Solutions for the Data Level’s Representation in a Decision Support System in Wind Power Plants

ADELA BÂRA, ANDA VELICANU, IULIANA BOTHA, SIMONA VASILICA OPREA
Economic Informatics Department
Academy of Economic Studies
Transelectrica National Power Grid Company
Bucharest
ROMANIA
{Simona.Oprea}@transelectrica.ro

Abstract: - This paper presents the solutions for the data level’s representation in the architecture of a decision support system that can be used in the National Power Grid Companies for integrating and analysing the energy produced in the wind power plants. We considered some informatics technologies, like XML and object-relational databases for storing data and also the spatial data for representing the locations and the coordinates of the wind power plants installed in the country. In the general context it is presented the architecture of the DSS prototype, which can be applied in the uncertain and unpredictable environments, like the production and the prediction of the wind energy.

Key-Words: - decision support system; spatial analysis, geographic information system, spatial data, object-relational databases, XML data

1 Introduction

Geographic Information Systems (GIS) allow creating maps, integrating information, visualizing scenarios, solving complicated problems and developing effective solutions in a new, easy to use way. A GIS is a system used for modeling information, processes and structures that reflect the real world, including past events, in order to understand, analyze and manage resources and facilities.

To achieve a performing GIS must assure that its functionalities are independently developed and then composed as a whole within a service-oriented architecture. Web services, delivered or not as electronic services, give the possibility to implement a system in which components are weakly-coupled, composite and independent.

An example of using GIS is developing an application for wind parks. Their particularities, like produced energy, noise, altitude where are located, and the external factors that affect the functionality, like wind speed, conflict areas, weather, must be taken into consideration. Thus, such a system will be complex and will need for processing the information spatial analysis as well as data mining tools.

Considering the fact that all business decisions include components of time and space, we can say that GIS represent good solutions in order to improve decision making.

Spatial objects are a combination of non-spatial attributes, which describe characteristics of each item and geo-spatial components, which describe the relative location and geometrical shape of each object.

Nowadays, most of the data used by organizations are stored as object data because this form has the greatest resemblance to the real world and can be easily stored in relational databases with object-oriented extensions.

Organizing spatial data into spatial databases can affect the response speed of GIS systems. An important factor that may have influence on GIS usage refers to the normalization process, which tends to spread data in multiple tables. The result consists in inefficient spatial queries and negative influence for the performance of GIS systems.

The most effective way to store spatial data is given by object-relational databases (ORDB). These are a hybrid type of databases, used to combine the main advantages of relational databases and object-oriented technology [4].
Management systems databases (DBMSs) have implemented support for spatial data type, directly or by extension.

In the following sections, we’ll present our solution for developing a decision support system in the national power grid company that will store and represent wind park locations in a GIS, but for advanced analyses will use the object oriented methods.

2 Problem Formulation
2.1 System requirements and constraints
The high level management, as it is in the National Power Grid companies, needs a real time analyses and forecast of the energy produced by the wind power farms. For better manage, transform, process and analyze this information, it is needed a decision support system (DSS). Many nationally and internationally companies developed various types of frameworks, architectures, solutions and systems to provide economic assistance to decision making processes and production environments with a relatively high degree of certainty.

For environments with low predictability, such as the energy system, that integrates wind resources which depend on natural factors, the traditional forecast applications built in the Enterprise Resource Planning (ERP) solutions is inefficient, being known cases of failure to implement decision support systems [3].

First problem is that data sources provided for the decision-making process must integrate three types of information: external information gathered from the wind parks’ production, information within the organization concerning the energy consumption and a combined forecast system both internally and externally. These sources need an efficient integrating framework.

The second problem is the accuracy of the existing forecast systems used to determine the wind park production. There are several studies and software tool developed nationally (studies conducted by research institutes SC ISPE SA, Tractebel Engineering SA, ICEMENERG) and European studies (EWIS study conducted in EU countries, DENNA, made in Germany). However, these studies have shown that the forecast errors propagated in the national energy system can lead to periods of maximum consumption to drop power products in certain regions. Thus, a new type of model based on data mining techniques has to be build that will be able to consider more efficiently the natural factors and also the specific characteristics of the wind power turbines and the environmental constraints.

The third problem is to gather, extract, transform and store the data into a central database from where the business intelligence application can present useful information to the high level management. Also, a data warehouse must be build to store historical data, so we need to implement a solution that will optimize the performance of queries and analyses.

2.2 System’s architecture
The system’s architecture has to be developed to solve the three major problems described in the previous section.

Thus, for wind park locations and production and also for energy consumption areas the data collected will be stored as spatial data in order to representing them on interactive maps monitoring the activities, and other data like multimedia data and LOB will use specific operations. The data sources will be integrated into a central database. The data will be received in eXtensible Markup Language (XML) format and will be stored in object-relational databases (ORDB), in order to realize further analysis, necessary for decision making. The section 3 and 4 will present detailed the solutions adopted for this level of the architecture.

For prediction and forecast component, based on our previous research as described in [5], [6] and [7], we’ll build a model based on data mining algorithms with 95% accuracy. The system will use data that can influence wind parks’ performances, like meteorological data or data resulted from national energy production regular monitoring. The software component will store predicted data to the central database in a relational form. The system will use a central data warehouse used by business intelligence tools to present and analyze wind park production and energy consumption, both for historical and forecast periods. An extract, transform and load process need to be developed. In the first phase it will have two components: one for integrate the data sources into the database and the other one for loading the data from the database into the data warehouse.

In the figure 1 is presented the system’s architecture.
The components of the architecture are structured on the four traditional levels of a decision support system, namely: 1) data level; 2) system of analytical, mathematical and statistical models; 3) interface; 4) component to ensure communication. The following sections describe some of the specific components of the data level: spatial analysis with GIS and data representation in an object-relational method.

3 Spatial Analysis with GIS

The main purpose of Geographic Information Systems (GIS) is to perform Spatial Analysis (SA) for geographically referenced data. Simply, we can say that GIS is a digital environment for spatial analysis. The information environment, in which the spatial analysis is made, is the map composed of layers and data. For spatial analysis, the data must be geographically referenced and other GIS related functions must be used: acquisition, editing, validation, storage, primary processing, visualization, posting. Spatial modeling is seen, in this case, as a special spatial analysis, which has spatial scenarios results. Spatial analysis must simultaneously fulfill the following goals: review and interpretation of data, getting extra seemingly hidden information, quantitative and qualitative assessment of the entities, processes and phenomena in the analyzed space, providing practical support for decision making situations. Making spatial analysis involves the use of analytical procedures, combined with: database management, statistical and geostatistical data analysis, image processing and computer mapping elements. Particularly important to the use of GIS and AS there are the knowledge of computer graphics and the specific of information technology in the GIS. Thus, geographic information system is seen as a spatial information science rather than as a technology [1].

Collecting data has long been one of the most expensive procedures in GIS. Now, this data gathering process has become faster, easier and cheaper thanks to the development of data acquisition technologies such as Global Positioning System (GPS) on aerial photos, automatic or semi-automatic scanners and barcode readers. Because of this, geographic information systems have evolved from technologies that use less data and computing
power, to technologies that use data-rich environments and high computing power.

According to [2] the basic concepts of spatial analysis are spatial dependence, spatial autocorrelation, statistical inference for spatial data, stationarity and isotropy.

Spatial dependence is a key concept for understanding and analysis of spatial phenomena. This means that certain properties of spatial objects may depend on the same or other properties of spatial objects. In [2] is cited the shadow that a tree can produce around it and how it can affect the growth of other trees nearby. Another example is the discovery of oil spots in a lake and the probability to also discover oil around that stain.

Spatial autocorrelation is a mathematical expression of spatial dependencies. The term comes from the statistical correlation, which is used to measure the dependence between two variables. Autocorrelation implies that the variables compared are the same, but the measurements are made in different places. Spatial autocorrelation verifies how the spatial dependence varies by comparing a sample and its neighbors. In [2] is presented three such formulas:

\[
\Gamma(d) = \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}(d) \xi_{ij}
\]  (1)

Where: \( \Gamma(d) \) is an expression of the link between two random variables as the product of two matrices. Posing a distance \( d \), the \( W_{ij} \) matrix measures spatial continuity between \( Z_i \) and \( Z_j \) random variables. This link indicates whether or not the two variables are at a distance smaller than \( d \). \( \xi_{ij} \) is a matrix that measures the correlation of these two random variables and it can be their product, as expressed in (2).

\[
I = \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}(z_i - \bar{z})(z_j - \bar{z})
\]

\[
\sum_{i=1}^{n} (z_i - \bar{z})^2
\]  (2)

Where: \( W_{ij} \) is 1 if the geographical areas associated with \( Z_i \) and \( Z_j \) spatially touch, and is 0 otherwise.

Another example is the variogram, where values are formed as squared differences:

\[
\hat{\gamma}(d) = \frac{1}{2N(d)} \sum_{i=1}^{N(d)} [z(x_i) - z(x_i + d)]^2
\]  (3)

Where: \( N(d) \) is the number of variables separated by the \( d \) distance.

An important consequence of spatial dependencies is that statistical inference on such data is not as effective as if the independent variables have the same size. In general, this is reflected in higher variances for estimators, lower levels of significance in testing hypothesis and less beneficial adjustment of estimated models. In other words, such a pattern of assumptions usually does not work as properly as when the data compared is the same size and appears independent.

Key statistical concepts that define a spatial data structure are based on primary and secondary order effects. Therefore, the primary order effect is the expected value, which is the average of process in space. Secondary order effect is the covariance between \( S_i \) and \( S_j \) surfaces. In this context, a process is considered stationary, if it’s stationary and if the covariance depends only on the distance between the points and not on their direction.

A Z stochastic process is stationary of order 2, if the expected value of \( Z(u) \) is constant in all areas studied, meaning it does not depend on its position.

\[
E\{Z(u)\} = m
\]  (4)

Also, the spatial covariance’s structure depends only on the relative vector of points difference \( h = u - u' \)

\[
C(h) = E\{Z(u) \cdot Z(u+h)\} - E\{Z(u)\} E\{Z(u+h)\}
\]  (5)

Standard GIS systems are very slow in the process of analyzing large volumes of data. It is therefore necessary to explore new methods of analysis that take into account both non-spatial and spatial objects stored in large databases.

### 4 Object-relational representation

The fast development of the IT&C domain was accompanied constantly by significant changes in economic informatics. The current informatic technologies have created the need for the use and storage of complex object types within databases, while keeping the power of the relational structured query language (SQL). Therefore, the new
generation of database systems supports a unified relational and object-oriented data model. Thus, the relational model is extended with the object-oriented concepts such as scalability, complex data types (large objects, multimedia data, XML data, spatial data, user defined object types, etc.) and the fundamental object characteristics like encapsulation, inheritance and polymorphism [10].

An advantage of using applications with object-relational databases, despite the relational ones, leads in the use of a unique data model, recognized in both database and programming language (which is most certainly an object-oriented programming language) [9].

![Diagram showing object-relational databases](image)

**Figure 2.** – Using a unique data model in object-relational databases (Source: authors)

As shown in the figure 2, object-relational databases can by-pass the necessity of using mapping techniques between relational database and application programs. However, it is needed for specific techniques in order to map the objects stored in relational tables into instances of classes used in application programs.

The proposed DSS will integrate data from various systems. There are source modules from which are received XML documents with wind park’s characteristics and destination modules represented by object-relational databases. XML documents are processed and the data are stored into database. The advantage of using an object-relational database is that we can get the benefits of both relational and object-oriented technologies, while the disadvantage translates into lower performance due to XML data mapping to the relational data, which can produce a database schema with many relations [11].

When used in object-relational databases, XML data must be mapped into relations. In order to transfer the data between XML documents and object-relational structures are used specific mapping methods. The study [8] makes a basic classification of these mapping methods, as follows:

- **Generic methods** – that are not using any schema of stored XML documents;
- **Schema-driven methods** - that are based on existing schema of stored XML documents;
- **User-defined methods** - that are based on user-defined mapping.

In order to integrate the XML data about wind parks into the database and realize the mapping, we will use the schema-driven methods. The mapping
process will consist in three phases. We will start with the conceptual model, then we will create the XML schema and finally we will map the specific elements on object-relational database schema.

5 Conclusions and future work
The paper presents the necessities for having real time analyses and forecast of the energy produced by the wind power farms. The high level management of the National Power Grid companies needs solutions like decision support systems (DSS) in order to better manage, transform, process and analyze the information.

The integration of the specific data (like multimedia, XML and spatial data) into object-relational databases is treated as a necessity for today’s enterprises that maintain informatics systems in the unpredictable environments.

In the future, our research will continue with the development and implementation of the conceptual model proposed in this paper. For this purpose, will be used specific technologies for each level of the model architecture [11]. Thus, the data level will use solutions for organizing and integrating data, the models level will use solutions of multidimensional analysis, forecasting models, simulation and extrapolation, and for the interface level will be used solutions for data analysis and dynamic presentation of data.

Acknowledgments
This paper presents some results of the research project PN II, TE Program, Code 332: “Informatics Solutions for decision making support in the uncertain and unpredictable environments in order to integrate them within a grid network”, financed within the framework of People research program.

This paper is also a result of the project POSDRU/6/1.5/S/11 „Doctoral Program and PhD Students in the education research and innovation triangle”. This project is co-funded by European Social Fund through The Sectorial Operational Program for Human Resources Development 2007-2013, coordinated by The Bucharest Academy of Economic Studies.

References: