International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:3, No:6, 2009

CSOLAP (Continuous Spatial On-Line Analytical Processing)

Taher Omran Ahmed, Abdullatif Mihdi Buras

Abstract**—**Decision support systems are usually based on multidimensional structures which use the concept of hypercube. Dimensions are the axes on which facts are analyzed and form a space where a fact is located by a set of coordinates at the intersections of members of dimensions. Conventional multidimensional structures deal with discrete facts linked to discrete dimensions. However, when dealing with natural continuous phenomena the discrete representation is not adequate. There is a need to integrate spatiotemporal continuity within multidimensional structures to enable analysis and exploration of continuous field data. Research issues that lead to the integration of spatiotemporal continuity in multidimensional structures are numerous. In this paper, we discuss research issues related to the integration of continuity in multidimensional structures, present briefly a multidimensional model for continuous field data. We also define new aggregation operations. The model and the associated operations and measures are validated by a prototype.

Keywords**—**Continuous Data, Data warehousing, Decision Support, SOLAP

I. Introduction

THE last few years have seen an explosive growth of the size of data produced by the different kinds of sensors and stored in different operational databases. Exploiting these huge volumes of data in a decisional context is the main function of data warehouses (DW). Data warehouse are used in analysis purposes that involve examining data and possibly identifying relationships that may exist between different elements [27]. However information collected in a decisional environment usually involve spatial positions. Therefore decision making must take in consideration this type of data. In operational databases, geographic information systems (GIS) have shown their ability in geographic data management. The effectiveness of GIS comes mainly from of linking their capacity different information in a spatial context and draw conclusions from the different relations that exist between different phenomena. However GIS are oriented towards spatial data management not towards effective analysis. Even though, they are considered as decision support systems, their contribution strategic levels of decision making is quiet limited [6].

GIS cannot provide decisional information mainly because they are not adequately designed for decision support. Providing a dependable spatial decision support system consists of coupling GIS and OLAP so that multidimensionality is provided by OLAP and spatial data are manipulated by GIS, [16]. In this paper we present our research towards integrating spatiotemporal continuity in decision support systems. In the second section we present the research objectives, issues and motivations. Section 3 contains a state of the art study of multidimensional structures and field based data. In section 4 we go over solutions proposed to the research issues. In the fifth section we briefly present a multidimensional model for field based data. In section 6, we present a prototype of a continuous spatial data warehouse. The conclusions and future work are presented in section 7.

II. Research Issues and Motivations

Multidimensional modeling proved to be a good solution for decision support systems for several reasons. First of all, with the use of pre-aggregation response time to lengthy queries that may involve millions of lines is significantly reduced. Second, users of decision support systems are not required to have databases systems knowledge. Third, the multidimensional representation is closer to the way of thinking of analysts where facts are analyzed with respect to various factors. Multidimensional systems provide users with the freedom to explore data and produce types of reports without restricting them to a pre-set format.

In conventional multidimensional structures all data involved are discrete. Dimensions are divided in discrete hierarchical levels with each level having finite set of discrete members. A discrete fact may be found at the intersection of members where members serve as a set of coordinates in a multidimensional space. However for natural phenomena the spatial and temporal dimensions are not discrete and must be treated as continuous. Phenomena take place everywhere and constantly without disruption. But of course they cannot be measured continuously at all points in space nor can these measurements be stored in databases due to several factors like the discrete nature of computers. Only samples are measured and stored which leads to discrete representation of the continuous phenomenon. This representation is reflected in discrete spatial and temporal dimensions in spatial data warehouse *(SDW)* and spatial OLAP (*SOLAP*). *SDW* and *SOLAP* have performed very well in the domain of spatial

Taher Omran Ahmed is a lecturer, head of Computers Science department and dean of Faculty of Science – Alzintan, Aljabal Algharbi University, Libya. (phone: +218 91 3673343; e-mail: fenneer@yahoo.com,).

Abdullatif Mihdi Buras is a lecturer, head of Mathematics department and dean of Faculty of Science – Gharian, Aljabal Algharbi University, Libya. (phone: +218 91 3758042 e-mail : buras@yahoo.com).

decision support systems. Among the domains that benefited or could benefit from *SDW* and *SOLAP* is environmental health, forestry, transport….etc.

It should be noted that all the above mentioned domains are discrete. Therefore since a distinction is made in *GIS* on the detailed data level with respect to the continuous and discrete representation, the same distinction should be made on the aggregated level. We genuinely believe that the missing piece of the puzzle is enabling decision support systems for continuous field data. Our main objective is integrating continuity within multidimensional structures in order to perform analysis and exploration on both continuous data and discrete data.

Despite the huge amounts of continuous data collected by all types of sensors there has not been much work done on data warehouse modeling for this type of data. Most of the work on *SDW* and *SOLAP* deal with discrete spatial data. Even when continuous field data are dealt with in a decisional context they are treated as discrete and they lack their main characteristic which is spatiotemporal continuity, [13], [19], [29]. The motivations for integrating spatiotemporal continuity within multidimensional structures point out the inadequacy of overlooking the characteristics of field based data [1]:

- Recovering Hidden Information : The discrete representation of continuous data in databases leads to discrete dimensions when data are modeled in multidimensional structures. This means that some information is lost or say at least is *hidden*. Hidden information can be very useful for analysis and exploration. Thus, having a continuous representation provides the means of recovering lost or hidden information by giving an estimation of unmeasured data.

- Analysis at Detailed Levels of Hierarchy: Decision support systems usually deal with aggregated data and completely or partly overlook detailed data but there are numerous cases where the need of low granularity analysis arises. In the cases of disaster management it is crucial that the behavior of the phenomenon at low levels of detail is known.

- Continuous Analysis and continuous representation : Due to the discrete representation, natural phenomena are analyzed as discrete spatial objects and as snapshots of data values over different periods of time. It would be more realistic for environmentalists to analyze a natural phenomenon as it occurs in real life (i.e. *naturally*) where a phenomenon evolves in space and in time rather than as a collection of discrete pieces.

Integrating spatiotemporal continuity in multidimensional structures raises several research questions and poses numerous challenges [1]:

- **Continuous multidimensional model**. Existing multidimensional models cannot support spatiotemporal continuity. All existing models deal with discrete dimensions related to discrete facts or measures. These models show great shortcomings when dealing with field-based data. To the best of our knowledge, no work has been done on modeling multidimensional structures for continuous data. The only work that discusses the use of continuous dimensions is found in [25] who focuses on using the known density of data to calculate aggregate queries without accessing the data. The representation reduces the storage requirements, but does not present the continuity in the same way we do since we estimate non-existing measures based on existing sample data values.

- **Range of continuity**. One of the most debatable issues we encountered was defining the range of continuity with respect to hierarchies of dimensions. The question was "where does the continuity begin and end *hierarchically*?" This question is clear in the temporal dimension since the hierarchical nature of time is more evident. In a simple form, time is ordered hierarchically as year, month, day, hour, minute, second, microsecond… etc. Does continuity imply producing new finer levels or should it only produce data for any instant in time? And can higher levels be continuous?

- **Determination of spatial and temporal interpolation methods**. Different spatial and temporal interpolation methods should be applied based on the modeled phenomena. The spatiotemporal interpolation methods differ in their assumptions, methodologies, complexity, and deterministic or stochastic nature. Each method has its merits and is applicable according to temporal length scale, spatial length scale, stationarity, and variability of the field under consideration [26].

- **Storage and optimization**. One of the objectives of *OLAP* is to guarantee fast response to users' queries. This is achieved by the use of pre-aggregation and therefore eliminating the overhead of calculating *SQL* aggregations during run time. This objective must be kept in mind for continuous multidimensional structures. The complexity of the problem is augmented because of interpolation. Having to interpolate means that pre aggregation can not be performed for several or all operators.

- **Navigation and operations in the continuous hypercube**. Navigation in continuous structures is different from that in the conventional ones. The classic *OLAP* operators need to be extended and redefined to be applicable to the new structures. The introduction of continuity could change the results of different operations. Also there will be need to add and formally describe new operators to facilitate the continuous navigation. As an example, the operation *sum* in the discrete structure will be *integration* in the continuous structure.

- **Result visualization**. Results must be displayed in an intuitive and attractive manner to provide a user friendly environment for decision making. This includes cartographic display, grid, graphic representation …etc. Since an important part of the results is estimated it is crucial to give an indication of how good the estimation is through the use of quality indicators. In addition to all that there is the issue of visualizing temporally continuous data which can be treated as a sequence of images (animations).

III. Related Work

In this section we go briefly over related work in multidimensional modeling and continuous field data representation.

A. Multidimensional Modeling

Multidimensional structures are based on the concepts of **facts** and **dimensions**. Facts are defined *"located"* by combinations of members of dimensions if a corresponding

value exists. Dimensions are usually organized in hierarchies that contain several levels where every level corresponds to a level of analysis.

A.A. Concepts and definitions

A data warehouse is defined as "subject oriented, integrated, time variant and non-volatile collection of data in support of management's decision making process" [14]. Most data warehouses consolidate data from different source systems that could have different data organization and format. Data are collected from several operational and external sources, they are then cleaned, transformed, integrated, and afterward they are loaded into the main data warehouse. In almost all applications one of the dimensions of analysis is time to enable analysis of evolution of data over different periods of time [14]. The most popular analysis mean is the *On-Line Analytical Processing (OLAP)* which enables users to examine data within a multidimensional model allowing the retrieval and summarization of data. Fast response to queries is one of the objectives of decision support systems. To speed up query response time some selected aggregations are materialized as summary table or materialized views. Accordingly the queries whose results are found in the pre-aggregated data are answered directly from the materialized views (*summary table*) without actually accessing neither the fact tables nor the dimensions.

A.B. Spatial Data Warehouse

Conventional data warehousing deals with alphanumeric data. However, in the real world spatial data make a large part of the data stored in corporate databases. It has been estimated that about 80% of data have a spatial component, like an address or a postal code [8]. To obtain maximum benefits of the spatial component of the stored data, there have been important advances in spatial data warehousing and spatial *OLAP*. According to [28] a spatial data warehouse is a conventional data warehouse that contains both spatial and non spatial data where both types complement
each other in the support of the decision making process. A spatial data warehouse contains three types of spatial dimensions: (a) Non geometric dimension, (b) mixed dimension and (c) geometric. The first type of spatial dimensions is seen as a hierarchy containing members that are only located with place names (an address or a postal code) and are not represented geometrically. The second type of dimensions is a hierarchy whose detailed level members have a geometric representation but general levels do not have one (at a certain level of aggregation). The last type is a hierarchy whose all members have a geometric representation. Spatial to spatial dimensions can be implemented in multidimensional architectures only when the cartographic representations and navigation are supported by the multidimensional database client [17]. In a spatial data warehouse, in addition to spatial dimensions, measures can also have a spatial representation. [28] and [4] distinguish two types of measures: a numerical measure containing only numerical data and spatial measures which contains a collection of pointers to spatial objects. Two types of spatial measures is made in [23] : the first is a geometric shape or a set of shapes obtained by a combination of several geometric spatial dimensions. The second type is a result of spatial metric or topological operators.

A.C. Spatial OLAP (SOLAP)

Conventional *OLAP* can be used for spatiotemporal analysis and exploration. However, the lack of cartographic representations leads to serious limitations (lack of spatial visualization, lack of map-based navigation, etc) [22]. Therefore to overcome these limitations, visualization tools and map-based navigation tools have to be integrated within conventional *OLAP*. The result would be Spatial OLAP *(SOLAP)* that can be seen as a client application on top of a spatial data warehouse. Spatial *OLAP* is defined as "a visual platform built especially to support rapid and easy spatiotemporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic displays as well as in tabular and diagram displays" [3]. The most important features of *SOLAP* as defined by [5],[22],[23] based on theoretical and implementation works are :

- A flexible interface that supports different data visualization formats.
- All navigation operations must be available in all forms of display (diagrams, maps, tables …).
- Providing the capabilities to define new calculated measures from existing ones.

B. Continuous Field Based Data

Geographic space can be conceptualized as an object (*vector*) or as a field (*raster*) [11]. The boundaries of a spatial object define this conceptualization. Some spatial variables have clearly defined boundaries (e.g. cities, rivers, roads …) and are modeled as spatial objects. Conversely, other spatial variables have fuzzy boundaries (e.g. soil type, vegetation, pollution …) and cannot be described as objects. It is an important requirement for *GIS* that the two approaches (fields and objects) co-exist to handle the different types of data [10]. A continuous field is defined as a portion of space where the applied force to a given point depends only on its position. The number of points in a field is infinite and thus the
values of this field are continuous [21]. The general mathematical definition of a field consists of 3 main parts [15], [24] :

- A domain **D** which is a continuous set,
- A range of values **R** and
- A mapping function ***f*** from the domain **D** to the range **R**.

Computers deal only with discrete data, therefore field-based data are measured in only selected points or zones of the field and are stored as sample points. Interpolation methods are then applied to estimate the values of the phenomenon at points where no measurements have been made. These sample points can be represented in different ways depending on data size and the phenomenon being modeled.

IV. Proposed Solutions

Most of the research issues were looked into nevertheless with different degrees of depth. Our main focus was on formalizing the integration of spatiotemporal continuity within multidimensional structures. This includes defining a formal model and the associated operators. Interpolation methods were also studied in order to determine the most appropriate methods and their most efficient way of application so that they provide good estimation without hindering the performance of the process of exploration and analysis. The issue of range of continuity was well discussed and thoroughly studied to determine at what level continuity is applied. Navigation in the continuous hypercube is dependant on the continuous representation and is a basic component of the multidimensional model therefore this issue was also studied. Certainly data visualization was looked into to provide an intuitive interface for data exploration and analysis. Some of the solutions to the research issues are presented briefly here. Some details can be found later in the paper however technical issues were used during the implementation of the prototype therefore only their results can be seen.

- **Continuous multidimensional model.** The work done on modeling continuous field data is mainly geared towards transactional spatial databases [9],[12],[20]. However in the domain of multidimensional modeling there is no work done on formalizing a model that integrates continuous field data. A survey of existing multidimensional models was performed and it showed that no existing model can handle field based data. There is a necessity to define a model that can both treat discrete hypercubes (sample data) and can support estimated data resulting from interpolation. Based on the survey a conventional multidimensional model was selected and extended to support the field based data type.

- **Range of continuity.** In a multidimensional context detailed data are associated with the lowest level of detail (lowest granularity of the hierarchy). General levels contain aggregated data (i.e. a single value by a combination of level members corresponding to an aggregated measure). As the navigation goes up in detail the number of values decreases leading to a single value that represent the fact being analyzed. As an example, let us take the analysis of the average profit for a chain of stores for a geographical region according to this hierarchy

  Store → City → Region → Country

  The profit is calculated for the lowest level of hierarchy *store* which gives *S* different values (one value per store). Rolling the profits up to the *city* level will give *C* different aggregated values (one value per city). Going up again in the hierarchy will give *R* different aggregated values (one value per region). Rolling the results up one more time will give *one* aggregated value associated with the unique member *Country*. The same principle holds for natural phenomena. At the most detailed level of the spatial or temporal dimensions there are the measured values (sample data) and at higher levels there are aggregated data. If both dimensions are treated as continuous, at the lowest level of detail there is an infinite number of members corresponding to infinite facts that are either measured or estimated. However as soon as we perform a roll up to a certain level we will have a single value corresponding to every member of this level. Certainly this single value is calculated from a continuous representation of the lowest level. All levels higher than the most detailed level have a finite number of members which is consistent with multidimensional approach. The notion of continuity must be confined to the lowest level of detail. For the temporal dimension, continuity is also confined to the lowest level. Applying interpolation methods results in a continuous temporal representation of the phenomena. No new levels should be created because their creation does not result in continuous representation but only in reducing the time intervals between measurements. That is, instead of having discrete values every hour there will be discrete values every minute or every second which is certainly not a temporal continuity. In addition to creating continuous representation, there is another concern that has to be dealt with which is the treatment of missing data. Missing data are synonym with natural phenomena since sensors which measure a phenomenon may malfunction, get blocked, break down or may be even stolen. Based on measured data, cross interpolation techniques are used to estimate missing data.

- **Determination of spatial and temporal interpolation methods.** In the literature there is a wide variety of spatial and temporal interpolation methods. Some methods provide information about the accuracy of estimation and hence they can be useful for indicating data quality. Generally interpolation methods can be applied or classified as local or global. An interpolation method can be applied over all sample values for estimation or they can be applied to a specific number of sample points lying within a defined diameter. Spatial interpolation methods are classified as deterministic (e.g. trend surface analysis, thin plate splines, inverse distance weighting…) or stochastic (e.g. ordinary kriging, simple kriging… etc). Examples of temporal and spatiotemporal interpolation methods include mean over time method, space time product and tetrahedral method. In our prototype the user can choose method to use for interpolation. This approach is more attractive for the experienced users however it can lead to bad modeling of the spatial phenomenon if the user has a limited experience in spatial analysis.

- **Storage and optimization.** For the discrete representation pre-aggregation is used to speed up query response time. For the continuous representation we

considered two solutions: (1) storing only sample points and when the user poses a query that involves locations or times where no values are measured, an interpolating function is invoked using sample points to calculate the required data on the fly. (2) Storing interpolated values at the lowest level of spatial and temporal granularities and sample values at higher levels of granularities so that there will be no need to calculate for non-existing data values. According to our multidimensional model where the cube at the most detailed levels contains sample data values the first choice is more appropriate. The storage cost of the second choice would be enormous. However with the first choice, the storage cost is kept at the possible minimum. Additionally depending on the interpolation methods applied the discrete representation can be used to compute higher level values. For example some interpolation methods give estimations within the minimum and maximum observed values therefore the aggregations *max* and *min* need not be calculated since they will be the same as the measured values and therefore pre aggregated values can be used to answer this type of queries.

- **Navigation in the continuous hypercube.** In addition to the conventional *OLAP* operators we have defined new operators especially for continuous field data. Since data exist all over the field of the study the aggregation function sum will be considered as the integral of the function representing the continuous phenomenon. Figure 2.1 shows the difference between the operation sum on dimension time in conventional *OLAP* and in continuous *OLAP*. In Figure 1a the sum of values for the period between t_1 and t_5 is the sum of the 5 observed values. On the other side (Figure 1b) the sum is the integral of the function *f*(t) for the same period where *f(t)* is the function representing the phenomenon.

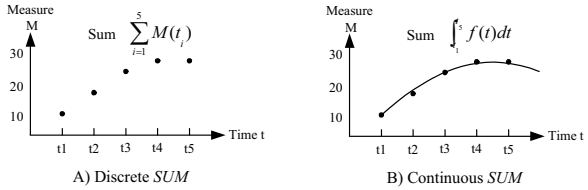

Fig. 1. *Discrete and continuous operation "SUM"*

It should be noted that depending on the interpolation methods used the continuous aggregation and the discrete aggregation may not necessarily have the same results. In addition to extending existing operators we have defined a new calculated measure that is evaluated at the lowest level of detail based on the continuous representation and are aggregated to higher levels.

- **Result visualization.** Complex spatial relationships can be easily and efficiently processed by decision makers when they are viewed as a map rather than as a table [18]. This leads to the importance of having visual and cartographic displays of data as required by the continuous structures. Animation will be used to restore field based data. Since a large part of the displayed information will be estimated, quality of data varies in terms of accuracy. Therefore information about data quality should be indicated to users to decide whether a result should be considered for decision making or not. Data quality is defined as fitness of use [32]; i.e. are the data accurate enough to be used as basis for decisions? This notion is very important particularly for decision making where a decision should be based on reliable information only. Consequently quality indicators will be used to show the quality of results. Based on the number of sample points used and the interpolation method applied, the estimated values will differ in quality. Data quality could be indicated in different forms : a counter, a percentage or using the metaphor of traffic lights as in [7]. The interface will be composed of two main parts as in [22] : a navigation panel and a visualization window. The navigation panel permits the user to select measures to be viewed with respect to members (predefined or not). The visualization space will show the information in different format : cartographic, different types of charts and tabular grids. Continuous data are compared with discrete data to reflect the changes that occur when hidden and missing data are recovered (*estimated*). For temporal continuity the results could be displayed in animated form to simulate continuity in time.

- **Metadata management.** As in conventional data warehousing, in continuous data warehouses metadata play also an important role. In addition to the traditional contents of metadata, the interpolation methods used and how they are used will be included in the metadata. As mentioned earlier, when data values are missing, due to malfunctioning of instruments for example, an indication to that will be specified in the metadata. Also the output formats could be explained in the metadata such as: explanation of functions displayed, quality indicators and their meanings.

V. Continuous Data Warehouse

The creation of a data hypercube for continuous field data is performed in several steps. First of all, data describing the natural phenomena are captured and sent by several sensors distributed over the space of the study. Usually, the sensors capture data at variable time intervals depending on the phenomena being analyzed and produce a discrete representation of the phenomena both spatially and temporally. Data sent by the sensors are stored in a database. The discrete multidimensional structure built from this database is the internal representation of the hypercube. The external representation which is what the user sees is achieved by applying spatial and temporal interpolation methods on data from the discrete hypercube. The user should deal with the continuous representation without in fact noticing the existence of a discrete vision of the hypercube. Results are represented

in different forms such as functions consisting from both actual values and estimated ones, tables, charts, bar graphs, cartographic forms… etc (Figure 2).

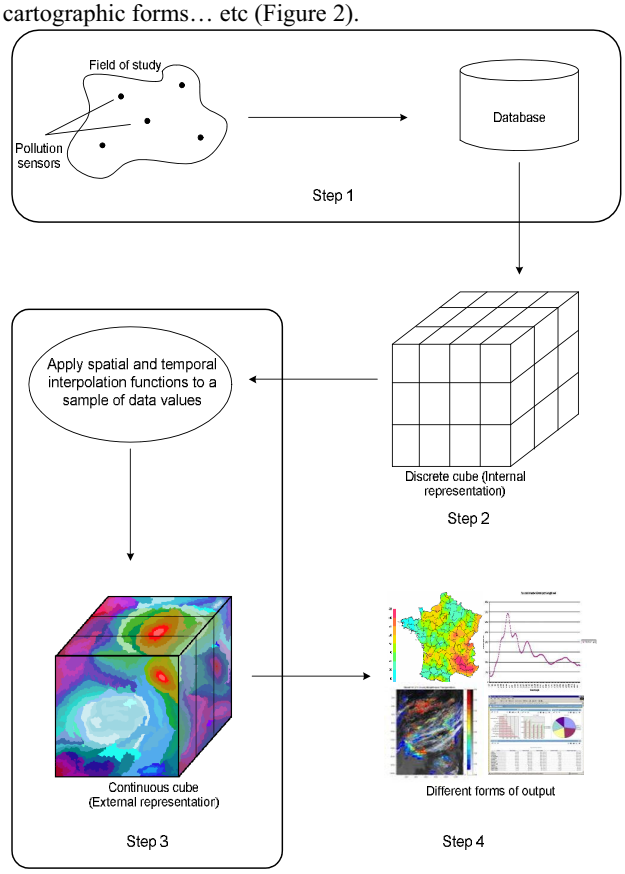

Fig. 2 Steps of building and exploiting continuous hypercubes

A. An MD model for continuous data

Existing multidimensional models were formalized for alphanumeric data and do not take in consideration spatiotemporal continuity. Thus a new multidimensional has to be defined or an existing one has to be extended so that it takes into account the spatiotemporal data characteristics. The model proposed in [31] is the best candidate for extension. In fact the model has the following strong points:

- It is the richest model with respect to contained multidimensional semantics. It is an expressive model that can be easily extended without loosing neither its expressiveness nor its analyzability.
- It contains the formal definition of most of the necessary operations and those that are not defined can be derived in terms of the defined operations. The operations defined in the model are classified in two groups : simple operations and complex operations. The complex operations are built on top of the simple operations.
- It is based on the idea of basic cube which allows for the serial performing of operations which is typical in *OLAP* applications.

The model presented in [31] was extended to take into account spatiotemporal continuity using the concept of basic cubes. In this paper we present some of the necessary elements of the model. For more details we refer the reader to [1].

B. Basic cubes

The model is based on the notion of basic cubes which are cubes at the lowest level of detail. We distinguish two types of basic cubes. The first type of basic cubes is the *discrete basic cube* $\boldsymbol{discC}_b$ which is a 3-tuple $<\boldsymbol{D}_b, \boldsymbol{L}_b, \boldsymbol{R}_b>$ where $\boldsymbol{D}_b$ is a list of dimensions including a dimension measure *M*. The list of the lowest levels of each dimension is denoted $\boldsymbol{L}_b$ and $\boldsymbol{R}_b$ denotes a set of cells data represented as a set of tuples containing both level members and measures in the form of $S = [s_1, s_2, s_3, \ldots s_n, m]$ where *m* is the dimension that represents the measure. In order to achieve a continuous representation of the basic cube, estimated values are derived from the $\boldsymbol{discC}_b$. Estimated measures related to the infinite members of a given spatial dimension or temporal dimension are estimated using actual cell values from the $\boldsymbol{discC}_b$. This of course necessitates applying interpolation functions to a sample of the $\boldsymbol{discC}_b$ values to calculate the measures corresponding to the new dimension members which will give the second type of basic cubes which are called the *continuous basic cube* $\boldsymbol{contC}_b$. The number of tuples in the *continuous basic cube* is theoretically infinite since a dimension level in the class of continuous dimensions contains theoretically an infinite number of members. We define $\boldsymbol{contC}_b$, as 4-tuple $<\boldsymbol{D}_b, (\boldsymbol{D'}_b, \boldsymbol{F}), \boldsymbol{L}_b, \boldsymbol{R'}_b>$. Where $\boldsymbol{D}_b$, $\boldsymbol{D'}_b$ are the discrete dimensions and continuous dimensions respectively.

It can be clearly seen that $\boldsymbol{discC}_b \subseteq \boldsymbol{contC}_b$ as sample values are included within the $\boldsymbol{contC}_b$. The *continuous basic cube* is built from the *discrete basic cube* by applying spatial and temporal interpolation functions.

C. Cubes

Cubes are built from basic cubes. A cube *C* is defined as 4-tuple $<\boldsymbol{D}, \boldsymbol{L}, \boldsymbol{contC}_b, \boldsymbol{R}>$ where, similar to the basic cubes, *D* is a list of dimensions including *M* as defined above, *L* is the respective dimension level, *R* is cell data and $\boldsymbol{contC}_b$ is the basic cube from which the cube *C* is built. Because of the nature of continuous field data, different aggregation functions are used to build the cube at higher dimension hierarchies. For example, the sum of the measure for a specific region or a specific period of time will be represented as the integration of the function representing the phenomenon. Other aggregation functions like *min*, *max* or *average* will be performed on $\boldsymbol{contC}_b$ and their results will be assigned to the higher levels of the hierarchy. In addition a new calculated measure applicable only to field based data is defined and will be calculated in the lowest level of detail and it can be aggregated to higher levels of the cube.

Based on data values used to obtain the aggregation, the aggregation operations on continuous multidimensional structures can be classified as either discrete or continuous. Discrete operations use only sample data values and their aggregations will correspond to the discrete higher levels. Continuous operations use all data values of the field (sample values and estimated values). Their aggregations will correspond to the higher levels resulting in aggregated values

based on continuous representation, [1]. The same multidimensional schema is used for both cases with the only difference being the detailed data used in calculating the aggregations. One can imagine the existence of a parallel hierarchy that is used for the continuous representations (Figure 3). It should be noted that continuous aggregated values and discrete aggregated values are not necessarily equal.

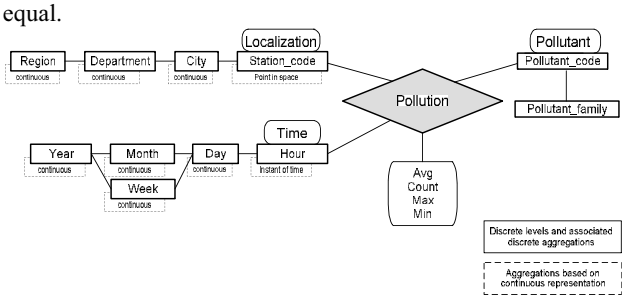

Fig. 3 Steps of building and exploiting continuous hypercubes

Discrete aggregations

They include the conventional *OLAP* operations and are calculated based on the real observed values. We list here the most common operations:

DiscMax = v_i such that $v_i > v_j \ \forall v_j \in V^*$

DiscMin = v_i such that $v_i < v_j \ \forall v_j \in V^*$

DiscSum = $\sum_{i=1}^{n} v_i$ for $v_i \in V^*$ where n is the number of observed values.

DiscAvg = $\frac{\sum_{i}^{n} v_i}{Card\ (V^*)}$ for $v_i \in V^*$ where n is the number of observed values

Continuous aggregations

The second category of operations concerns the operations that involve all domain values of the phenomenon i.e. all observed and estimated values of the phenomenon are used to produce a hypercube based on continuous representation. In addition to aggregation functions a new calculated measure has been defined that apply only on continuous data. Continuity is considered on either an interval of time or on a specified region. The continuous aggregations on the measure of the continuous field are [1] :

ContMax = v_i such that $v_i > v_j \ \forall v_j \in V$

ContMin = v_i such that $v_i < v_j \ \forall v_j \in V$

ContSptSum = $\iint f(x, y, t)\,dxdy$ where $(x, y) \in D$

ContTmpSum= $\iiint f(x, y, t)\,dxdydt$ where $(x,y) \in D$

ContSptAvg= $\frac{\iint f(x, y, t)\,dxdy}{area}$ where $(x, y) \in D$

ContTemporalAvg = $\frac{\iiint f(x, y, t)\,dxdydt}{t_2 - t_1}$ where $[t_1; t_2]$ is an interval of time and $(x, y) \in D$

[1] Definition of continuous field from [1].

The calculated measure *gradient*, which is the change of the value of the field that results by a change of a unit of space, can be applied to most detailed level of data and its results can be aggregated to higher levels :

Gradient = grad($(x, y), t$) = v= $\left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial t}\right]$ where $(x, y) \in D$.

Gradient is the change of the value of the field that results by a change of a unit of space.

VI. Continuous SOLAP Prototype

One of the applications concerned by continuous multidimensional structures is air pollution analysis. We designed and implemented a prototype of an application that observes air pollution in order to validate the model defined above and to show the potentials of what we termed *continuous spatial on-line analytical processing CSOLAP*. In order to validate our model and test the performance of our data warehouse a considerable volume of data is required. We used data published by AIRPARIF, [2]. However this set covers only the Parisian region. For the rest of France we had to simulate pollution values so that we ended up with a large set of data that we stored in an SQL Server data warehouse. To create a continuous representation all dimensions have to be at their lowest level then we choose an interpolation function from a list of different functions. This will create a thematic map on the fly. The thematic map all values of the field for the chosen dimensions (Figure 4).

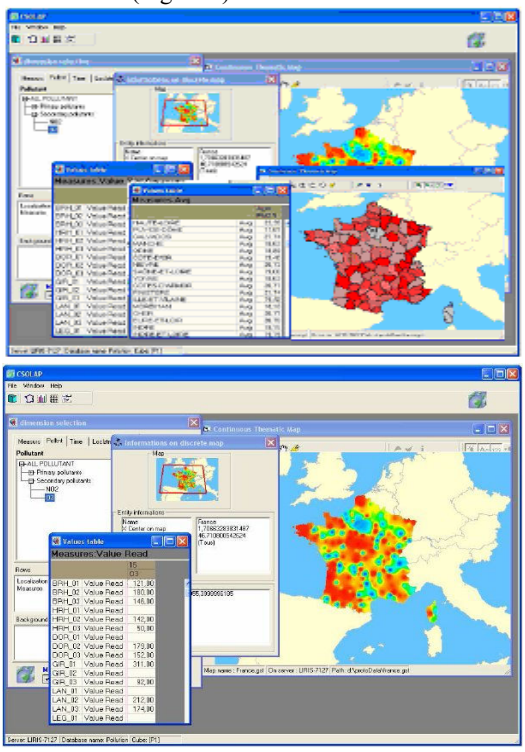

Fig. 4 Screens from CSOLAP prototype

VII. Conclusions and Future Work

Spatial decision support systems have benefited from advances made in OLAP and data warehouses technologies. The emergence of SOLAP has made spatial decision support easy and flexible. However the perception of space and time is still limited to the discrete perception which does not represent natural phenomena correctly. Our work aims at the integration of spatiotemporal continuity in multidimensional structures. Starting from a discrete hypercube, a continuous hypercube is created by applying interpolation functions over cell data. We have two representations : the first is an internal representation (machine level), the second is an external representation (user level).

A multidimensional model dedicated to continuous field data was defined. Along with model, a set of operations were defined. The model and the prototype are validated by a prototype of a data warehouse of air pollution. However not all research issues listed in this paper have been looked into. Our main objective was defining a formal model. Defining additional operators, metadata management, dealing with different intervals of time, storing continuous hypercube and ad-hoc hierarchies are a list of the perspectives of this research.

References

[1] Ahmed, T. O. and Miquel, M. Multidimensional Structures Dedicated to Continuous Spatiotemporal Phenomena. Proc *22nd British National Conf. on Databases (BNCOD22),* Sunderland. 2005, 29-40.

[2] AIRAPRIF, Monitoring the quality of air in Ile de France. http://www.airparif.asso.fr.

[3] Bédard, Y. "Spatial OLAP." *Vidéo-conférence*. 2ème Forum annuel sur la R-D, Géomatique VI: Un monde accessible, 13-14 Nov., 1997, Montréal [online].

[4] Bédard, Y., Merrett, T. and Han, J. Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery in Geographic Data Mining and Knowledge Discovery. *Research Monographs in GIS series* edited by Peter Fisher and Jonathan Raper. 2001. 53-73.

[5] Bédard, Y., Proulx, M.J. and Rivest, S. Enrichissement du OLAP pour l'analyse géographique: exemples de réalisations et différentes possibilités technologiques. *1ere journée francophone sur les entrepôts de données et l'analyse en ligne*, Lyon, 2005. In French.

[6] Cowen, D. J. "GIS versus CAD versus DBMS: What are the differences?" *Fotogrammetric Engineering and Remote Sensing*, 54, 1988, 1551-1555.

[7] Devillers, R., Gervais, M., Bédard, Y. and Jeansoulin, R. 2002. Spatial Data Quality: From Metadata to Quality Indicators and Contextual End-user Manuel, in OEEPE-ISPRS, Istanbul. Joint Workshop on Spatial Data Quality.

[8] Franklin, C. An Introduction to geographic Information Systems: Linking Maps to databases. *Database,* 1992, 13-21.

[9] Filho, J. L. and Iochpe, C. 1999. "Specifying analysis patterns for geographic databases on the basis of a conceptual framework." In the Proc. of the 7th ACM int. symposium on Advances in GIS, Kansas City, Missouri, United States, pp 7 – 13.

[10] Galton, A. Integrating Fields and Objects in Geographic Information Science. *Workshop on Fundamental Issues in Spatial and Geographic Ontologies*, Switzerland, 2003.

[11] Goodchild, M. Geographical information science. *International Journal of GIS*, 2003, *6*, 31-45.

[12] Gordillo, S. "Modélisation et manipulation de phénomènes continus spatio-temporels." Thèse de doctorat. Université Claude Bernard Lyon I. 2001. In French.

[13] Hasenauer, H., Haslik, I., Rosenthaler, R., Pernul, G. and Stangl, D. Conceptual framework of a data warehouse for the National park Hohe Tauern. Proc. *13th Int. Symposium "Informatik für den Umweltschutz" der Gesellschaft für Informatik* (GI), Magdeburg, 1999, 478-480.

[14] Inmon, W. H. Building the Data Warehouse. John Wiley and sons. 1992.

[15] Kemp, K. and Vckovski, A. Towards an Ontology of Fields. Proc. of the *3rd Int. Conf. on GeoComputation*, Bristol, UK, 1998.

[16] Kouba, Z., Matousek, K. and Milkovsky, P. On Data Warehouse and GIS integration. Proc. of the *11th Int. Conf. and Workshop on Database and Expert Systems Applications*, Greenwich, 2000, 604-613.

[17] Marchand, P., Brisebois, A., Bedard, Y. and Edwards, G. Implementation and Evaluation of a Hypercube-Based Method for Spatiotemporal Exploration and Analysis. *ISPRS journal of photogrammetry and remote sensing*. 59 (1,2), 2004, 6-20.

[18] Mennecke, B. and Higgins, G. 1999. "Spatial Data in the Data Warehouse: A Nomenclature for Design and Use," the *5th Ann. Americas Conf. on Information Systems*. pp. 274 - 276.

[19] Morgan, D. G. and Glover, T. Distributing Data Ownership: The Northwestern Geospatial Data Network. GIS 2001. Vancouver, B.C., February 10-22.

[20] Mostaccio, C. A. 2003. "Organisation physique des bases de données pour les champs continues." Thèse de doctorat, Université Claude Bernard Lyon I, France. In French.

[21] Pariente, D. Estimation, modélisation et langage de déclaration et de manipulation de champs spatiaux continus." Thèse de doctorat. Institut National des Science Appliquées de Lyon. 1994. In French.

[22] Rivest, S., Bédard, Y. and Marchand, P. Towards Better Support for Spatial Decision Making: Defining the Characteristics of Spatial On-Line Analytical Processing (SOLAP). *Geomatica, the journal of the Canadian Institute of Geomatics*, 55, 2001, 539-555.

[23] Rivest, S., Bedard, Y., Proulx, M.J. and Nadeau, M. SOLAP: A New Type of User Interface to Support Spatiotemporal Multidimensional Data Exploration and Analysis. Proc. of *ISPRS workshop on Spatial, Temporal and Multi-Dimensional Data Modeling and Analysis*, Québec City, Canada, 2003.

[24] Schabenberger, O. and Gotway, C. A. Statistical Methods for Spatial Data Analysis. Chapman & Hall/CRC Pres. 2005.

[25] Shanmugasundaram, J., Fayyad, U. M. and Bradely, P. S. Compressed data cubes for OLAP Aggregate Query Approximation on Continuous Dimensions. Proc. of the *5th ACM SIGKGG International Conf. on Discovery and Data Mining (KDD99)*, New York, 1999, 223-232.

[26] Shen, S., Dzikowski, P., Li, G. and Griffith, D. Interpolation of 1961-97 Daily Temperature and Precipitation Data onto Alberta Polygons of Ecodistrict and Soil Landscapes of Canada. Journal of. Applied Meteorology, 40, 2162 – 2176.

[27] Staudt, M., Vaduva, A. and Vetterli, T. The Role of Metadata for Data Warehousing." *Technical Report* 99.06, Department of Information Technology, University of Zurich, September, 1999.

[28] Stefanovic, N., Han, J. and Koperski, K. Object-based Selective Materialization for Efficient Implementation of Spatial Data Cubes. *IEEE Transactions on Knowledge and Data Engineering*, *12*(6), 2000, 938 - 957.

[29] Tan, X. Data Warehousing and Its Potential Using in Weather Forecast. *Proc. 22nd Int. Conf. on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*. Atlanta, GA. 2006.

[30] Tchounikine, A., Miquel, M., Laurini, R., Ahmed, T.O., Bimonte, S. and Baillot, V. Panorama de travaux autour de l'intégration de données spatio-temporelles dans les hypercubes. *Revue des Nouvelles Technologies de l'Information (RNTI)*, Editions Cepaduès, numéro spécial, Juin, 2005. In French.

[31] Vassiliadis, P. Modeling Multidimensional Databases, Cubes and Cube Operations. Proc. of the *10th Int. Conf. on Scientific and Statistical Database Management (SSDBM)*, Capri, Italy, 1998.

[32] Veregin, H. 1999. "Data Quality Parameters." In *Geographical Information Systems*, Vol. Principles and Technical Issues (Eds, Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W.) John Wiley & Sons, Inc., pp. 177-189.