRIS to monitor the death rates for the different departments of France [28].

Within the framework of the monitoring of air pollution, some researches were performed like the work realized by [28,33] in LIRIS INSA Lyon-France and the Imperial College of London United Kingdom respectively [16].

In this study, we use the SOLAP approach to develop a spatial decision-support system for regional air pollution monitoring over Morocco. The data collection in our case is heterogeneous as it is obtained from three main sources: Data produced by a numerical atmospheric model, concentrations captured by in-situ detectors and observed by satellite [35]. Within this framework, we will present the SOLAP architecture we defined as well as the data integration mechanisms used for extraction, analysis and exploration purposes. We will define specific queries that can be either spatial or spatiotemporal. The cartographic results will be illustrated too. We will also discuss the benefits and advantages of this representation for supporting the decision making process, analysts or any end user queries.

2 General framework and case study

2.1 Numerical model

This study is primarily based on data obtained from an air pollution prediction model, this type of system called Chemistry Transport Model (CTM) is described in [27]. The model used is CHIMERE in its 2007 version, forced by a meteorological model MM5 developed in the United States by the University Of Pennsylvania (PSU) and the National Center for Atmospheric Research (NCAR) [8] at the regional scale.

The purpose of CHIMERE is to simulate gas-phase chemistry, aerosol formation, transport and deposition [6,10,24,25] at European and urban scales. It has been designed with the aim of both performing episodic and long-term simulations at various spatial scales, ranging from local to regional scale, on a personal computer or a workstation. The general performances of the model for the simulation of ozone and aerosols are given by [10,25].

CHIMERE proposes many different options for simulations which make it also a powerful research tool for testing parameterizations, hypotheses. Its use is relatively simple so long as input data is correctly provided. It can run with several vertical resolutions, and with a wide range of complexity. It can run with several chemical mechanisms, simplified or more complete[27].

The objective of this study is to realize long-term simulations of the air quality for Morocco. Being a country of 16 regions of a total surface of 710.850km², with a diversified weather; Mediterranean in the North, Atlantic on the West and Saharan in the South, of about 2000 km of coast and varied soil types, Morocco constitutes a very complex case study. Moreover, the population densities are different from a region to other so implying non uniform distributions of pollutants. Operations of nesting were then necessary on the regions of high pollution such Casablanca and the regions of Fez, Mohammedia and Rabat.

Many studies were conducted in this context all over the world, and such models are used in real time and in operational mode in many European regions. Being the official model for the air quality monitoring and prediction on France and the region of Paris [3,11,26]. Other establishments use the models for the monitoring of the air quality in their regions (simulations and forecasts over different regions and country: China [8] Italy [1], Portugal [4,5], Switzerland [9], Spain [17].

Results showed a reasonable skill for daily maximum forecasts of ozone with an averaged root mean square (RMS) error of about 16 ppb and 0.8 of correlation, which is in agreement with the ozone forecast model intercomparison experiment described in [32].

2.2 Domain configuration

In the present study, the model turns at a regional scale over a domain covering Morocco. The domain spans from 20 to 37°N and 0 to 17°E and the grid size resolution is of 10 km which gives at all 100*200 grid points as shown in the (figure 1) and 8 height levels in the vertical resolution extending to about 500 hPa. The lowest level is of 50 m, the last one is about 5 km. The realized simulations spread out over the period going from January 1st till December 31st of 2008 [35].

The simulations spread on intervals of 5 successive days and every new period is initialized by the concentrations obtained in the end of the previous period. So the concentrations are continuous in the time.

The measured data are provided, both by the Surveillance and Risk Prevention Directorate (SRPD) from the Secretariat of Water and Environment from the Ministry of Energy, and the National Directorate of Meteorology (NDM).

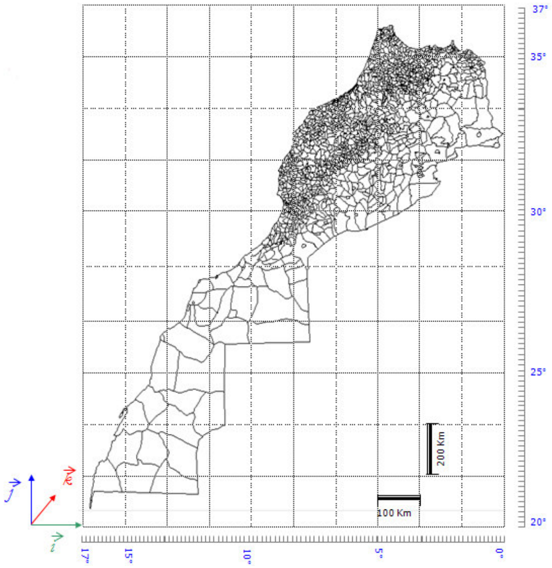


Figure 1 : showcase Grid point resolution of all Moroccan regions.

2.3 System architecture

The data warehouse that forms the core of our DSS incorporates information that comes from three main sources: The outputs of CHIMERE, captors and satellites. Therefore, the production data base will contain: Pollutant's concentrations, the wind's components and temperature, over the whole three-dimensional application domain.

We have chosen a combination between the Geographical Information Systems (GIS) and OLAP tools. This coupling process called SOLAP [15,30,31,34] will allow a thorough spatial analysis by providing thematic maps as results to spatio-temporal queries, additional operations like aggregation and drilling of data can also be performed. The following schema (figure 2) presents the main components of the developed system.

The first component includes the system data sources; it shows heterogeneous information obtained from satellite observations, sensors measures and numerous prediction simulations. The second component is the environmental data warehouse. The data are extracted from transactional databases and/or flat files, transformed to multidimensional architectures like star, unified flake or constellation, and then loaded within data marts, created during the transformation process. The third component relates to the SOLAP architecture. The data is geo-referenced under spatio-temporal cubes. The last component of the system includes the user's interface which is used for queries, reports generation and spatio-temporal data analysis. The interface offers different options based on the profile of the connected user like analysts, administrators or decision makers.

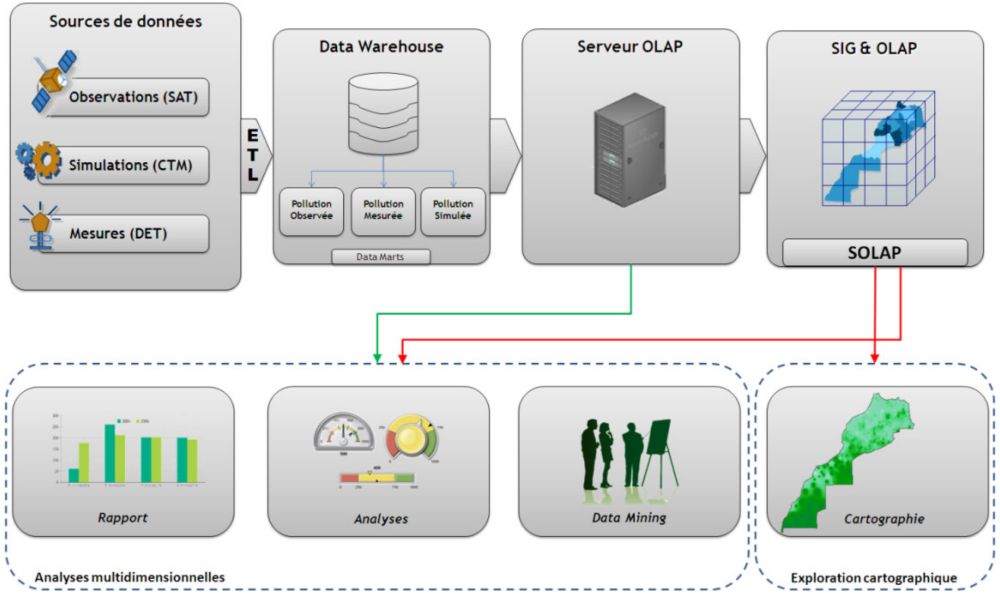


Figure 2 : The overall architecture of the air pollution decision support system.

3 Environmental database architecture

3.1 Description of the OLAP context

OLAP technologies allow the implementation of multidimensional databases. This approach is based on concepts like “dimensions” which represent the themes of interest and which often are organized hierarchically into different levels of granularity (ex: Pollutant, Time, Detector…) and “measures” which are numerical facts analyzed according to specific axis. For example, we can visualize the values of the NO2 pollutant per area and per date.

Each dimension is described by a set of “members” (ex: Semester1 of year 2008, Fez_Medina, Volatile Organic Components…), the members of one level (different areas) may be aggregated to constitute the members of the next higher level (the City for example). The “facts” represent the subject of interest that relates specific measure to specific dimensions on a particular aggregation level [21,22,23,29].

The common OLAP architecture usually comprises three components: the multidimensional structured database, the OLAP server and the OLAP client that accesses the database via the OLAP server. Depending on the technology (relational, multidimensional…) used to implement the OLAP database; it is possible to distinguish three OLAP approaches: relational OLAP (ROLAP), multidimensional OLAP (MOLAP) or hybrid OLAP (HOLAP), which is an optimized combination of the two previous approaches [35].

The OLAP schema (figure 3) depicts the overall multidimensional structure we have used in the context of this study. The Schema shows how data cubes are deployed. This multidimensional structure provides predicted concentrations of different pollutants per date and per location. For example, we may know the average of concentrations of PM10, which are a Particles Matter, in January 2008 in Sidi Brahim Area; an industrial area of Fez city.

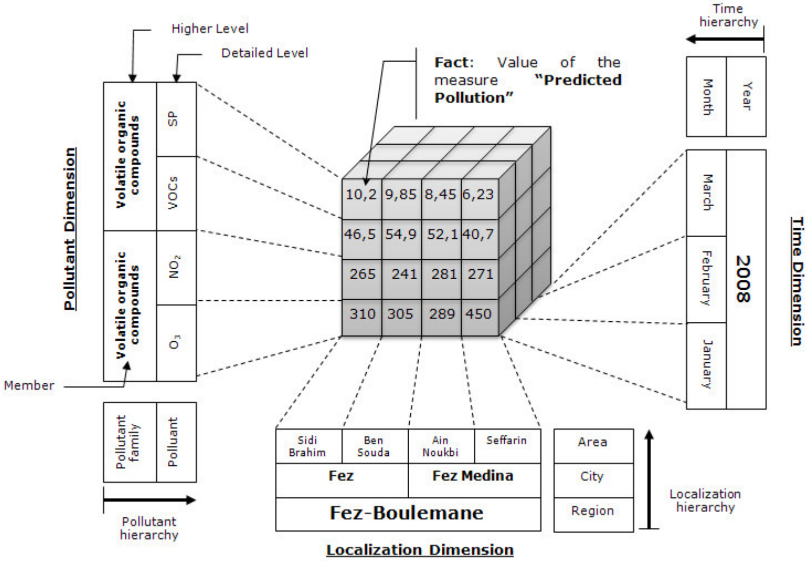


Figure 3 : Example of an air pollution multidimensional data cube

3.2 Spatio-temporal granularities and OLAP operators

The OLAP technology offers efficient operations for navigating among “hierarchies” and “levels” and also interactivity in term of data exploration. Popular operations that are applied to multidimensional cubes are Drill Down, Roll Up, Drill Across, Slice and Pivot. Drill Down operation allows visualizing a more detailed level within a dimension (drilling down the Fez-Boulmane Region to obtain values in the detailed level City). Roll Up corresponds to a group-by on one of the dimensions (ex. rolling up all semesters to have the value of O3 in a year). Drill Across consists to visualize another member or another dimension at the same level of detail. The Pivot operation consists in exchanging dimensions in order to modify the content of the analysis axis. The Slice operation consists in reducing the dimensionality of the data.

The figure (figure 4) shows dimensions hierarchies we are using for the pollution multidimensional model.

In the present study, the themes of analysis considered are: The concentrations of pollutants captured by pollution sensors, those observed by satellites and those simulated by the numerical atmospheric model.

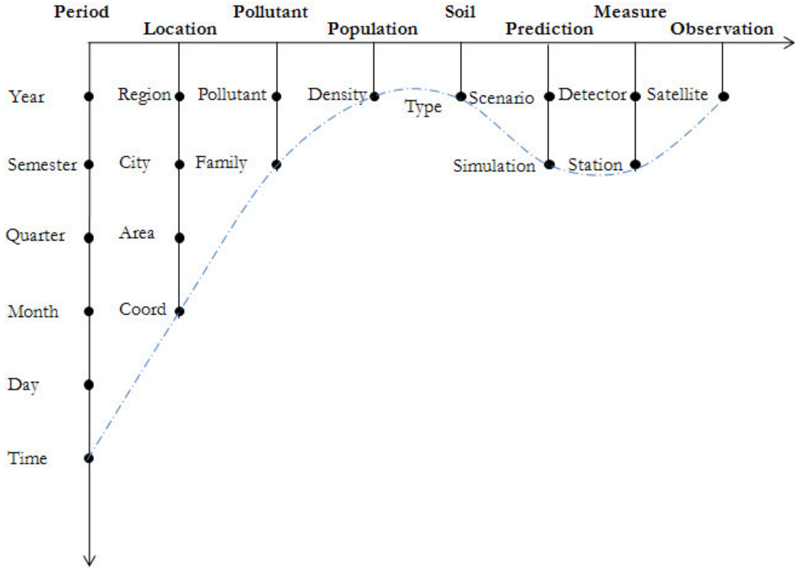


Figure 4 : Hierarchies and granularities of the dimensions of the multidimensional schema

4 Geographic information system and SOLAP specifications

4.1 GIS specifications

The Geographic Information Systems (SIG) are powerful tools to handle, query and navigate within a spatial databases [7,34].

A spatial data is often made of three parts: Geometrical (ex. Line, polygon…), descriptive (ex. City_Name…) and calculated (ex. Sidi_Brahim_Global_Surface or Perimeter). The geographic information is represented either in vector format or Raster format.

4.2 SOLAP specifications

The SOLAP solution corresponds to the coupling of an OLAP server and a GIS tool. [30,31,34] defines the SOLAP as “a category of software that allows rapid and easy navigation within spatial databases and that offers many levels of information granularity, many themes, many epochs and many display modes synchronized or not: Maps, tables and diagrams”.
SOLAP technology allows the dynamic creation of charts which are based on multidimensional data stored in the spatial cubes and offers the possibility of spatiotemporal exploration and analysis.

4.3 Data loading and interfaces

Data are extracted from hyper-cubes and geo-referenced on layer maps. The user uses different SOLAP operations like Rolling up or drilling down into Areas for the localization dimension for example. Thematic charts are obtained as a result of the user’s actions. They can also determine different aggregated values like Average, Max, Min…so on. (figure 5) shows an instance of the extraction and geo-referencing processes.

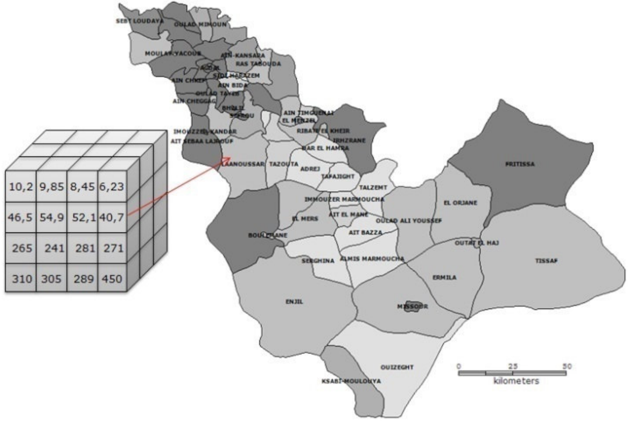


Figure 5 : Extracting and geo-referencing multidimensional data.

4.4 Data exploration and analysis

In terms of functionalities, our system offers, the system aggregated information of the air pollution according to various axes of analysis. We give bellow some examples of how the cubes were deployed [20].

4.5 Tabular format

Table 1 shows the average of pollutant values obtained by applying the aggregate operations; Sum and Count in a day per captor, then the results were rolled-up to the month level and we applyed a second roll-up operator to the semester level.

Table 1 : the average of captured values of different pollutants per sensor in the first Semester of 2008

Time	Detector	Pollutant	AVG
Sem1/2008	DET_SidiBrahim	O_3	14.51
		PM10	120,33
		NO_2	301.5
	DET_BenSouda	O_3	22,08
		PM10	128,39
		NO_2	297.66
	DET_AinNoukbi	O_3	38,96
		PM10	80,44
		NO_2	305.33

Table 2 shows the values of the pollutants for the period between 20/09/2008 to 22/09/2008, with roll up to the City level.

Table 2: Predicted values of different pollutants per station from 20-22 September, 2008 in Fez city

City	Time	Pollutant	Station	Value
Fez	20/09/2008	NO_2	Sidibrahim	400
			Bensouda	360
			Doukarat	157
	21/09/2008	O_3	Sidibrahim	180
			Bensouda	300
			Doukarat	150
	22/09/2008	PM10	Sidibrahim	121
			Bensouda	88
			Doukarat	96

4.6 Diagrams

The figure (figure 6) shows the time series of the daily mean of O3 and NO2 concentrations respectively, simulated by CHIMERE model and observed at Aljahid station (industrial), located in Casablanca area. Even if some discrepancies occurred, the time variations are relatively well captured by the model[35].

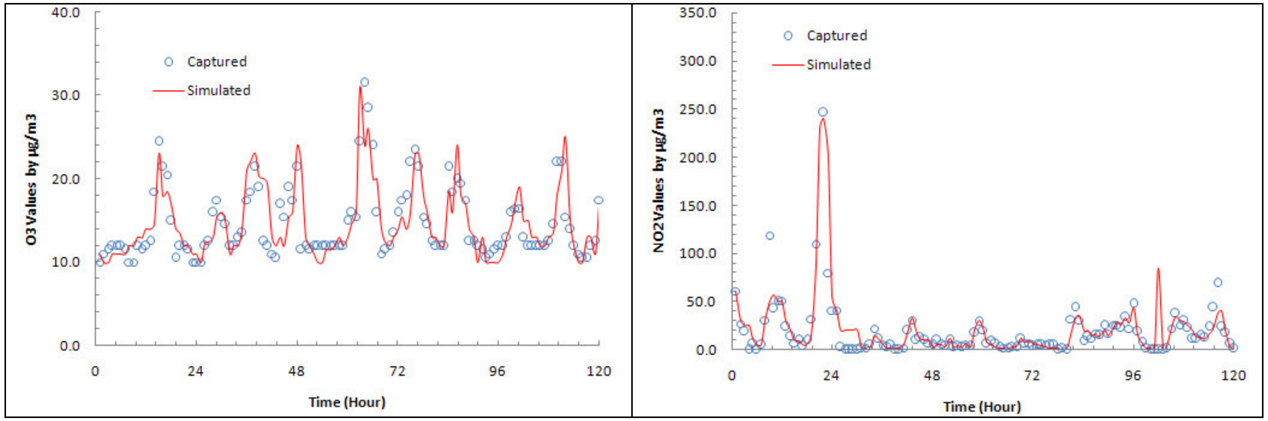
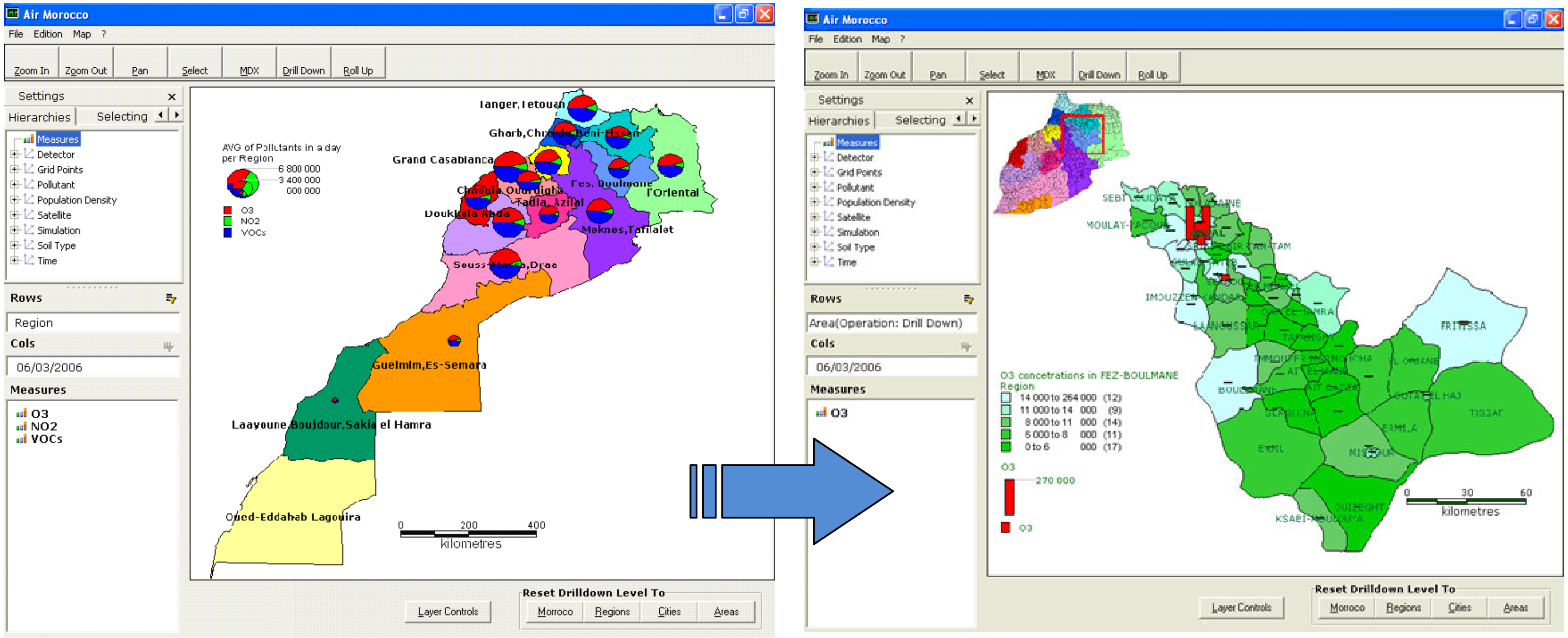


Figure 6 : Time series of O3 and NO2 concentrations (µg/m3) simulated and measured in Casablanca (Aljahid Station) from 1 to 5 August 2008.

4.7 Thematic maps

The SOLAP architecture above which our application is based allows a cartographic representation of the selected indicators within the application domain. Different levels of thematic spatiotemporal maps can be generated as a result to user's actions like drilling down or rolling up spatial dimension. The user can also analyze diverse aggregated values such as AVG, Min… The figure (figure 7) shows an example of the extraction and geo-coding process to calculate the average of some pollutants concentrations in FEZ-Boulmane Region. Spatial OLAP operations are also illustrated.



Left window:

Rows
Region

Cols
06/03/2006

Measures
O3
NO2
VOCs

Right window:

Rows
Area(Operation: Drill Down)

Cols
06/03/2006

Measures
O3

Figure 7 : Drilling down the spatial dimension and thematic maps generation

5 CONCLUSION

This study was completed within the framework of a SOLAP study for the monitoring of the air pollution in all Moroccan regions and cities. The integration of the warehouse was done in a MOLAP environment and the tools of analysis and interrogation of data were established with the aim to construct a decision-making tool that will help to act properly and in time. The decision maker can also use it as a support when defining short-and long-term actions to control the air pollution phenomena in the county. We have also developed the cartographic component of our application that makes possible the use of geographic and thematic tools of interrogation and analysis.

We are currently developing a more comprehensive decision support system for air pollution monitoring over the country of Morocco, the latest has data processing layer and an end user presentation layer. Specification and navigation options of the system are the subject of our future study.

6 REFERENCES

[1] A DE MEIJ, P. THUNIS, B. BESSAGNET AND C. CUVELIER. “*The sensitivity of the CHIMERE model to emissions reduction scenarios on air quality in Northern Italy*”. Atmospheric Environment, vol. 43, 2009, pp. 1897–1907, doi:10.1016/j.atmosenv.2008.12.036.

[2] A. HODZIC, H. CHEPFER, , R. VAUTARD, , P. CHAZETTE, , M. BEEKMANN, B. BESSAGNET, B. CHATENET, J. CUESTA, , P. DROBINSKI, , P. GOLOUB, , M. HAEFFELIN, , Y. MORILLE, "*Comparison of aerosol chemistry-transport model simulations with lidar and sun-photometer observations at a site near Paris*", J. Geophys. Res., 109, D23201, 2004, doi:10.1029/2004JD004735.

[3] A. HODZIC, R. VAUTARD, H. CHEPFER, P. GOLOUB, L. MENUT, P. CHAZETTE, J. L. DEUZE, A. APITULEY, P. COUVERT, “*Evolution of aerosol optical thickness over Europe during the August 2003 heat wave as seen from CHIMERE model simulations and POLDER data*”. Atmospheric Chemistry and Physics. Vol. 6, 2006, pp.1853–1864.

[4] A. MONTEIRO, A.I. MIRANDA, C. BORREGO, R. VAUTARD, J. FERREIRA, A.T. PEREZ, *“Long-term assessment of particulate matter using CHIMERE model”*. Atmospheric Environment, Vol. 41, 2007, pp. 7726–7738. doi:10.1016/j.atmosenv.2007.06.008.

[5] A. MONTEIRO, R. VAUTARD, C. BORREGO AND A.I. MIRANDA, *“Long-term simulations of photo oxidant pollution over Portugal using the CHIMERE model”*, Atmospheric Environment, Vol. 39, June 2005, pp. 3089-3101, doi:10.1016/j.atmosenv.2005.01.045

[6] B. BESSAGNET, A. HODZIC, R. VAUTARD, M. BEEKMANN,S. CHEINET, C. HONORE,C. LIOUSSE AND L. ROUIL, *“Aerosol modeling with CHIMERE - Preliminary evaluation at the continental scale”,* Atmospheric Environment, Vol 38, Issue 18, June 2004, pp 2803-2817, doi:10.1016/j.atmosenv.2004.02.034.

[7] C. FRANKLIN, *“An Introduction to Geographic Information Systems: Linking maps to databases”*. Database, Vol. 15, No. 2, 1992, pp. 12-21.

[8] D. ZHONGYUAN, R. WANG, L. FANG, C. DEROGNAT, G. GUERINOT, M. BEEKMANN, B. DAMEZ-FONTAINE AND A. ALBERGEL, *“Sulphur chemistry and acid rain over china. How to compute the contribution of each province?”*. 9th Int Conf. on armonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, 2007.pp. 194-198.

[9] E. CHAXEL, AND J.P. *“Cholleta, Ozone production from Grenoble city during the August 2003 heat wave”*, Atmospheric Environment, Volume 43,Oct 2009, pp. 4784-4792. doi:10.1016/j.atmosenv.2008.10.054.

[10] H. SCHMIDT, C. DEROGNAT, R. VAUTARD, M. BEEKMANN. *“A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in Western Europe”*. Atmospheric Environment, vol. 35, Dec 2001, pp.6277–6297. doi:10.1016/S1352-2310(01)00451-4.

[11] I. PISON, L. MENUT, *“Quantification of the impact of aircraft traffic emissions on tropospheric ozone over Paris area”*. Atmospheric Environment, vol. 38, 2004, pp. 971–983. doi:10.1016/j.atmosenv.2003.10.056

[12] J. CURRIE, , M. NEIDELL AND J.F. SCHMIEDER. *“Air pollution and infant health: Lessons from New Jersey”*. Journal of Health Economics, vol. 28, May 2009, pp. 688–703. doi:10.1016/j.jhealeco.2009.02.001.

[13] J. JUST, L. NIKASINOVIC, Y. LAOUDI AND A. GRIMFELD. *“Air pollution and asthma in children”*. Revue française d’allergologie et d’immunologie Clinique, Volume 47, April 2007, pp. 207-213, doi:10.1016/j.allerg.2007.01.012.

[14] K.E. PRECY AND M. FERETTI, *“Air pollution and forest health: toward new monitoring concepts”*. Environmental Pollution, vol 130, july 2004, pp. 113–126, doi:10.1016/j.envpol.2003.10.034.

[15] M. MIQUEL, M., Y. BEDARD AND A. BRISEBOIS, *“Conception d'entrepôts de données géospatiales à partir de sources hétérogènes, exemple d'application en foresterie”*, Ingénierie des Systèmes d'information, Vol. 7, 2002, pp. 89-111

[16] M. RICHARDS, M. GHANEM, M. OSMOND, Y. GUO AND J. HASSARD, *“Grid based analysis of air pollution data”*. Ecological Modelling, Volume 194, March 2006, pp. 274-286, doi:10.1016/j.ecolmodel.2005.10.042.

[17] M.G. VIVANCO, I. PALOMINOA, R. VAUTARD, B. BESSAGNET, F. MARTIN, L. MENUT AND S. JIMENEZE, *“Multi-year assessment of photochemical air quality simulation over Spain”*. Environmental Modelling & Software vol. 24, 2009, pp. 63–73. doi:10.1016/j.envsoft.2008.05.004.

[18] NCAR, National Center for Atmospheric Research numerical model home page, MM5 http://www.mmm.ucar.edu/mm5/mm5-home.html

[19] P. ISABELLE, M. LAURENT. *“Quantification of the impact of aircraft traffic emissions on tropospheric ozone over Paris area”*. Atmospheric Environment, Volume 38, March 2004, pp. 971–983, doi:10.1016/j.atmosenv.2003.10.056.

[20] P. JANUS, *“ProClarity Analytics Server”,* Pro PerformancePoint Server 2007, Building Business Intelligence Solutions, Springer-Verlag, 2008, ISBN 978-1-59059-961-7, pp. 211-254.

[21] P. VASSILIADIS, “*Modeling multidimensional databases, cubes and cube operations”*. Proc of the 10th International Conference on Statistical and Scientific Database Management (SSDBM), Capri, Italy, 1998, pp. 53-62. ISBN:0-8186-8575-1.

[22] R. AGRAWAL, A. GUPTA AND S. SARAWAGI. *“Modeling Multidimensional Databases”*. Proc. of the 13th Int'l Conference on Data Engineering, Birmingham, U.K., April 1997 pp. 232-243, DOI 10.1109/ICDE.1997.581777.

[23] R. KIMBALL, M. ROSS, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, 2nd Edition. John Wiley, ISBN: 978-0-471-20024-6,464 pages.

[24] R. LAURENCE. *“PREV’AIR : un système opérationnel de prévision et de cartographie de la qualité de l’air en France et en Europe”*. http://www.prevair.org/fr/index.php

[25] R. VAUTARD, B. BESSAGNET, M. CHIN AND L. MENUT, *“On the contribution of natural Aeolian sources to particulate matter concentrations in Europe: Testing hypotheses with a modeling approach”*. Atmospheric Environment, vol.39,Jun 2005.pp. 3291-3303, doi:10.1016/j.atmosenv.2005.01.051

[26] R. VAUTARD, M. BEEKMANN, J. ROUX AND D. GOMBERT, *“Validation of a deterministic forecasting system for the ozone concentrations over the Paris area”*, Atmospheric Environment, Vol. 35, 2000, pp. 2449–2461, doi:10.1016/S1352-2310(00)00466-0.

[27] R.VAUTARD, *“The Chimere chemistry-transport model”*, 2007, http://euler.lmd.polytechnique.fr/chimere/

[28] S. BIMONTE, A. TCHOUNIKINE AND M. MIQUEL. *“GeoCube, a Multidimensional Model and Navigation Operators Handling Complex Measures, Application in Spatial OLAP”*. Advances in Information Systems, Databases and Datawarehouses, Springer Berlin / Heidelberg, 2006, ISSN 0302-9743 (Print) 1611-3349, pp100-109.

[29] S. CHAUDHURI AND U. DAYAL, *“An overview of data warehousing and OLAP technology. SIGMOD”*. 1997. Rec., 26: 65-74

[30] S. RIVEST ET AL, *“SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data”*. ISPRS Journal of Photogrammetry & Remote Sensing, vol. 60, 2005 pp.17–33.

[31] S. RIVEST ET AL, *“Toward better support for spatial decision making: Defining the characteristics of spatial on-line analytical processing”*. Geomatica, vol. 55, 2003, pp. 539-555.

[32] S. TILMES, J. BRANDT, F. FLATOY, R. BERGSTRÖM, J. FLEMMING, J. LANGNER, J.H. CHRISTENSEN, L.M. FROHN, O. HOV, I. JACOBSEN, E. REIMER, R. STERN AND J. ZIMMERMANN, *“Comparison of five Eulerian air pollution forecasting systems for the summer of 1999 using the German ozone monitoring data”* . Journal of Atmospheric Chemistry, Springer Netherlands 2002, pp. 42:91–121.ISBN: 0167-7764.

[33] T.O. AHMED AND M. MIQUEL, *“Multidimensional Structures Dedicated to Continuous Spatiotemporal Phenomena”*. Lecture Notes in Computer Science, Database: Enterprise, Skills and Innovation, Springer Berlin / Heidelberg, Jun 2005, , ISSN 0302-9743 (Print) 1611-3349, pp. 29-40.

[34] Y. BEDARD ET AL., *"Merging hypermedia GIS with spatial on-line analytical processing: Towards Hypermedia SOLAP"*, Lecture Notes in Geoinformation and Cartography, Geographic Hypermedia, E. Stefanakis et al., Springer-Verlag, 2006, ISBN: 978-3-540-34237-3, 167-187.

[35] Y. OUBENAALLA, O. EL BEQQALI, *"Long term simulations of atmospheric poullution over morocco using Chimere / WRF models"*. (accepted)