Multidimensional Structures for Field Based Data

A Review of Models

Taher Omran Ahmed
IT Department, College of Applied Sciences,
Sultanate of Oman,
Computer Science Department, Faculty of Science
Aljabal Algharbi University, Azzentan, Libya
e-mail: taher.ibr@cas.edu.om, fenneer@yahoo.com

***Abstract*— The growth in size of geographical data collected by different types of sensors and used in diverse applications have led to the adoption of spatial data warehouses (SDW) and spatial OLAP (SOLAP). Spatial data are represented as either discrete (objects) or continuous (raster), also called field-based data. The latter representation deals with data about natural phenomena that exist continuously in space and time. Continuity was not taken into account during the early days of SDW and SOLAP, however, during the last decade there were attempts to include this concept within SOLAP. In this paper, we present some of the concepts, issues and discuss the models proposed to represent spatiotemporal continuity in a decisional context and attempt to foresee where the research in this area is heading.**

Keywords-Data Warehousing; Field Based Data; Models; Multidimensional; SOLAP; Spatial Data.

I. INTRODUCTION

Data warehouses rely on multidimensional structures, which are based on the concept of facts or measures and dimensions. Facts are the subject of analysis and they are usually numeric values. Facts are defined by the combination of values (members) of dimensions if a corresponding value exists. Dimensions represent the context in which measures or facts are analyzed. Usually, dimensions are organized in hierarchies composed of several levels; each level represents a level of detail as required by the expected analysis. A hierarchy is a set of variables which represent different levels of aggregation of the same dimension and which are interlinked by a mapping [18]. Traditional data warehouses deal with alphanumeric data; however, most businesses take geographical location seriously when they seek good decisions and hence a large segment of data stored in corporate databases is spatial. It has been estimated that about 80% of data have a spatial component to it, like an address or a postal code [8]. In order to obtain maximum benefits of the spatial component of data, there had been important efforts that led, eventually, to the introduction and implementation of spatial data warehousing (SDW) and spatial OLAP (SOLAP).

Spatial data warehousing has been recognized as a key technology in enabling the interactive analysis of spatial data sets for decision-making support [11][19]. According to [22] a spatial data warehouse is a subject oriented, integrated, time-variant and non-volatile collection of both spatial data and non-spatial data in support of management's decision-making process. In plain terms, it is a conventional data warehouse that contains both spatial and non-spatial data where these two types of data complement each other in the support of the decision making process.

OLAP is a tool for analysis and exploration of conventional (alphanumeric) data warehouses. It can also 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, and so on) [19]. To overcome these limitations, visualization tools and map-based navigation tools have to be integrated within the conventional OLAP. The result would be a OLAP that can be seen as a client application on top of a spatial data warehouse [2]. The presence of these components would give specialists from multiple disciplines (forestry, public health, transport,…etc.) a new exploration and analysis potential known as Geographic Knowledge Discovery (GKD) [3].

This paper sheds a light on the research carried on spatial OLAP for continuous field data and the different models that have been proposed. In order to achieve our objective, a brief overview of spatial data warehousing is presented in Section 2. The rest of the paper is organized as follows: In Section 3, we discuss the major concepts of several multidimensional models for continuous field data that have been proposed. A comparison between models is presented in Section 4 and we conclude the paper in Section 5.

II. SPATIAL DATA WAREHOUSINGA AND SPATIAL OLAP

A spatial data warehouse contains three types of spatial dimensions; non-spatial, mixed and spatial dimensions.

The first type is a hierarchy containing members that are only located with place names (an address or a postal code) and not represented geometrically. The absence of geometrical representation handicaps the spatiotemporal exploration and analysis but it is still possible for the users to carry out the spatial cognition [13]. The second type is a hierarchy whose detailed level members have a geometric representation but general levels do not have one (at a certain level of aggregation). An example of this type of dimensions is using maps with polygons for cities and regions but neither

 ISBN: 978-1-61208-486-2

for states nor country. In the last type of spatial dimensions, all members have a geometric representation.

In addition to spatial dimensions, two types of spatial measures are also distinguished, the first being 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 [20].

A literature review of SOLAP shows an extensive amount of work, unfortunately most of the published research focused on the implementation side and on showing the advantages and potential uses of SOLAP. To the best of our knowledge, there is not enough solid attempts to go in-depth in the theory behind the concepts or to propose a sound mathematical model. Hence, SOLAP remained just an application that can be seen - by many - merely as a coupling between OLAP and GIS. Moreover, SOLAP solely deals with the discrete representation of GIS and hence it cannot grasp the essence of continuity in natural phenomena. Even when continuous field data are dealt with in a decisional context [14] they are treated as discrete and hence they lack their main characteristic that is, spatiotemporal continuity. For instance, [24] proposed a logical model to integrate spatial dimensions representing incomplete field data at different resolutions in a classical SOLAP architecture. Nonetheless, during the last ten or more years, serious attempts were made at integrating continuity in multidimensional structures.

III. MODELS

There is no consensus on one specific model to represent multidimensional structures for field-based data. In fact, even conventional multidimensional structures, which are more established, use different models. Over the last few years, several models were proposed for spatiotemporal multidimensional structures such as [21], where a logical multidimensional model to support spatial data on SDW is proposed. But, most of them concentrate on spatial continuity rather than spatiotemporal continuity.

In this section we present an overview of the main concepts of (four) major models.

A. *Ahmed and Miquel, 2005*

The model presented in [1] is one of the earliest models for multidimensional structures that attempts to include continuous field data as measures and dimensions. The researchers present the concepts, research issues and potentials of continuous multidimensional structures and propose a model for continuous field data. The model is based on the concept of basic cubes, which are used as the lowest level of detail of dimensions. To imitate the behavior of natural phenomena a continuous data warehouse will be treated as a second layer on top of the discrete data warehouse.

1) *Basic Definitions*

There are n dimensions with r being the rank of dimension levels starting from (*level 1*) all the way up to (*level r*) and k being the cardinality (number of members) of a given dimension level DL_i. The domain of values $dom(DL_i)$ for a given dimension level DL_i may contain two types of members: predefined members and any possible value between any two given members to give a continuous representation of the dimension level. A value x belonging to a specific dimension level DL_i can have ancestors and descendants which are specific instances related to x at higher and lower dimension levels respectively.

2) *Basic Cubes*

A basic cube is a cube at the lowest level of detail. The discrete basic cube $\boldsymbol{discC_b}$ is a 3-tuple $<\boldsymbol{D_b, L_b, R_b}>$ where $\boldsymbol{D_b}$ is a list of dimensions including a dimension measure M. L_b is the list of the lowest levels of each dimension and $\boldsymbol{R_b}$ is a set of cell data represented as a set of tuples containing level members and measures in the form of $x = [x_1, x_2,...x_n, m]$ where m is the dimension that represents the measure.

To obtain a continuous representation of the basic cube, estimated measures related to the infinite members of a given spatial and temporal levels are calculated using actual cell values from $\boldsymbol{discC_b}$. This involves applying interpolation functions to a sample of $\boldsymbol{discC_b}$ values to calculate the measures corresponding to the new dimension members, which will result in continuous basic cube $\boldsymbol{contC_b}$.

A continuous basic cube contCb is 4-tuple $<\boldsymbol{D_b, (D'_b, F), L_b, R'_b}>$. Discrete and continuous dimensions are represented by $\boldsymbol{D_b}$, $\boldsymbol{D'_b}$ respectively. In addition, $\boldsymbol{F}$ is a set of interpolation functions associated with the continuous dimensions and has the same cardinality as $\boldsymbol{D'_b}$. The lowest levels of dimensions are represented by $\boldsymbol{L_b}$. The continuous representation is defined over a spatial and/or temporal interval. Hence, $\boldsymbol{R'_b}$ is a set of tuples of the form $x = [x_1, x_2,...\ x_n, m]$ where $x_i \in [minDom(Lb_i) - \Delta, maxDom(Lb_i) + \Delta]$ with Δ being a small predefined value used to allow for continuous representation around the values of the domain of the dimension levels, and to predict values outside the specified interval. The measure m $\subseteq$ M can be either interpolated (approximated) or exact. When measures are interpolated m is defined as:

- $m= f(m_1, m_2,...,m_k)$ where f is a spatial or temporal interpolation function,
- $m= f(m_1, m_2,...,m_k)$ using values lying within a predefined spatial or temporal distance d, or
- $m = f_1 \circ f_2 (m_1, m_2,...,m_k)$ or $m = f_2 \circ f_1 (m_1, m_2,..., m_k)$. The order of applying interpolation functions.

It can be concluded that $discC_b \subseteq contC_b$

3) *Cubes*

Cubes at higher levels are built by applying a set of operations on data at the basic cube level. A cube C is defined as 4-tuple $<\boldsymbol{D}, \boldsymbol{L}, \boldsymbol{contC_b}, \boldsymbol{R}>$ where, D is a list of dimensions including M as defined before, L is the respective dimension level, R is cell data and $\boldsymbol{contC_b}$ is the basic cube from which C is built. Because of the nature of the continuous field data, different aggregation functions are used to build the cube at higher dimension hierarchies. For example, the sum of measures for a specific region or a specific period will be represented as an integral. Other aggregation functions like *min*, *max* or *average* will be performed on ***contCb*** and their results will be assigned to the higher levels of the hierarchy.

4) *Aggregations*

On a continuous field, two classes of operations are defined. The first deals with discrete operations and the second groups the continuous operations:

Discrete operations. Only the sample points are used.

$$\text{DiscMax} = v_i \text{ such that } v_i \geqslant f(s_k)\ \forall\, s_k \in S$$
$$\text{DiscMin} = v_i \text{ such that } v_i \leqslant f(s_k)\ \forall\, s_k \in S$$
$$\text{DiscSum} = \sum_{e_k \in E} f(S_k)$$
$$\text{DiscCount} = \text{Card}(S)$$
$$\text{DiscAvg} = {}^{DiscSum}/_{DiscCount}$$

Continuous operations. All values of the field are used.

$$\text{ContMax} = v_i \text{ such that } v_i \geqslant f(S_k)\ \forall\, S_k \in D \times T$$
$$\text{Cont Min} = v_i \text{ such that } v_i \leqslant f(S_k)\ \forall\, S_k \in D \times T$$
$$\text{ContSpatSum} = \int f(s)\,dp$$
$$\text{ContTempSum} = \int f(s)\,dt$$
$$\text{ContSpatAvg} = \int \frac{f(s)\,dt}{(region\ area)}$$
$$\text{ContTempAvg} = \int \frac{f(s)\,dt}{(t_2 - t_1)}$$

where $[t_1{:}t_2]$ *is a time interval*

B. *Vaisman and Zimányi, 2009*

Vaisman et al. [23] base their multidimensional model for continuous fields in spatial data warehouses on MultiDim model presented in [12]. A multidimensional schema consists of a finite set of dimensions and fact relationships. A dimension consists of at least one hierarchy, containing at least one level. A basic hierarchy is hierarchy with only one level. Several levels are related to each other through a binary relationship that defines a partial order $\preceq$ between levels. For any two consecutive related levels l_i, l_j, if $l_i \preceq l_j$ then l_i is called child and l_j is called parent. A level representing the less detailed data for a hierarchy is called a leaf level.

For spatial levels, the relationship can also be topological requiring a spatial predicate, e.g., intersection. A fact relationship may contain measures that can be spatial or thematic. The thematic measures are numeric that are analyzed quantifiably whereas the spatial measure can be represented by a geometry or field, or calculated using spatial operators, such as distance or area.

The dimension levels have two types of attributes (category and property). For parent level, the category attribute defines how child members are grouped. In the leaf level, the category attribute indicates the aggregation level of a measure in the fact relationship. The property attribute can be spatial (represented by geometry or field) or thematic (descriptive, alphanumeric data types). Hence, property attribute provides additional features of the level. A level is spatial if it has at least one spatial property attribute. Likewise, a hierarchy is spatial if it has at least one spatial level.

A *field type* is defined as a function from the spatial domain to a base type. Field types are obtained by applying the field(.). Therefore, the result of field(real) (e.g. representing a natural phenomenon) is a continuous function f : point → real. There are two types of fields (temporal and nontemporal). Field types are partial functions, i.e., they may be undefined for certain regions of space. Along with field types a set of operations over fields are defined and classified as in Table 1.

TABLE I. FIELDS AND OPERATION SETS

Class	Operations
Projection to Domain/Range	defspace, rangevalues, point, val
Interaction with Domain/Range	atpoint, atpoint, atline, atregion, at, atmin, atmax, defined, takes, concave, convex, flex
Rate of change	partialder_x, partialder_y
Aggregation operators	integral, area, surface, favg, fvariance, fstdev
Lifting	Operations on discrete types are generalized for field types

1) *Relational Calculus*

To express SOLAP queries [23], use a query language based on the tuple relational calculus (as in [7]) extended with aggregate functions and variable definitions. They show that this language expresses standard SOLAP queries and that, by extending the calculus with *field* types multidimensional, queries over fields can be expressed. The query language is introduced by example.

Figure 1 and Figure 2 show a query over discrete data and SOLAP operations respectively.

Query 1: Display the name and population of counties in California.
Answer: {c.name, c.population | County(c) ∧ ∃s (State(s) ∧ c.state = s.id ∧ s.name = \`California')}

Query 1: Compute the total population of counties in California.
Answer: sum({c.population | County(c) ∧ ∃s (State(s) ∧ c.state = s.id ∧ s.name = \`California')})

Figure 1. Query over discrete data

Query 3: Total area of land plots located within 10 km from Orange County that intersect San Diego County.
Answer: {c.name, nbLandPlots | County(c) ^nbLandPlots =
count({*l*.id|LandPlot(*l*)| ∧ ∃ *p* (Crop(*p*) ∧
sum$_2$({*y*.id, *y*.production | Yield(*y*) ∧ *y*.crop= *p*.id ∧
***y*.landPlot= *l*.id ∧ *l*.county = *c*.id ∧ ∃ *t* (Time(*t*) ∧**
***y*.time= *t*.id ∧ *t*.date ≥ 1/3/2008 ∧ *t*.date ≤ 31/3/2008)})>100,000)})}**

Figure 2. SOALP operations

2) Extending OLAP Operations with Continuous Fields

In addition to operations defined on fields specified in 3.2.1, they also define another set of operations that allows the interaction with domain and range. These operations are listed in Figure 3.

atpoint, atpoints, atline, atregion	restrict the function to a given subset of the space defined by a spatial value.
at :	restricts the function to a point or to a point set (a range) in the range of the function.
concave, convex:	restrict the function to the points where it is concave or convex, respectively
flex :	restricts the function to the points where convexity changes

Figure 3. Domain and range operators

There are also operators to compute how the field changes in space. Moreover, there are three aggregate operators defined as follows:

Field average favg : integral/area

Field variance fvariance : $\iint_{g} \frac{(f(x,y)-\text{favg})^2}{area}$ dxdy

Field Standard deviation fstdev : $\sqrt{\text{fvariance}}$

A class of multidimensional queries over fields denoted SOLAP-CF queries is also defined. Fields are classified as *temporal* identified by f(🌍,🕓) pictogram and *Non-temporal* identified by *f*(🌍) pictogram. In the Non-temporal field, each point in the space has a value (e.g., soil type) whereas for temporal field there is a value that changes with time instant at each point in the space (e.g. temperature). The model also supports field measures that can be pre-computed in the pre-processing stage as a function of many factors.

C. Bimonte and Kang, 2010

Prior to introducing their model, Bimonte et al. [4] define four requirements for definition of a formal model for field data as dimension and measures:

1. Measures as continuous field data
2. Hierarchy on continuous field data
3. Aggregation functions as Map Algebra functions
4. Independence of implementation

They argue that no existing model satisfies all 4 requirements.

1) Definitions

As in the previous models, Bimonte et al. [4] start by providing a uniform representation for field and vector data, which are used to define measures and dimensions members of the multidimensional model.

Real world entities are represented by 3 types of objects described by alphanumeric attributes. An object can represent levels and members of dimensions.

a) **An object**

An Object Structure S_e is a tuple $\langle a_1 \ldots an \rangle$ where $\forall\, i \in [1,\ldots n]$ a_i is an attribute defined on a domain $dom(a_i)$

An Instance of an Object Structure S_e is a tuple $\langle val(a_1),\ldots val(a_n)\rangle$ where $\forall\, i \in [1,\ldots n]$, $val(a_i) \in dom(a_i)$

'$I(S_e)$' denotes the set of instances of S_e

b) **Geographic Object**

A geographic object is a geometry (geom) and an optional set of alphanumeric attributes ([$a_1, \ldots a_n$]) whose values are associated to the whole geometry according to the vector model.

Let $g \subset R^2$. An Object Structure $S_e = \langle geom, [a_1,\ldots a_n]\rangle$ is a Geographic Object Structure if the domain of the attribute geom is a set of geometries: $dom(geom) \in 2^g$

geom is called 'geometric support'

c) **Field Object**

A Field Object extends a Geographic Object with a function that associates each point of the geometry to an alphanumeric value.

Let $S_e = \langle geom, field, [a_1,\ldots a_n]\rangle$ a Geographic Object Structure. S_e is a Field Object Structure if the domain of the attribute field is a set of functions defined on m sub-sets of points of geom having values in an alphanumeric domain dom_{field} : $dom(field)= \{f_1 \ldots f_m\}$

An Instance of a Field Object Structure S_e is a tuple $\langle g, f_j, val(a_1),\ldots val(a_n)\rangle$ where:

- *$\forall\, i \in [1,\ldots n]$ $val(a_i) \in dom(a_i)$, $g \in dom(geom)$*
- *$f_j : g \rightarrow dom_{field}$ and $f_j \in \{f_1, \ldots, f_m\}$*

'field support' is the input domain of f_j

2) Spatio-multidimensional Model for Field Data

A spatio-multidimensional model uses data as dimensions composed of hierarchies, and facts described by measures. A hypercube is an instance of the spatio-multidimensional model.

a) Hierarchies and facts

Vector objects are organized in a hierarchical way. A Spatial Hierarchy organizes the Geographic Objects into a hierarchy structure using a partial order $\preceq_h$ where $S_i \preceq_h S_j$ means that S_i is a less detailed level than S_j. Measures are aggregated according to the groups of spatial members defined by the tree $<_h$.

Field Hierarchy is defined as a hierarchy of field objects. A Field Hierarchy Structure, $\mathcal{H}_h$, is a tuple $\langle \mathcal{L}_h, \angle_h, \ulcorner_h, \preceq_h \rangle$ where:

- *$\angle_h$, $\ulcorner_h$, are of Field Object Structures, and $\mathcal{L}_h$ is a set of Field Object Structures*
- *$\preceq_h$ is a partial order defined on $\mathcal{L}_h$, $\angle_h$, $\ulcorner_h$.*

An Instance of a Field Hierarchy Structure $\mathcal{H}_h$ is two partial orders: $<_h$ and $<_f$ such that:

- *$<_h$ is defined on the instances of $\mathcal{L}_h$, $\angle_h$, $\ulcorner_h$. Noted as $<_h$ 'geographic objects order'*

- $<_f$ is defined on the field supports of the instances of $\mathcal{L}_h$, $\angle_h$, $\ulcorner_h$ such that:

- if $cood_i <_f cood_j$ then $S_i \leq_h S_j$, where $cood_i$ belongs to a field support of an instance of S_i, and $cood_j$ belongs to a field support of an instance of S_j, ($cood_i$ and $cood_j$ are geometric coordinates)

- $\forall$ $cood_i$ which does not belong to the field supports of the instances of $\ulcorner_h$, $\exists$ one $cood_j$ belonging to the field support of an instance of S_j such that $cood_i <_f cood_j$

- $\forall$ $cood_i$ which does not belong to the field supports of the instances of $\angle_h$, $\exists$ $cood_j$ belonging to the field support of an instance of S_j such that $cood_j <_f cood_i$.

$<_f$ is 'field objects order'

The set of leafs of the tree represented by $<_h$ with root t_i are denoted as leafs($\mathcal{H}_h$, t_i).

The set of leafs of the tree represented by $<_f$ with root $cood_i$ are denoted as leafsFieldSupport($\mathcal{H}_h$, $cood_i$).

Based on the definitions above, the model uses a concept of Field Cube Structure that represents the spatio-multidimensional model schema. The model supposes the existence of only one spatial dimension and one field measure.

b) Field Cube

A Field Cube Structure, $\mathcal{FC}_c$, is a tuple $\langle \mathcal{H}_1, \ldots \mathcal{H}_n$, $\mathcal{FieldObject} \rangle$ where:

- $\mathcal{H}_1$ is a Field Hierarchy Structure (Spatial dimension)

- $\forall i \in [2,\ldots n]$ $\mathcal{H}_i$ is a Hierarchy Structure.

- $\mathcal{FieldObject}$ is Field Object Structure.

An Instance of a Field Cube Structure $\mathcal{FC}_c$, I($\mathcal{FC}_c$), is a set of tuples $\{\langle t\mathcal{b}_1, \ldots t\mathcal{b}_n, t\mathcal{b}_f \rangle\}$ where:

- $\forall i \in [1,\ldots n]$ $t\mathcal{b}_i$ is an instance of the bottom level of $\mathcal{F}_i$ ($\angle_i$).

- $t\mathcal{b}_f$ is an instance of $\mathcal{FieldObject}$.

The instance of the spatio-multidimensional model is a hypercube. A hypercube can be represented as a hierarchical lattice of cuboids.

Field measures are aggregated from fact table data (basic cuboid) to represent non-basic cuboids.

Aggregations from cuboids to higher levels are classified as in Figure 4.

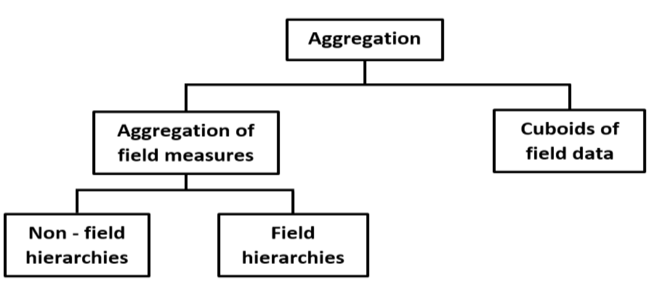

Figure 4. Types of aggregations

3) Aggregation of Field Measures

I. spatial aggregation

Let $\mathcal{G}$ be the geometric attribute. Its aggregation is defined by means of a function $O^{\mathcal{G}}$ that has as input n geometries of the attribute $\mathcal{G}$, and that returns one geometry:

$O_{\mathcal{G}}$: dom($\mathcal{G}$)$\times\ldots\times$dom($\mathcal{G}$) $\rightarrow 2^{\mathcal{G}}$ where $\mathcal{G}$ is a subset of the Euclidian Space R^2

II. alphanumeric aggregations

Let A be an alphanumeric attribute. Its aggregation is defined by means of a function O_A that has in input n values of the attribute A, and that returns one value of the attribute A:

$O_{\mathcal{A}}$: dom($\mathcal{A}$)$\times\ldots\times$dom($\mathcal{A}$) $\rightarrow$ dom($\mathcal{A}$)

4) Gómez and Gómez, 2011, 2012

The model proposed in [23] is extended in [9][10]. The extension includes proposition of a closed generic map algebra over continuous fields. The algebra serves as basis for a language that allows analyzing continuous field data and OLAP cubes using traditional OLAP operations. For the sake of briefness, we will present cube operations for continuous data starting by basic definitions.

a) Spatial Dimension Schemas

A spatial schema is a tuple $\langle$*name$\mathcal{DS}$, $\mathcal{L}$*, $\rightarrow\rangle$ where:

(a) name$\mathcal{DS}$ is a literal;

(b) $\mathcal{L}$ is a non-empty finite set of names called levels (e.g. province, country) which contains a distinguished level name *All;*

(c) Each level $l \in \mathcal{L}$ has a non-empty finite set of names, called level descriptions *LD*(*l*);

(d) Each level description is associated with a base type, called its domain;

(e) $\rightarrow$ is a partial order (*rollup*) relation on the levels in $\mathcal{L}$.

(f) The closure of rollup has a unique bottom level and a unique upper level called *all* such that *LD(all)={all}*

Based on this, a spatial dimension schema is a dimension schema $\langle$*name$\mathcal{DS}$, $\mathcal{L}$*, $\rightarrow\rangle$ where at least one level $l \in \mathcal{L}$ has exactly one level description with domain of type geometry.

b) Cube Schema and Operations over Fields

A cube schema is a tuple $\langle$*nameCS, D, M*$\rangle$ where *nameCS* is a literal, *D* is a finite set of dimension schemas and *M* are also a finite set measures.

According to the definition of discretized fields, the value of the field at that point is the measure and the coordinates of the point are the dimension. The fact is represented by the field and its schema $\langle$*cubeName, D, M*$\rangle$. Therefore, a field can be seen as an OLAP Cube. It is shown that DFields and traditional OLAP cubes can be easily integrated for data analysis.

They also define a set of aggregate functions $\mathcal{A}$ = {Max, Min, Avg, Sum}.

Dice operator. This operation selects values in dimensions or measures based on a Boolean condition σ which may introduce discontinuity. To avoid discontinuity, the operator is defined by setting the value of samples with ⊥ values where the σ operator is not satisfied.

ROLL-UP operator. It aggregates facts according to dimension hierarchy. In this model the spatial dimension is a single-level dimension which implies that rollup hierarchies must be introduced externally. Three roll-up operations are defined (spatial over spatial field, spatial over spatiotemporal field and temporal over spatiotemporal fields).

Drill-Down. This operations aggregates facts according to dimension hierarchy. In this model, the spatial dimension is a single-level dimension which implies that rollup hierarchies must be introduced externally. They define 3 roll-up operations (spatial over spatial field, spatial over spatiotemporal field and temporal over spatiotemporal fields).

Roll-Up. This operation reverses the effect of **Roll-Up** so it is just the inverse of the mentioned operator.

IV. DISCUSSION

The approaches used in the above listed proposals (and in others) are diverse to say the least. Moreover, it seems that there is no build up on previous work or any attempts to criticize or enhance what has already been done which leads to a plethora of models without a clear attempt to define a mainstream model.

To compare the 4 models, one needs to define criteria and search which model meets the necessary requirements to become an applicable model. According to [6], true data models should have 3 components: A *structure* that defines how data are structured, *integrity rules* that define how data are kept correct and consistent and *operators* which define how data are manipulated. Therefore we will compare these models using the above mentioned conditions in addition to the requirements defined by [4] and additional requirements defined in [16][17]. The different models are evaluated against these six requirements (Table 2):

1) Structure

The common structure between all models is the hypercube. Each model proposes a different way of building its hypercube.

2) Integrity rules

Multidimensional structures are not concerned with integrity as much as they are concerned with fast response to analytical queries.

3) Operators

This complexity of SOLAP queries implies long query processing time. Therefore, most queries are run in advance and the results are stored as materialized views. Operators are either for navigation along the analysis dimensions or for returning previously computed results.

4) Continuous data as measures

Usually measures are numerical values that are analyzed according to axis of analysis (the dimensions). In the models we reviewed only one proposal imposes continuous data as measures constraint.

5) Explicit hierarchies in dimensions

The hierarchy should be explicit to allow the user to navigate with clear knowledge of the relationship between the different levels [15].

6) Symmetric treatment of dimensions and measures:

Since measures can be a level of a dimension, it is essential that measures can be transformed into attributes and vice versa [5]. This will provide an important functionality to any OLAP system.

TABLE II. MULTIDIMENSIONAL MODEL COMPARISON CRITERIA

Criteria	Ahmed et al. [1]	Vaisman et al. [23]	Bimonte et al. [4]	Gómez et al. [9][10]
Structures	✓	✓	✓	✓
Integrity rules	x	x	x	x
Operators	✓	✓	✓	✓
Continuous data as measures	x	x	✓	x
Explicit hierarchies	✓	x	✓	x
Symmetric treatment of dimensions and measures	✓	x	x	x

To clarify some of the differences between the 4 models, we will present examples of how the aggregate *average* is handled by each model.

In [1], to find the average pollution for a specific region, the aggregation is done by applying interpolation functions to the basic cubes to obtain a continuous representation of the field. The sum of all values is calculated as in integral of the function representing the field. The average is then obtained by dividing the sum by the area of the field. In [23], the average monthly temperature for a land plot is calculated as follows :

{*l*.number, *m*.month, temp | LandPlot(*l*) ^ Month(*m*)^ first=min({*t*.date|Time(*t*) ^ *t*.month = *m*.id}^last= max({*t*.date|Time(t)^*t*.month=*m*.id}^ temp = avg ({atperiods (atregion(*t*.geometry,*l*.geometry), range(first, last)) | Temperature(*t*)})} Where *first* and *last* represent first and last day of the month.

 ISBN: 978-1-61208-486-2

In [4], the aggregation is performed by applying the average on the Field Hierarchy H_{regres}. $F_4(x;y)=AVG(leavesFieldSupport(H_{deptres}, (x_2;y_2))) = AVG(f_3(x;y), f_3(x_1;y_1))$.

For more details about the examples we refer the reader to [1][4][9][23].

From Table 2, none of the proposed models satisfies all criteria. Models proposed by Vaisman et al. [23] and Gómez et al. [9][10] satisfy only structure and operators constraints. The major weakness is the lack explicit hierarchies, which is essential for navigation in cubes. The model presented by Ahmed et al. [1] lacks treatment of continuous data as a measure which is supported in [4]. However, Bimonte et al. [4] does not treat dimensions and measures symmetrically.

V. CONCLUSION

Spatial data warehouses have been around for some time now. Most of the early work on this topic was oriented towards discrete spatial data. The combination of cartographic display and OLAP resulted in SOLAP. However it was also limited to discrete spatial representation. Attempts at integrating continuous or field based data in multidimensional structures began about 10 years ago. During this period, a number of models to represent spatial and/or spatiotemporal continuity were proposed. In this paper, we studied some of these models and compared them with respect to different criteria and conditions for multidimensional models. None of the different proposals covered all comparison criteria and hence there is still a considerable amount of work to be done on the subject. The other remark is that most of the work concentrated on the theoretical side without a mention of solid model that can be used in real life application. To the best of our knowledge, even though there is still an undiscovered wealth mine for research and development, there is a lack of recent work on this domain.

REFERENCES

[1] T. O. Ahmed and M. Miquel. “Multidimensional Structures Dedicated to Continuous Spatiotemporal Phenomena”. In proceedings of the 22nd British National Conference on Databases (BNCOD), (Sunderland, UK, July 5-7, 2005) 29-40.

[2] T. O. Ahmed. “Spatial On-line Analytical Processing (SOLAP): Overview and Current Trends”. In proceedings of Internationa Conference on Advanced Computer Theory and Engineering (ICACTE), (Phuket, Thailand, December, 20-22, 2008) 1095-1099.

[3] Y. Bédard, T. Merrett and J. Han. “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.

[4] S. Bimonte and M. A. Kang. “Towards a Model for the Multidimensional Analysis of Field Data”. In proceedings of the 14th East European Conference (ADBIS10) (Novi Sad, Serbia, Sep 20-24, 2010. 300-311.

[5] L. Cabibbo and R. Torlone. “From a procedural to a visual query language for OLAP”. In proceedings of the 10th International Conference on Scientific and Statistical Database Management (SSDBM’98), IEEE Computer Society Press, Washington, DC, USA pp. 74–83, 1998.

[6] E. F. Codd. “Data Models in Database Management”. ACM SIGMOD Record 11(2), 1981, pp. 112-114.

[7] R. Elmasri. and S. Navathe. Fundamentals of Database Systems. Pearson, 7th edition, 2015.

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

[9] L. Gómez, S. Gómez and A. Vaisman. “Analyzing Continuous Fields with OLAP Cubes”. In proceedings of the 14th Workshop DOLAP (DOLAP’11), (Glasgow, Scotland, Oct 24-28, 2011). pp. 89-94

[10] L. Gómez, S. Gómez and A. Vaisman. “A Generic Data Model and Query Language for Spatiotemporal OLAP Cube Analysis”. In proceedings of the 15th International Conference on Extending Database Technology (EDBT2012) (Berlin, Germany) Mar 27-30, 2012, pp. 300-311.

[11] J. Han, R. Altman, V. Kumar, H. Mannila and D. Pregibon. “Emerging Scientific Applications in Data Mining”. Communication of the ACM, 45,8 (Aug 2002), pp. 54-58.

[12] E. Malinowski and E. Zimányi. “Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications”. Data-Centric Systems and Applications, Springer, 2008.

[13] P. Marchand, A. Brisebois, Y. Bedard and G. Edwards. “Implementation and Evaluation of a Hypercube-Based Method for Spatiotemporal Exploration and Analysis”. ISPRS Journal of Photogrammetry and Remote Sensing. (Aug 2004) 59(1-2), pp. 6-20.

[14] F. Moreno, J. A. E. Arias and B. M. Losada. “A Conceptual Spatio-temporal Multidimensional Model”. Revista Ingenierías Universidad de Medellín. (2011) 9(17), pp. 175-184.

[15] T. B. Pederson. “Managing complex multidimensional data”. Business Intelligence: Lecture Notes in Business Information Processing. (2013) vol. 138. pp. 1-28.

[16] T. Pedersen and C. Jensen. “Multidimensional Data Modeling for Complex Data”. In proceedings of the 15th International Conference on Data Engineering (ICDE99)(Sydney, Australia) March 23-26 1999, pp. 336-345.

[17] G. Pestana and M. Mira da Silva. “Multidimensional Modeling based on Spatial, Temporal and Spatio-Temporal Stereotypes”. 5th ESRI International User Conference (ESRI) (San Diego, USA) Jul 25-29, 2005, pp. 5-15.

[18] E. Pourabbas and M. Rafanelli. “Characterization of Hierarchies and Some Operators in OLAP Environment.” In proceedings of the 2nd ACM International Workshop on Data Warehousing and OLAP. (Kansas City, USA). Nov 2-6, 1999, pp. 54 - 59.

[19] S. Rivest, Y. Bedard and P. Marchand. “Towards better Support for Spatial Decision Making: Defining the Characteristics of Spatial On-Line Analytical Processing (SOLAP)”. Journal of Canadian Institute of Geomatics, 55(4), 2001, pp. 539-555.

[20] S. Rivest, Y. Bedard, M. J. Proulx and M. Nadeau. “SOLAP: A New Type of User Interface to Support Spatiotemporal Multidimensional Data Exploration and Analysis”. In proceedings of ISPRS workshop on Spatial, Temporal and Multi-Dimensional Data Modeling and Analysis, (Québec, Canada) Oct 2-3, 2003.

[21] M. Sampaio, A. Sousa and C. Baptista. “Towards a Logical Multidimensional Model for Spatial Data Warehousing and OLAP”. In proceedings of 9th ACM International Workshop on Data Warehousing and OLAP (New York, USA) Nov 5-11, 2006, pp. 83-90.

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

[23] A. A. Vaisman and E. Zimányi. "A Multidimensional Model Representing Continuous Fields in Spatial Data Warehouses". In proceedings of the International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2009)(Seattle, USA) Nov 4-6, 2009, pp. 168-177.

[24] M. Zaamoune, S. Bimonte, F. Pinet and P. Beaune. "A New Relational Spatial OLAP Approach for Multi-resolution and Spatio-multidimensional Analysis of Incomplete Field Data". In proceedings of International Conference on Enterprise Information Systems (ICEIS 2013)(Angers, France) 4-7 Jul 2013, pp. 145 -152.