

Proposition pour l'intégration de l'analyse spatiale et de l'analyse multidimensionnelle

Sandro Bimonte*, Anne Tchounikine**, Maryvonne Miquel**, Robert Laurini**

LIRIS (Laboratoire d'InfoRmatique en Image et Systèmes d'information), INSA-Lyon
Bâtiment Blaise Pascal, 7 Avenue Jean Capelle,
69621 Villeurbanne CEDEX, France

sandro.bimonte@imag.fr*
prenom.nom@insa-lyon.fr**

Résumé. L'introduction de l'information spatiale dans les modèles multidimensionnels a donné naissance au concept de Spatial OLAP (SOLAP). Dans cet article, nous montrons en quoi les spécificités de l'information géographique et de l'analyse spatiale ne sont pas entièrement prises en compte dans l'analyse et les modèles multidimensionnels SOLAP. Pour pallier ces limites, nous proposons le concept de dimension géographique et décrivons les différents types de hiérarchies associées. Nous proposons l'introduction de nouveaux opérateurs qui permettent d'adapter les opérateurs d'analyse spatiale au paradigme multidimensionnel. Enfin, nous présentons notre prototype qui offre une interface web de navigation spatiale et multidimensionnelle, et permet l'intégration de ces nouveaux concepts.

1 Introduction

Un entrepôt de données est une structure informatique dans laquelle est centralisé un volume important de données. L'organisation des données permet l'analyse multidimensionnelle qui consiste à explorer tout ou partie des données à un niveau détaillé et/ou agrégé grâce aux outils OLAP (On-Line Analytical Processing). L'information géo-référencée, souvent contenue dans les données, est intégrée sous forme textuelle dans les modèles multidimensionnels classiques. Des modèles plus récents, de type "OLAP Spatial" (SOLAP), visent à intégrer la donnée spatiale dans l'OLAP et à enrichir les systèmes OLAP classiques grâce, par exemple, à la visualisation cartographique, permettant ainsi d'expliciter la distribution géographique d'une information et/ou de mettre en relation des informations à diverses granularités géographiques.

L'information géographique est la représentation d'objets, ou de phénomènes, localisés dans l'espace ; les modèles SOLAP se concentrent généralement sur cette composante spatiale de l'information géographique. Or celle-ci est aussi caractérisée par un ensemble d'aspects sémantiques (ses attributs descriptifs alphanumériques et ses relations avec d'autres objets) qui sont pertinents à la fois dans la modélisation de la donnée géographique multidimensionnelle et de ses hiérarchies, et aussi dans l'analyse. L'analyse spatiale offerte par les Systèmes d'Information Géographique est par nature flexible et itérative : les données géographiques peuvent être modifiées ou remplacées grâce aux méthodes de transformations spatiales tout au long du processus d'analyse. Or les opérateurs spatiaux fournis par les différents systèmes SOLAP sont souvent des opérateurs orthogonaux aux opérateurs

multidimensionnels, autrement dit, ils ne sont pas utilisés dans la navigation OLAP, ce qui contribue à figer les scénarii de navigation.

Dans nos travaux, nous proposons un modèle "OLAP Géographique" qui reformule les concepts du Spatial OLAP pour prendre en compte la composante spatiale et sémantique de l'information géographique et le caractère itératif de l'analyse spatiale. L'OLAP Géographique définit les concepts de mesure, de dimension et de hiérarchie géographiques, ainsi que de nouveaux opérateurs spatio-multidimensionnels. Nous montrons en quoi l'OLAP géographique peut permettre d'intégrer analyse spatiale et analyse multidimensionnelle. Nous nous focaliserons dans cet article sur les concepts de dimensions géographiques et les opérateurs associés. En utilisant des données environnementales de la lagune de Venise, nous présentons notre système OLAP Géographique, GeWolap, qui offre une interface web de navigation spatio-multidimensionnelle.

Cet article est organisé de la façon suivante : la section 2 présente les caractéristiques majeures de l'analyse multidimensionnelle et de l'analyse spatiale. Un état de l'art critique des modèles et des systèmes SOLAP est décrit en section 3. La section 4 explicite le concept de dimension géographique, les opérateurs multidimensionnels associés et fournit une brève présentation du modèle formel GeoCube. La description de notre prototype est donnée en section 5.

2 Analyse multidimensionnelle et analyse spatiale

Les Systèmes d'Aide à la Décision (SAD) sont définis comme des systèmes d'information flexibles et interactifs qui aident les décideurs dans l'extraction d'informations utiles pour identifier et résoudre des problèmes et pour prendre des décisions (Alter, 1980). Cette connaissance est obtenue à partir de données brutes, de connaissances personnelles, et de modèles analytiques. Les SAD présentent ces informations, provenant de différentes sources, dans un environnement unique, uniforme et familier à l'utilisateur. Ils fournissent un environnement qui combine, uniformise et synchronise les bases de données, les modèles d'analyse et les techniques de visualisation. Ils permettent de comparer différents résultats, de concevoir et de valider des hypothèses.

2.1 Entrepôts de données, OLAP et analyse multidimensionnelle

Les entrepôts de données associés à des outils d'analyse OLAP représentent une solution effective pour l'informatique décisionnelle (Inmon, 1996). Les données sont organisées dans des hypercubes en axes d'analyse appelés "dimensions" et en sujets d'analyse appelés "faits". Les faits sont caractérisés par des mesures pré-calculées à l'aide de fonctions d'agrégation selon les différentes granularités définies par le schéma hiérarchique de chaque dimension. Dans le cas classique, une mesure est une valeur numérique qui décrit quantitativement le fait, par exemple le montant et/ou le volume d'une vente. Le processus d'analyse multidimensionnelle consiste en l'exploration de l'hypercube. L'utilisateur parcourt les données de l'hypercube selon les différents axes d'analyse à la recherche d'informations utiles, dans un processus fortement interactif, qui comprend des étapes de formulation des hypothèses, expérimentation et analyse. Les utilisateurs interagissent avec les données multidimensionnelles pour formuler, modifier et valider leurs hypothèses. Les chemins d'analyse sont imprédictibles, contrairement aux données qui sont pré-définies (dimensions

et mesures détaillées) et/ou pré-calculées (mesures agrégées). Chaque résultat d'analyse est la conséquence des résultats précédents. Chaque étape du processus d'analyse est effectuée par une navigation dans l'hypercube, ou par une requête multidimensionnelle. Ces requêtes utilisent les opérateurs OLAP (Roll-up, Drill-down, Slice, etc...) permettant de calculer les mesures pour des ensembles de membres à des niveaux de granularité et selon des prédicats sélectionnés par l'utilisateur.

Les architectures des systèmes OLAP sont des architectures trois tiers. Le premier tiers est un serveur d'entrepôt de données. Les données, sélectionnées pour leur pertinence, sont extraites des bases de données transactionnelles, nettoyées, transformées puis intégrées dans l'entrepôt. Le deuxième tiers est un serveur OLAP qui, pour optimiser le temps de réponse, pré-calcule les différents agrégats en utilisant des fonctions d'agrégation classiques de l'algèbre relationnelle (SUM, MIN, MAX, AVG ou COUNT). Le dernier tiers est un client OLAP qui offre une interface utilisateur avec des outils de reporting, d'analyse interactive, et parfois de fouille de données. Le paradigme de visualisation le plus adopté par les clients OLAP est la table de pivot. Il s'agit d'un tableau multidimensionnel auquel sont associés les totaux et les sous-totaux, et qui offre une vue de données imbriquées sur plusieurs niveaux. Les tables de pivot sont généralement couplées avec des affichages graphiques (histogrammes, courbes, etc.). Les actions de l'utilisateur sur ces différentes composantes (clic de souris, sélection graphique, drag-and-drop) se traduisent par l'appel aux opérateurs OLAP.

L'uniformisation de données hétérogènes, la présentation de l'information à différents niveaux de détail et sur différents axes d'analyse, la rapidité d'accès aux données et l'interface interactive, simple et intuitive font des entrepôts de données associées aux outils OLAP de véritables SAD.

2.2 Systèmes d'Information Géographique et analyse spatiale

L'information géographique est la représentation d'un objet ou d'un phénomène réel localisé dans l'espace à un moment donné. L'information géographique est caractérisée par une composante purement spatiale et une composante sémantique (Degréne et Salgé, 1997). La composante spatiale représente la position sur la surface terrestre et la forme d'un objet du monde réel. Une position est décrite dans un système de référence explicite comme par exemple un système de coordonnées. Cette composante permet de représenter la forme de l'objet lui-même et de positionner celui-ci par rapport aux autres phénomènes ou objets du monde réel. La composante sémantique représente l'information relative à la nature, l'aspect et les propriétés descriptives d'un objet ou d'un phénomène du monde terrestre. Cette information peut aussi inclure des relations spatiales, descriptives et de généralisation cartographique avec d'autres objets ou phénomènes. En effet, un des aspects sémantiques qui distingue l'information géographique des données classiques est sa représentation multiple à différentes échelles ou selon différents thèmes secondaires (Weibel et Dutton, 2001). Les cartes généralisées sont obtenues grâce aux opérateurs de généralisation (Regnault et McMaster, 2007).

Les Systèmes d'Information Géographique (SIG) permettent de générer, mémoriser et visualiser les données géographiques. Les SIG offrent des fonctions d'analyse spatiale, qui les transforment en véritables outils pour l'analyse spatiale. Ainsi, les utilisateurs de SIG disposent d'un ensemble d'outils statistiques (par exemple des méthodes de Point Pattern Analysis) et non statistiques (par exemple le buffer), qui peuvent créer ou modifier les

données géographiques (Longley et al., 2001). Le processus d'analyse spatiale est itératif et comprend les étapes suivantes : (1) identification du problème et des buts de l'analyse, (2) identification des problématiques spatiales et des outils pour les résoudre, (3) identification des données et leur préparation pour les opérations spatiales, (4) création d'un plan d'analyse, (5) exécution du plan et visualisation des résultats (Mitchel, 2005). Chaque étape du processus peut être modifiée, itérée ou éliminée en fonction des résultats obtenus. En d'autres termes, les données, les sujets et les outils du processus d'analyse spatiale ne sont pas fixés a priori. Dans ce processus, la visualisation des résultats joue évidemment un rôle très important car elle stimule l'utilisateur dans son processus de découverte de patterns, relations et tendances. La «géo-visualisation» intègre ainsi les techniques de visualisation scientifique, de cartographie, d'analyse des images, d'exploration de données, pour fournir une théorie, des méthodes et des outils pour la représentation et la découverte de la connaissance spatiale (MacEachren et al., 2001).

3 L'OLAP Spatial

Yvan Bédard définit l'OLAP Spatial (SOLAP) comme «une plate-forme visuelle spécialement conçue pour supporter l'analyse et l'exploration spatio-temporelles rapides et faciles des données multidimensionnelles composées de plusieurs niveaux d'agrégation à l'aide d'affichages cartographiques aussi bien qu'à l'aide de tableaux et diagrammes statistiques» (Bédard, 1997). Le SOLAP se propose comme un SAD dans lequel les fonctionnalités OLAP sont associées à des fonctionnalités SIG et à des techniques de géo-visualisation (Rivet et al., 2005). La prise en compte de la composante spatiale améliore l'analyse OLAP classique, la représentation cartographique permettant de mettre en évidence des relations spatiales entre différents faits et/ou dimensions qu'une simple étiquette textuelle ou un affichage graphique n'aurait pas, ou mal, montrées. Les modèles SOLAP redéfinissent les concepts principaux de l'OLAP : mesure, dimension et opérateurs de navigation multidimensionnelle.

3.1 Concepts principaux

Le terme de dimension spatiale désigne la prise en compte de l'information spatiale dans un axe d'analyse d'une application décisionnelle. Plusieurs travaux s'intéressent à la modélisation et la prise en compte des dimensions spatiales, par exemples (Bédard, et al., 2001), (Fidalgo et al., 2004) et (Malinowski et Zimányi, 2005). L'introduction explicite de la composante spatiale dans les niveaux des hiérarchies de dimensions permet de bénéficier d'une représentation cartographique des membres de dimensions, d'utiliser de prédicats spatiaux dans les opérations de coupe, et de prendre en compte les relations topologiques pendant les processus d'agrégation (Jensen et al., 2004), (Malinowski et Zimányi, 2005).

L'introduction des données spatiales dans les dimensions d'entrepôts de données a mené différents auteurs à la définition d'opérateurs d'analyse spatio-multidimensionnelle. (Rivet et al., 2005), (Sampaio et al., 2006), (Matias et Moura-Pires, 2007), (Scotch et Parmanto, 2005), (Hernandez et al., 2005) reformulent les opérateurs de forage définissant les opérateurs "spatial-drill down" et "spatial roll-up" qui permettent de naviguer dans une dimension spatiale. (Sampaio, et al., 2006), (Colonnese et al., 2005) et (Matias et Moura-Pires, 2007) appellent "spatial slice" une opération de coupe qui utilise un prédicat spatial.

(Scotch et Parmanto, 2005) introduisent deux nouveaux opérateurs de coupe : le "buffer" et le "spatial drill-out". Le "buffer" utilise l'opérateur d'analyse spatiale de buffer, qui crée une zone tampon autour d'un membre, pour sélectionner des membres de la dimension spatiale. Le "spatial drill-out" sélectionne tous les membres adjacents au membre sur lequel cet opérateur est appliqué. Ces opérateurs spatio-multidimensionnels permettent de naviguer dans un hypercube spatial en utilisant les attributs géométriques des dimensions spatiales.

3.2 Les outils SOLAP

Un outil SOLAP repose sur l'intégration des fonctionnalités SIG et OLAP (Kouba, et al., 2000), (Rivest et al., 2005). Cette intégration nécessite une reformulation des trois tiers de l'architecture OLAP classique, en utilisant et/ou en ajoutant des fonctionnalités SIG. Le premier tiers est un entrepôt de données spatiales qui permet de modéliser les structures de données associées aux dimensions et aux mesures spatiales. Le deuxième tiers est un serveur OLAP qui implémente l'hypercube spatial. Il fournit les opérateurs de navigation spatio-multidimensionnelle de coupe et de forage, gère les dimensions spatiales, permet la définition des mesures spatiales et fournit un mécanisme pour calculer les vues matérialisées. Le client SOLAP combine et synchronise des affichages tabulaires, graphiques et cartographiques. Le résultat d'une requête multidimensionnelle doit être représenté sous forme de tableau, de graphiques et à travers la composante cartographique. Cette dernière montre les membres des dimensions spatiales sur lesquels sont affichées les mesures numériques grâce à divers affichages graphiques (i.e., histogrammes, camemberts, etc.). Les cartes, dans un outil SOLAP, doivent rendre visible l'information et permettre d'explorer et d'analyser la base de données spatio-multidimensionnelle sans connaître le langage de requêtes et/ou la structuration des données. Par conséquent, la composante cartographique offre des cartes "sensibles" où une simple interaction génère des actions d'exploration et où la visualisation s'adapte aux manipulations de l'utilisateur. Ainsi, la visualisation est une des caractéristiques la plus importante et en même temps une des plus problématiques d'un outil de navigation spatio-multidimensionnel. L'interface utilisateur doit supporter plusieurs dimensions spatiales, des cartes qui mixent différents types spatiaux (point, ligne, polygone) et doit permettre de naviguer entre les niveaux textuels et cartographiques. De plus, elle doit supporter la synchronisation, en d'autres termes une action (par exemple un drill-down ou un roll-up) sur une composante (par exemple une carte) doit être propagée aux autres composantes (table de pivot et graphiques). La synchronisation est une tâche compliquée car ces composantes ont toutes des pouvoirs d'expression différents.

Plusieurs outils qualifiés de SOLAP ont été proposés. CommonGis (Hernandez et al., 2005) est un logiciel de géo-visualisation étendu pour supporter les bases de données multidimensionnelles. Il fournit des fonctionnalités d'analyse multicritère, d'analyse spatiale et des techniques visuelles d'analyse spatio-temporelle en utilisant les séries temporelles. CommonGis a été adapté pour l'analyse des données multidimensionnelles spatiales, où l'information spatiale est incluse en tant qu'axe d'analyse. L'interface est flexible et interactive, et possède une vaste collection de techniques avancées de géo-visualisation et autorise les opérations SOLAP de forage et de coupe. En revanche, les fonctionnalités OLAP sont limitées car CommonGis n'utilise pas de serveur OLAP.

L'intégration d'une composante cartographique statistique ou interactive à un outil OLAP classique est à la base de nombreux travaux. Le système Polaris (Stolte et al., 2003) permet la visualisation des mesures alphanumériques à travers des cartes thématiques non

interactives, englobées dans les cellules de la table de pivot. PostGeOLAP (Colonnese et al., 2005) est un système SOLAP qui gère les mesures numériques et les dimensions spatiales. PostGeOLAP fournit des méthodes pour créer la base de données spatiales multidimensionnelles et les vues matérialisées. Plusieurs solutions qui utilisent les cartes interactives ont été développées, dont (Silva et al., 2006) et (Sampaio et al., 2006). (Silva et al., 2006) présentent un outil web SOLAP dont la caractéristique principale est l'usage des services web géographiques pour la définition de GeoMDQL, un langage de requêtes qui étend le langage MDX de Microsoft pour les entrepôts de données spatiales. Le prototype est basé sur le serveur OLAP Mondrian modifié pour gérer les requêtes GeoMDQL et le client OLAP JPivot couplé avec une carte interactive pour la représentation des dimensions spatiales. L'outil utilise la modélisation logique présentée en (Fidalgo et al., 2004) et ne gère pas les mesures spatiales. Dans (Sampaio et al., 2006), les auteurs décrivent un système web SOLAP qui permet d'interroger les entrepôts de données spatiales avec des opérateurs de forage et de sélection sur la dimension spatiale et gérer les mesures spatiales. Dans cette solution, l'interface web est composée d'une interface cartographique, un navigateur pour sélectionner les membres de dimensions, et une zone de texte pour éditer les requêtes multidimensionnelles. Dans (Rivest et al., 2005) les auteurs décrivent un outil SOLAP qui fournit de nombreux types de visualisation cartographique tabulaire et graphique accessibles par plusieurs fenêtres et des fonctionnalités SIG et OLAP. Les composantes visuelles sont synchronisées pour former une interface unique, flexible et interactive. L'outil gère les mesures classiques et les mesures spatiales définies comme l'ensemble de toutes les géométries représentant les objets spatiaux correspondant à une combinaison particulière des membres spatiaux. De nouvelles fonctionnalités ont été récemment introduites, comme par exemple l'association des documents hypermédia aux éléments multidimensionnels spatiaux et/ou non spatiaux (Bédard et al. 2006). SOVAT (Scotch et Parmanto, 2005) est un outil OLAP-SIG intégré qui combine des fonctionnalités OLAP, SIG et de fouille de données classiques et spatiales comme par exemple le clustering. Il fournit des méthodes de statistiques et d'analyse spatiale. L'outil synchronise un ensemble de cartes interactives, et d'affichages graphiques et tabulaires. SOVAT gère les dimensions spatiales et les mesures spatiales comme de collections de géométrie. Le prototype SOLAP présenté dans (Matias et Moura-Pires, 2007) se distingue par la possibilité de couper l'hypercube en utilisant des relations spatiales et par la possibilité de superposer deux hypercubes en utilisant leurs dimensions spatiales. Cet outil intègre de fonctionnalités SIG de base aux fonctionnalités purement OLAP.

3.3 Limites des solutions SOLAP

3.3.1 Dimension spatiale

Les modèles proposés en littérature (Bédard et al. 2001), (Fidalgo et al. 2004) (Malinowski et Zimányi, 2005) qui introduisent l'information géographique comme axe d'analyse, sont caractérisés par la présence de l'attribut géométrique dans les membres des différents niveaux. Les hiérarchies spatiales sont définies en utilisant les attributs spatiaux ou alphanumériques des dimensions. Ces relations définissent, comme pour les hiérarchies classiques, une relation d'inclusion entre les membres de niveaux différents. Les hiérarchies spatiales sont donc des hiérarchies classiques de classification ou de spécialisation (Lujian-Mora et al. 2002) avec des attributs géométriques. Ainsi, entre deux membres spatiaux de

deux niveaux différents, il existe toujours une relation topologique d'inclusion ou intersection (Malinowsky et Zimányi, 2005). Ces hiérarchies représentent différentes granularités de l'information géographique, chaque niveau géographique représentant une information géographique différente.

Les dimensions spatiales, basées exclusivement sur les relations spatiales, ne reflètent pas la sémantique des relations entre les membres de différents niveaux. Par exemple, un des aspects sémantiques qui caractérise l'information géographique est sa représentation à différentes échelles ou selon différents thèmes secondaires. Or, les relations hiérarchiques de généralisation cartographique ne représentent pas toujours de relations d'intersection ou inclusion. Grâce à ce type de hiérarchie, le décideur pourrait, par exemple, naviguer à travers différentes représentations à diverses échelles de la même information géographique, et visualiser les mesures à différents degrés de précision spatiale ; les méthodes d'agrégation devraient alors être différentes de celles utilisées pour les hiérarchies spatiales classiques car il n'est pas possible de quantifier l'apport d'un membre par rapport à son ancêtre dans le cas d'une hiérarchie de généralisation.

Selon nous, la prise en compte de la sémantique des relations entre les différents membres des dimensions spatiales est fondamentale pour le processus décisionnel, car elle peut être utilisée pour caractériser les processus d'agrégation, comme nous le montrons dans la section suivante.

3.3.2 Opérateurs multidimensionnels

Dans le processus d'analyse multidimensionnelle, l'utilisateur navigue dans l'hypercube à travers les hiérarchies des dimensions, en comparant les mesures, qui sont agrégées à différentes granularités avec des fonctions d'agrégation. Les chemins d'analyse sont donc imprédictibles, mais le contexte d'analyse, i.e. les données et le modèle multidimensionnel, est défini dès la phase de conception. Le processus d'analyse spatiale, lui, est un processus itératif dans lequel les données sont modifiées ou remplacées à chaque itération grâce aux méthodes d'analyse spatiale de transformation (i.e., le buffer, l'overlay, etc.). Ainsi, il apparaît souhaitable d'introduire et d'adapter les opérateurs d'analyse spatiale au contexte OLAP afin de profiter pleinement du caractère itératif de l'analyse spatiale. Ainsi les dimensions qui incluent l'information géographique doivent pouvoir être reformulées selon les exigences de l'utilisateur et les mesures doivent alors être recalculées.

Dans la majorité des cas, les outils SOLAP offrent d'une part les fonctionnalités OLAP de navigation, et d'autre part les fonctionnalités SIG de visualisation et d'analyse spatiale. Les cartes interactives sont couplées avec un ensemble d'outils d'analyse spatiale (i.e. ajout des cartes en format matriciel ou vectoriel, personnalisation des affichages, opérateurs métriques, requêtes spatiales, opérateurs d'analyse spatiale ou de fouille de données, etc.) qui ne sont pas intégrées au processus de navigation multidimensionnel.

En d'autres termes, ces outils ne modifient pas la structure de l'hypercube ce qui limite le caractère itératif du processus d'analyse.

4 Un cadre conceptuel pour l'OLAP géographique

Dans cette section, nous introduisons les concepts de dimension géographique et les opérateurs multidimensionnels associés ainsi que les principales définitions de notre modèle formel multidimensionnel.

Pour présenter nos contributions, nous utiliserons comme cas d'étude un projet concernant l'étude de la pollution de la lagune de Venise réalisé dans le cadre d'une collaboration avec l'organisation internationale CORILA (Consorzio per la Gestione del Centro di Coordinamento delle Attività di Ricerca inerenti il Sistema Lagunare di Venezia), qui a comme but la sauvegarde environnementale, architecturale et économique de la lagune de Venise.

4.1 Dimension Géographique

Nous appelons objet complexe une entité du monde réel (un patient, un produit, etc.) décrite par un ensemble d'attributs descriptifs alphanumériques (âge, nom, type, etc.). Un objet géographique est un objet complexe (une ville, un bâtiment, etc.) qui présente un attribut spatial (i.e. une géométrie) en plus de ses attributs descriptifs. Un exemple est une unité environnementale de la lagune de Venise décrite par les attributs suivants : son nom ("Murano", "Chioggia", etc.), la liste de plantes qui la recouvrent ("Spartina Maritima", etc.), son type (industrielle, agricole, etc.), un index de salinité (valeur numérique), un attribut spatial décrivant sa géométrie et sa surface.

Définition 1.

Une dimension est dite géographique si les membres d'au moins un niveau sont des objets géographiques.

Une dimension géographique structure l'information à différentes granularités représentées par ses niveaux. Les membres de ces niveaux peuvent être liés par des relations spatiales, et/ou des relations de généralisation cartographique et/ou des relations descriptives. Une dimension géographique peut présenter une ou plusieurs de ces hiérarchies, l'opération d'agrégation appliquée à la mesure dépend alors de la hiérarchie sur laquelle on navigue. On peut ainsi imaginer appliquer une opération additive lorsqu'on navigue sur une hiérarchie spatiale, et aucune agrégation sur une hiérarchie de généralisation cartographique. La modélisation des hiérarchies non-strictes, non-couvrantes et non-onto est une de caractéristiques avancées des modèles multidimensionnels classiques. Une hiérarchie est non-couvrante si une branche de la hiérarchie peut "sauter" un ou plusieurs niveaux, non-onto si les feuilles ne sont pas toutes au même niveau, et non-strictes si une relation n-n peut exister entre les membres de différents niveaux. Notre modèle doit permettre de modéliser ces hiérarchies complexes afin de prendre en compte les caractéristiques spécifiques des relations entre objets géographiques. Nous introduisons maintenant les différents types de hiérarchies applicables à une dimension géographique.

4.1.1 Hiérarchie descriptive

Une hiérarchie descriptive organise l'information géographique à différentes granularités thématiques.

Définition 2.

Une hiérarchie descriptive d'une dimension géographique est une hiérarchie OLAP de classification ou de spécialisation. Elle est définie en utilisant les attributs descriptifs des objets géographiques.

Un exemple de hiérarchie descriptive pour la lagune de Venise est présenté dans la Figure 1. Cette hiérarchie classe les unités ("Unit") de la lagune par rapport à leur type ("Type").

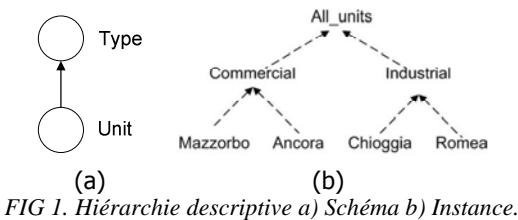


FIG 1. Hiérarchie descriptive a) Schéma b) Instance.

Un exemple d'application multidimensionnelle utilisant cette hiérarchie descriptive est présentée dans la Figure 2 selon le formalisme introduit par (Malinowsky et Zimányi, 2005).

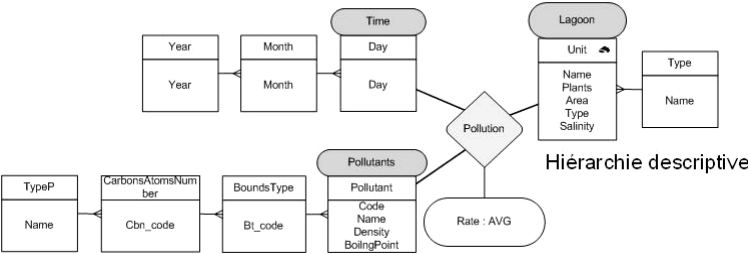


FIG 2. Modèle multidimensionnel comportant une dimension géographique représentant la lagune de Venise décrite par une hiérarchie descriptive.

Les dimensions sont le temps, les polluants et une dimension géographique avec la hiérarchie descriptive. Notons que certains niveaux de la dimension géographique peuvent être alphanumériques (eg. le niveau "Type"). Cette application permet d'analyser la pollution des eaux de la lagune en fonction du temps, des polluants et des unités de la lagune. La valeur de pollution est agrégée en utilisant la moyenne. Une requête multidimensionnelle peut être : "Quelle est la valeur moyenne de la pollution pour les unités de type industriel pour chaque année et pour les polluants de type organique ? ". Ce modèle multidimensionnel utilise la moyenne comme fonction d'agrégation sans contraintes particulières.

4.1.2 Hiérarchie spatiale

Définition 3.

Une hiérarchie spatiale d'une dimension géographique est une hiérarchie où les membres de différents niveaux sont liés par des relations topologiques d'inclusion et/ou d'intersection.

Une hiérarchie spatiale peut éventuellement être calculée grâce à l'attribut géométrique des membres de la dimension. Si les mesures peuvent être redistribuées sur la surface des

membres, les relations topologiques qui caractérisent cette hiérarchie peuvent permettre de quantifier l'apport d'un membre par rapport à son ancêtre dans le calcul de l'agrégation. Un exemple de schéma et d'instance d'une hiérarchie spatiale est montré dans Figure 3 où plusieurs unités de la lagune de Venise "Unit" sont regroupées de façon topologique en régions "Zone". Une relation d'inclusion existe entre les unités et les zones. Le calcul de la mesure, ici la moyenne de la valeur de pollution, peut prendre en compte ces relations topologiques.

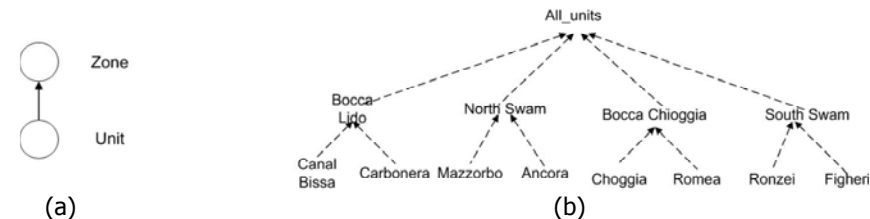


FIG 3. Hiérarchie spatiale a) Schéma b) Instance.

Un exemple utilisant cette hiérarchie est présenté dans la Figure 4. Cette application permet de répondre aux requêtes telles que : "Quelle est la valeur moyenne de la pollution pour chaque unité de la lagune en septembre 2007 et pour chaque polluant ?".

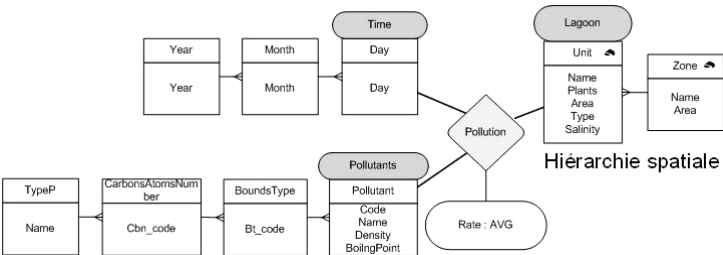


FIG 4. Modèle multidimensionnel comportant une dimension géographique représentant la lagune de Venise décrite par une hiérarchie spatiale.

4.1.3 Hiérarchie de généralisation cartographique

Une hiérarchie de généralisation cartographique représente une information géographique à différentes échelles ou selon différents thèmes secondaires.

Définition 4.

Une hiérarchie de généralisation cartographique d'une dimension géographique est une hiérarchie où les membres des niveaux représentent l'information géographique à différentes échelles/ ou selon différents thèmes secondaires et dont les membres d'un niveau sont les résultats de la généralisation des membres du niveau directement inférieur.

Dans la même application multidimensionnelle que celle de la Figure 4, nous substituons à la hiérarchie spatiale une hiérarchie de généralisation cartographique qui représente les unités de lagune de Venise à deux différentes échelles 1:1000 et 1:500 (Figure 5). Cette

application multidimensionnelle permet de répondre aux questions telles que : "Quelle est la pollution moyenne par polluant, année et par unité de la lagune à l'échelle 1:500 ?"

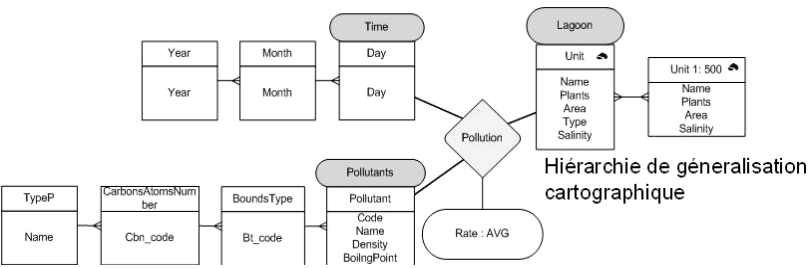


FIG 5. Modèle multidimensionnel comportant une dimension géographique représentant la lagune de Venise décrite par une hiérarchie de généralisation cartographique.

La Figure 6 montre un exemple de cette hiérarchie et les deux cartes associées. La carte généralisée est obtenue en utilisant les opérateurs de généralisation, d'agrégation, de simplification et de sélection (Regnaud et McMaster, 2007) : la forme de "Palude Maggiore" est simplifiée, "Campo" et "Ruzolo" sont fusionnées dans une seule grande zone dont la forme est simplifiée et "Colpo" est éliminée ("Colpo" est associé à la racine de la hiérarchie). Notons que cette relation est multi-valuée.

La représentation des mesures à travers des cartes à différentes échelles ou selon différents thèmes secondaires permet à l'utilisateur d'avoir un aperçu visuel global et simplifié du phénomène, en excluant les informations n'étant pas primordiales pour la compréhension de ses caractéristiques générales. Le processus d'agrégation pour les hiérarchies de généralisation cartographique doit tenir compte des relations de multi-association (Spaccapietra et al., 2007) résultant des opérateurs de généralisation, et ainsi être effectué avec un degré d'imprécision.

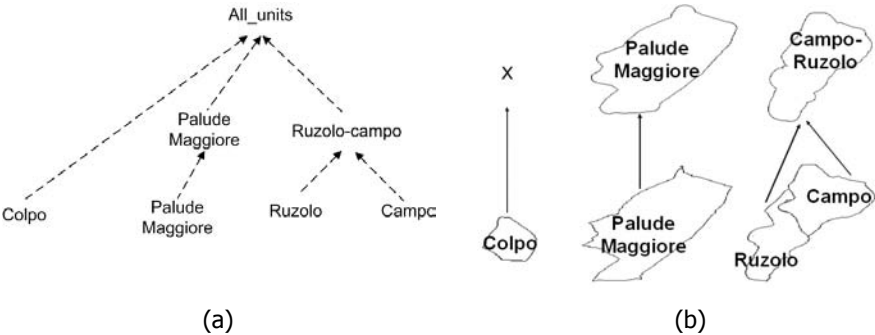


FIG 6. Hiérarchie de généralisation a) Instance b) Représentation cartographique

4.2 Opérateurs spatio-multidimensionnels portant sur la dimension géographique

Nous classifions les opérateurs spatio-multidimensionnels qui s'appliquent aux dimensions géographiques en trois catégories :

- *Les opérateurs de forage* permettent la navigation dans les dimensions géographiques. Dans le cas où plusieurs hiérarchies co-existent, ils doivent permettre de préciser la hiérarchie de navigation utilisée.
- *Les opérateurs de coupe* permettent de sélectionner une partie de l'hypercube en utilisant l'interaction avec la carte et/ou des relations topologiques, métriques et/ou directionnelles entre les membres géographiques.
- *Les opérateurs de modification dynamique* de l'hypercube permettent à l'utilisateur de créer de nouveaux membres géographiques à la volée grâce à des opérateurs d'analyse spatiale.

Les opérateurs de modification dynamique de l'hypercube permettent d'introduire de la dynamique dans la structure de l'hypercube. Au contraire des opérateurs de forage et de coupe, ils représentent une nouvelle approche dans l'analyse spatio-multidimensionnelle. En effet, comme nous l'avons montré dans la section précédente, les opérateurs d'analyse spatiale des solutions SOLAP gardent une définition "orientée" SIG. Or, fournir une vision multidimensionnelle de ces opérateurs est indispensable pour que les processus d'analyse multidimensionnelle et d'analyse spatiale s'enrichissent l'un l'autre. L'adaptation des opérateurs d'analyse spatiale au modèle multidimensionnel implique que les dimensions géographiques et les mesures associées puissent être calculées à la volée. La connaissance du phénomène étudié est bien sûr fondamentale pour définir le mode de calcul des mesures après insertion et/ou transformation d'un membre dans la dimension géographique.

Nous détaillons ici un de ces opérateurs. L'overlay est un opérateur d'analyse spatiale de transformation qui permet de mettre en relation des informations de nature différente. A partir de deux cartes, il génère une carte dans laquelle les géométries des objets géographiques sont recalculées grâce à l'opération topologique d'intersection. L'ensemble des objets géographiques de la carte résultat dépend de l'opérateur logique utilisé (AND ou OR). Un exemple d'overlay qui utilise l'opérateur logique AND est montré dans la Figure 7.

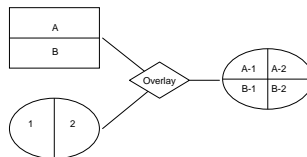


FIG 7. Overlay utilisant l'opérateur logique AND.

Dans un contexte spatio-multidimensionnel, il est possible d'appliquer l'overlay entre la carte qui représente un niveau d'une dimension géographique et une autre couche choisie par l'utilisateur. Le résultat de cette opération crée de nouveaux membres de la dimension géographique. Pour ces nouveaux membres, les attributs des mesures doivent être recalculés

en utilisant les parties de membres issues de l'opération d'overlay, les membres originaux des deux couches et les valeurs des attributs des mesures dans la table de faits. De la même façon, les opérateurs spatiaux de buffer, clipping, etc peuvent être utilisés pour modifier la structure et le contenu (i.e. membres et mesures) de l'hypercube.

4.3 Le modèle GeoCube

Les concepts de l'OLAP Géographique sont formalisés dans le modèle conceptuel GeoCube (Bimonte et al., 2005), (Bimonte, 2007). GeoCube prend en compte les composantes spatiale et sémantique de l'information géographique en dimension et en mesure et introduit la dynamique de l'analyse spatiale dans le processus d'analyse multidimensionnelle.

Ainsi, parallèlement aux concepts de dimension géographique et des opérateurs multidimensionnels géographiques associés, l'OLAP Géographique définit le concept de mesure géographique. Dimension et mesure sont traitées de façon symétrique, ce qui permet d'étendre le type de requêtes effectuées par l'utilisateur. Le concept de mesure géographique étant partie intégrante du modèle et de son algèbre, nous en synthétisons ci-après les grandes lignes. Le lecteur pourra trouver une présentation plus détaillée dans (Bimonte et al., 2006).

La mesure géographique est un objet géographique, décrit par un ensemble d'attributs descriptifs et un attribut spatial, et appartenant à une ou plusieurs hiérarchies géographiques, ceci dans une définition parfaitement symétrique avec celle des dimensions. L'information géographique peut donc être utilisée en dimension comme en mesure. L'agrégation d'une mesure géographique (lors d'une opération de forage) consiste en l'agrégation différenciée de ses attributs : par exemple, sur une mesure "unit", on appliquera la fusion des géométries, aucune agrégation aux noms, une liste pour les plantes, la moyenne pour la salinité et, pour le type, une moyenne pondérée par la surface. Ces attributs descriptifs pourront être utiles au processus décisionnel, par exemple un index de salinité très élevé peut se révéler toujours associé à une forte pollution en fer. D'autre part, la mesure géographique appartient à une structure hiérarchique et doit donc pouvoir être analysée à différentes granularités qui correspondent aux niveaux des hiérarchies. Ainsi, le passage d'une mesure détaillée à une mesure agrégée s'effectue lors d'une opération de forage ; le passage d'une mesure à une mesure de granularité différente s'effectue lors d'une opération de navigation dans la mesure. Cette vision complètement symétrique entre mesures et dimensions, associée à des opérateurs de navigation dans la mesure augmente encore la dynamique de l'analyse OLAP géographique.

Dans le reste de cette section nous présentons les concepts principaux du modèle de données et de l'algèbre de GeoCube.

4.3.1 Modèle de données

GeoCube représente toutes les données de l'application multidimensionnelle, membres de dimension et mesures, à travers les concepts de Schéma d'Entité S_e et d'Instance d'Entité $I(S_e)$. Un Schéma d'Entité définit la structure d'un objet de l'univers d'analyse par un ensemble d'attributs parmi lesquels, éventuellement, un attribut spatial, et de fonctions permettant de calculer des attributs dérivés. Par exemple, le schéma d'entité représentant les unités de la lagune est :

$S_{unit} = \langle \text{name_unit, geometry, salinity, list_of_plants, type, } \langle f_{area} \rangle \rangle$

Ce Schéma d'Entité décrit une unité de la lagune telle que définie paragraphe 4.1.

Les Entités sont organisées en hiérarchies grâce aux concepts de Schéma et d'Instance de Hiérarchie. Un niveau d'une hiérarchie est un Schéma d'Entité, et un membre est une Instance d'Entité. Le Schéma de Hiérarchie organise les niveaux d'une hiérarchie dans un treillis. Il est décrit par un tuple identifiant la liste des niveaux intermédiaires de la hiérarchie, le niveau le plus détaillé, le niveau le plus haut, et la relation d'ordre du treillis. Par exemple, $H_{\text{lagoon_spatial}}$ est un schéma de hiérarchie spatiale comportant 3 niveaux et décrit par :

$H_{\text{lagoon_spatial}} = \langle L_{\text{lagoon_spatial}}, S_{\text{unit}}, S_{\text{all_unit}}, \#_{\text{lagoon_spatial}} \rangle$ où :

- $L_{\text{lagoon_spatial}} = \{S_{\text{zone}}\}$, S_{unit} , $S_{\text{all_unit}}$ sont les niveaux de la hiérarchie,
- $\#_{\text{lagoon_spatial}}$ est la relation d'ordre du treillis telle que $(S_{\text{unit}} \#_{\text{lagoon_spatial}} S_{\text{zone}})$ et $(S_{\text{zone}} \#_{\text{lagoon_spatial}} S_{\text{all_unit}})$

Le Schéma du Cube de Base représente le schéma conceptuel de l'application multidimensionnelle et l'Instance de Cube de Base représente le cuboïde de base, i.e. une table de faits où toutes les dimensions sont aux niveaux les plus détaillés. Le Schéma du Cube de Base est défini par un ensemble de Schémas de Hiérarchie et une fonction booléenne. Cette fonction peut être représentée dans un espace multidimensionnel où les instances des membres des niveaux les plus détaillés des dimensions sont projetées sur les axes, et les points sont des valeurs booléennes représentant l'existence du fait. Grâce à cette modélisation de l'espace multidimensionnel, les Schémas de Hiérarchie utilisés dans la définition du Cube de Base jouent tous le rôle de dimensions. La distinction entre les Schémas jouant le rôle de dimensions et le Schéma jouant le rôle de la mesure se fait de façon dynamique, au moment de l'énoncé de la requête multidimensionnelle.

Le Cube de Base pour l'application de la figure 4 est $BC_{\text{corila}} = \langle H_{\text{pollutants}}, H_{\text{time}}, H_{\text{lagoon_spatial}}, H_{\text{rate}}, \delta \rangle$ où :

- $H_{\text{pollutants}}$ est la hiérarchie qui décrit les polluants,
- H_{time} est la hiérarchie temporelle,
- H_{rate} est la hiérarchie des intervalles de valeurs de pollution,
- $H_{\text{lagoon_spatial}}$ est la hiérarchie spatiale
- δ une fonction booléenne définie sur $I(S_{\text{pollutant}}) \times I(S_{\text{day}}) \times I(S_{\text{unit}}) \times I(S_{\text{rates}})$ soit le produit cartésien des instances des niveaux les plus détaillés des hiérarchies du Cube.

4.3.2 L'algèbre

Le Mode d'Agrégation permet de définir un ensemble de fonctions d'agrégations qui lient les valeurs des attributs des Instances d'Entité représentant les mesures détaillées et les valeurs des attributs des Instances d'Entité représentant les mesures agrégées. Il est décrit par un tuple identifiant le Schéma d'Entité de la mesure détaillée, le Schéma d'Entité de la mesure agrégée, et les fonctions d'agrégation utilisées.

Une Vue représente le résultat d'une requête multidimensionnelle. Le Schéma d'une Vue est un tuple qui identifie le Cube de Base utilisé, l'ensemble des niveaux de dimension de la vue, le Mode d'Agrégation, et une fonction booléenne. Contrairement au Cube de Base, une Vue identifie donc l'Entité utilisée comme mesure au travers du Mode d'Agrégation. Un exemple de vue qui représente les valeurs de pollution par mois, polluant et unité est $V_{\text{corila-rate-month}} = \langle BC_{\text{corila}}, \langle S_{\text{pollutant}}, S_{\text{month}}, S_{\text{unit}} \rangle, \Theta_{\text{avg}}, \gamma \rangle$ où :

- BC_{corila} est le Cube de Base,
- S_{unit} , $S_{pollutant}$, S_{month} sont les niveaux de la vue,
- Θ_{avg} le Mode d'Agrégation
- γ est une fonction booléenne qui joue le même rôle de la fonction booléenne du Cube de Base.

L'instance de $V_{corila-rate-month}$ est représentée dans le tableau 1.

$S_{pollutant.name_pollutant}$	$S_{month.month}$	$S_{unit.name_unit}$	$S_{rate.value}$
Zinc	9-05	Mazzorbo	2617
Sulfure Trioxide	9-05	Mazzorbo	2624
Méthane	9-05	Tessera	2618
Zinc	9-05	Canal Fondello	2619

TABLEAU 1. Instance de la Vue $V_{corila-rate-month}$

Il est aussi possible de définir, à partir du même Cube de Base, une Vue qui utilise les unités de la lagune comme mesures : $V_{corila-unit-day} = \langle BC_{corila}, \langle S_{pollutant}, S_{day}, S_{rate} \rangle, \Theta_{unit-fusion}, \gamma \rangle$ représentant les unités polluées par polluant, valeur de pollution et jour.

GeoCube propose un ensemble d'opérateurs qui s'appliquent aux dimensions, des opérateurs qui permettent de naviguer dans la hiérarchie de la mesure, et un opérateur qui inverse mesure et dimension. Tous les opérateurs s'appliquent à une Vue et donnent comme résultat une autre Vue. Soit $V_v = \langle BC_{bc}, L, \Theta_k, \gamma \rangle$ où L est un ensemble de Schémas d'Entité, alors $Op(V_v) [paramètres] = V'_v = \langle BC'_{bc}, L', \Theta'_k, \gamma' \rangle$ où γ' est calculée en utilisant l'algorithme de l'opérateur.

Les opérateurs définis dans GeoCube sont :

- Un opérateur de forage ρ qui permet de monter dans une hiérarchie et d'agréger les mesures. La syntaxe d'écriture est la suivante : $\rho(V_v)[S_e]$ où le paramètre S_e représente le nouveau niveau de la vue résultat.
- Un opérateur de coupe σ qui permet de sélectionner un sous-hypercube en filtrant les membres des dimensions, les mesures détaillées, ou les mesures agrégées selon un prédicat fourni en paramètre ($\sigma(V_v) [prédicat(S_e)]$).
- Des opérateurs qui modifient la structure de l'hypercube en ajoutant ou modifiant des membres géographiques grâce à des opérateurs d'analyse spatiale de buffer β et d'overlay ζ . La syntaxe de l'opérateur β est la suivante : $\beta(V_v) [H_h, f]$ où H_h est la hiérarchie contenant le membre géographique résultant de l'opérateur de buffer et f est une fonction qui calcule les mesures pour ce nouveau membre géographique. L'opérateur ζ utilise les mêmes paramètres.
- Deux opérateurs qui permettent de changer la granularité de la mesure. Ces opérateurs permettent de monter dans la hiérarchie de la mesure en remplaçant un ensemble de mesures détaillées par des mesures appartenant à un niveau moins détaillé. L'opérateur Classify, noté ω , nécessite que tous les descendants des nouvelles mesures soient présents dans la table de faits pour effectuer le remplacement. L'opérateur Specialize, noté Δ , nécessite un seul descendant. La

syntaxe est la suivante : $\omega(V_v) [S_i, \Theta_i]$ et $\Lambda(V_v) [S_i, \Theta_i]$ où S_i est le niveau de la mesure et Θ_i le Mode d'Agrégation associé.

- Un opérateur Π qui permute mesure et dimension. La syntaxe est : $\Pi(V_v) [S_i, \Theta_i]$ où S_i est le niveau le plus détaillé d'une dimension qui va être utilisé comme mesure et Θ_i est le Mode d'Agrégation associé.

5 GeWolap : un prototype pour l'OLAP Géographique

Dans cette section, nous décrivons notre prototype GeWolap. GeWolap est basé sur le modèle conceptuel GeoCube. GeWolap est un outil pour l'OLAP Géographique qui supporte les dimensions géographiques et les opérateurs multidimensionnels géographiques introduits dans la section 4.2, et présente une interface web interactive qui synchronise les différentes composantes cartographique et tabulaire. Il est développé sur une architecture trois tiers composée d'un SGBD spatial (Oracle), un serveur OLAP (Mondrian) et une interface web qui intègre un client SIG (MapInfo) et un client OLAP (JPivot).

5.1 Présentation générale de l'interface utilisateur

L'interface utilisateur (Figure 8) présente un environnement interactif qui encapsule la structure de l'application multidimensionnelle et traduit les interactions avec l'interface vers les opérateurs d'analyse et de navigation. L'interface est composée de deux panneaux principaux :

Le *Control Panel* est divisé en trois onglets : *Navigation* (liste les hypercubes disponibles), *Display/Export* (personnalisation de l'affichage et exportation vers différents formats) et *Analysis*. L'onglet *Analysis* contient un navigateur qui affiche les dimensions, les niveaux et les membres de l'hypercube. Il est utilisé pour sélectionner les dimensions visualisées sur les axes de la table de pivot et pour faire les opérations de coupe.

Le *Data Panel* contient la table de pivot et l'*OLAP toolbar* (Figure 8A), les graphiques, et la carte avec la *GIS Toolbar* et la *GeWolap Toolbar*. La *GIS Toolbar* (Figure 8B) propose des fonctionnalités purement SIG (zoom, pan, etc.). La *GeWolap Toolbar* (Figure 8C) présente les opérateurs tels que définis dans la section précédente, une fonctionnalité (*Properties Member*) qui affiche les propriétés d'un membre de la dimension géographique et une autre (*Measure displays*) qui permet de paramétrer le type de diagramme représentant les mesures.

5.2 Navigation avec GeWolap

Dans cette section, nous illustrons le processus de navigation dans l'hypercube de la Figure 4. Nous montrons les opérateurs de forage, les opérateurs de coupe et les opérateurs de modification dynamique de l'hypercube. La navigation dans un hypercube avec des mesures géographiques a été décrite en détail dans (Bimonte et al., 2006).

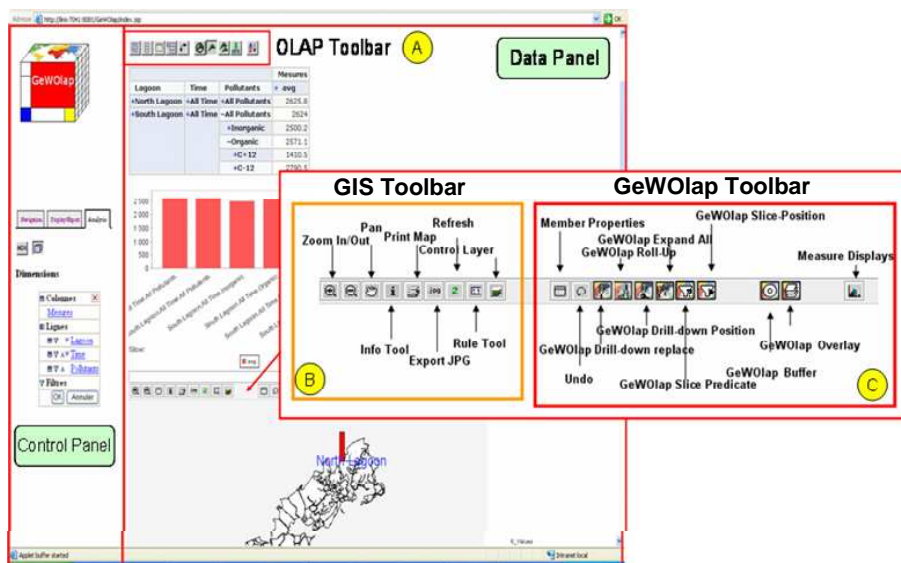


FIG 8. Interface GeWolap ; A) OLAP Toolbar, B)GIS Toolbar C) GeWolap Toolbar

Les opérateurs de forage (*GeWolap Roll-up*, *GeWolap Drill-down Position*, etc.) sont disponibles dans GeWolap à travers la table de pivot et la composante cartographique. Lors de ces opérations, GeWolap synchronise les composantes tabulaire et cartographique, en évitant que des membres cachent (recouvrent) leur fils sur la carte. On peut ainsi obtenir une carte dont les zones géographiques sont de grains différents. Un exemple d'utilisation d'un opérateur de forage est montré dans la Figure 9. La carte initiale (Figure 9.1) montre 3 mesures qui représentent la pollution moyenne pour tous les polluants, les polluants inorganiques et organiques sur la région sud. Après forage sur le membre géographique "North Lagoon", la carte de la Figure 9.2 présente la pollution moyenne sur des zones détaillées de la région nord de la lagune (North Swam, South Venice, North Venice, Bocca Lido et Litorale de Cavallino).

Cette opération réalise l'opération $p(V_{corila-rate})[S_{zone}] = V_{corila-rate-zone}$ où $V_{corila-rate}$ est la Vue qui représente la valeur de la pollution pour chaque unité, tous les polluants et tous les jours, et $V_{corila-rate-zone}$ représente la valeur de la pollution pour chaque zone, tous les polluants et tous les jours. Il est important de souligner le fait que puisque la table de pivot mélange plusieurs granularités d'une même dimension, sa représentation formelle correspond à un ensemble de Vues.

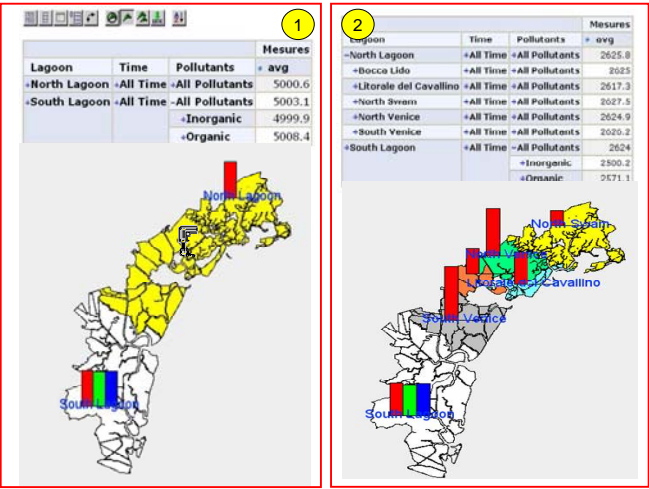


FIG 9. 1) Données initiales 2) Résultat avec l'opérateur de GeWolap-Drill Down Position sur la région "North Lagoon"

Les opérateurs de coupe *GeWolap-Slice Predicate* et *GeWolap-Slice Position* étendent les opérateurs de coupe et permettent de formuler une requête multidimensionnelle en sélectionnant un sous-ensemble des membres de la dimension géographique. Ceux-ci peuvent être spécifiés par pointage de la souris ou grâce à une requête utilisant un prédicat spatial (relations topologiques et/ou métriques) ou un prédicat "classique". Par exemple, il est possible de ne visualiser les mesures que pour les zones d'une surface supérieure à 1000 m². Formellement, cette opération de *GeWolap-Slice Predicate* correspond à l'opérateur $\sigma(V_{corila-rate})[S_{unit}, f_{area} > 1000]$.

Les opérateurs de modification dynamique de l'hypercube (*GeWolap Buffer* et *GeWolap Overlay*) créent de nouveaux membres de dimensions géographiques et permettent de recalculer les mesures pour ces nouveaux membres. *GeWolap Buffer* permet d'exprimer des requêtes sur des membres de la dimension géographique définis à la volée par l'utilisateur, grâce à la désignation d'une zone tampon. Pour ces nouveaux membres, les mesures doivent être recalculées grâce à une fonction fournie par l'utilisateur. Le calcul de la mesure prend en compte les parties de membres qui sont recouverts totalement et/ou partiellement par la région tampon, les membres originaux et les valeurs de mesures dans la table de faits associés. Supposons ainsi que la concentration des polluants est calculée comme une moyenne pondérée sur les surfaces. La Figure 10.2 montre la pollution moyenne pour "Bocca Chiogia", "Chioggia", "Litorale Pellestrino", "South Swam" et une zone tampon de 3 km autour de "Bocca Malmocco". Cette opération implémente l'opérateur $\beta(V_{corila-rate})[H_{lagoon_spatial_zoneb}, f_{avgbuffer}]$. Il est appliqué à $V_{corila-rate}$ en utilisant comme paramètres une hiérarchie qui contient le résultat de l'opération spatiale de buffer et la fonction pour le re-calcule de la mesure. On note que la modification issue de l'application de cet opérateur se répercute à la fois sur la représentation tabulaire et sur la cartographie.

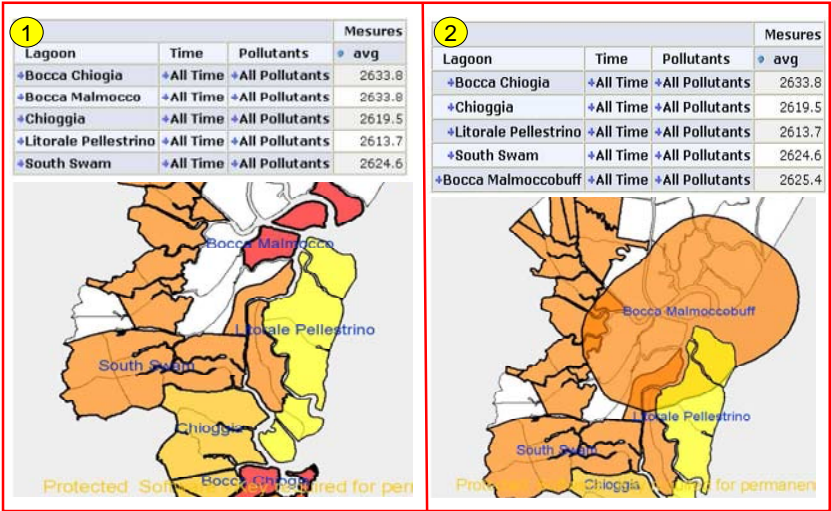


FIG 10. GeWolap Buffer 1) Données initiales 2) Résultat

L'opérateur *GeWolap Overlay* permet de faire des requêtes sur des membres calculés en utilisant une autre information géographique issue d'une autre couche et une fonction qui est fournie par l'utilisateur. Le calcul de la mesure prend en compte les parties de membres issus de l'opération d'overlay, les membres originaux des deux couches et les valeurs de mesures dans la table de faits associés. Supposons que l'utilisateur veuille ajouter aux informations de la pollution une autre source de données, par exemple une carte thématique de la lagune définissant les zones soumises à épuration. Supposons que la valeur de la nouvelle mesure peut être estimée comme une moyenne pondérée sur la surface des zones non épurées. La Figure 11.2 montre les valeurs modifiées de la pollution pour les zones de la lagune "North Venice" soumises à épuration. Nous notons que la pollution dans "Isola delle Tresse" avant l'overlay était de 2631.4 (Figure 11.1). GeWolap overlay nous permet de mettre en évidence une zone d'épuration ("Isola delle Tresse-ZoneRD") où la valeur de la pollution a pu être ré-estimée. Cette opération implémente l'opérateur $\zeta(V_{corilla-rate})[H_{lagoon_spatial_zoneov}, favgover]$. Il utilise comme paramètres une hiérarchie qui contient le résultat de l'opération spatiale d'overlay et la fonction pour le re-calcul de la mesure.

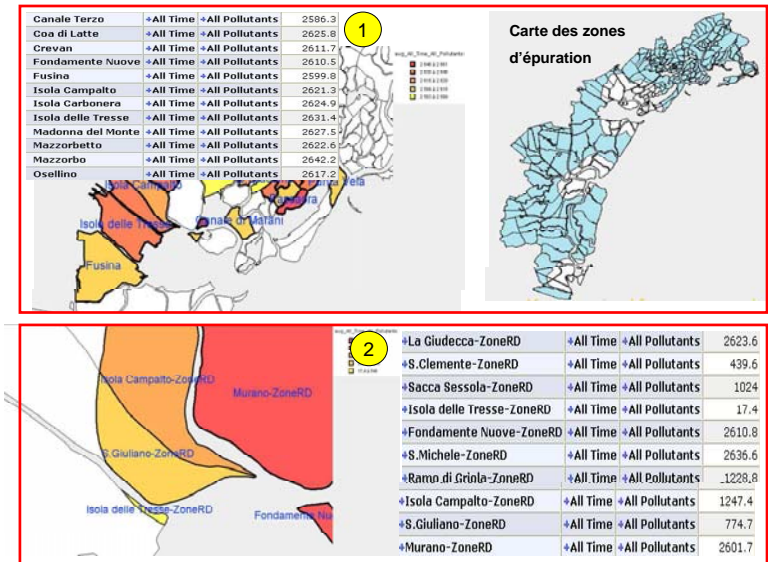


FIG. 11. 1) Données initiales 2) Résultats du GeWolap Overlay

6 Conclusion

L'OLAP spatial vise à intégrer la donnée spatiale dans les systèmes OLAP pour offrir un processus de décision spatial et multidimensionnel. En partant d'une réflexion sur l'analyse multidimensionnelle et spatiale, nous montrons que les modèles SOLAP ne prennent pas complètement en compte la composante sémantique de l'information géographique et le caractère itératif de l'analyse spatiale. Nous redéfinissons les concepts de dimension spatiale et les opérateurs spatio-multidimensionnels afin de pallier cette lacune. Nous introduisons le concept de dimension géographique et spécifions les différents types de hiérarchies possibles en considérant la sémantique des relations entre les membres. Nous identifions également de nouveaux opérateurs spatio-multidimensionnels qui intègrent l'analyse spatiale dans le paradigme multidimensionnel. Nous présentons notre prototype GeWolap, une solution web qui intègre des fonctionnalités OLAP géographiques dans un environnement unique et synchronisé en supportant les dimensions et les mesures géographiques. Ce prototype propose des opérateurs qui intègrent l'analyse multidimensionnelle et l'analyse spatiale. L'introduction de ces opérateurs présente une double difficulté : la prise en compte d'une gestion dynamique de la structure de l'hypercube (en termes de membres et de hiérarchie) et le calcul des mesures associées aux nouveaux membres. GeWolap vise à combiner effectivement l'analyse spatiale et multidimensionnelle dans le même paradigme d'exploration et d'analyse. Actuellement, nous travaillons sur l'intégration des hiérarchies de généralisation cartographique dans GeoCube et sur des politiques d'agrégation ad-hoc pour ce type de hiérarchies. De plus, un cadre global qui détermine automatiquement les variables visuelles à utiliser selon des règles sémiologiques paraît nécessaire. En effet, GeWolap

n'utilise aucune règle sémiologique particulière. L'accent dans ce travail est mis sur la synchronisation de la table de pivot et la composante cartographique. Cette problématique soulève encore de nombreux défis car le nombre d'attributs qui peuvent être représentés sur une carte "lisible" (Bertin et Bonin, 1992) est inférieur à ceux de la table de pivot. Une piste intéressante est l'utilisation des cartes multiples et/ou dynamiques.

7 Bibliographie

- Alter, S. L. (1980). Decision support systems: current practice and continuing challenges. Addison-Wesley.
- Bédard, Y. (1997). Spatial OLAP. 2ème Forum annuel sur la R-D, Géomatique VI: Un monde accessible, 13-14 November, 1997, Montréal.
- Bédard, Y., Merrett T. H., and Han J. (2001). Fundaments of Spatial Data Warehousing for Geographic Knowledge Discovery. *Geographic Data Mining and Knowledge Discovery*. London: Taylor & Francis, 53-73.
- Bédard, Y., Proulx, M., Rivest, S., and Badard, T. (2006). Merging Hypermedia GIS with Spatial On-Line Analytical Processing: Towards Hypermedia SOLAP. *Geographic Hypermedia: Concepts and Systems*. Berlin: Springer, 167-185.
- Bertin, J. Bonin, S. (1992). La graphique et le traitement graphique de l'information. Paris : Flammarion, 277
- Bimonte, S., Tchounikine, A., and Miquel, M., (2005). Towards a spatial multidimensional model. *ACM 8th International Workshop on Data Warehousing and OLAP*, 4-5 November, 2005, Bremen. New York, NY, USA: ACM press, 39-46.
- Bimonte, S., Wehrle, P., Tchounikine, A., and Miquel, M. (2006). GeWOlap: a Web Based Spatial OLAP Proposal. *Int. Workshop on Semantic-based Geographical Information Systems*, Montpellier, France, 29-30 October, 2006. Berlin Heidelberg: Springer-Verlag, 1596-1605.
- Bimonte, S. (2007). Vers l'intégration de l'information géographique dans les entrepôts de données et l'analyse en ligne : de la modélisation à la visualisation. PhD thesis, INSA Lyon.
- Colonese, G., Manhães, R., Montenegro, S., Carvalho, R., and Tanaka, A. (2005). PostGeoOlap: an Open-Source Tool for Decision Support. *2nd Simpósio Brasileiro de Sistemas de Informação*, 26-30 September, 2005, Florianópolis, Brasil
- Denegre, J., and Salgé, F. (1997). *Les systèmes d'information géographique*. Paris : Presses Universitaires de France
- Fidalgo, R., Times, V., Silva, J., and Souza, F. (2004). GeoDWFrame: A Framework for Guiding the Design of Geographical Dimensional Schemas. *6th International Conference on Data Warehousing and Knowledge Discovery*, 1-3 September, 2004, Saragossa, Espagne. Berlin Heidelberg: Springer, 26-37.
- Hernandez, V., Voss, A., Göhring, G., and Hopmann, C. (2005). Sustainable decision support by the use of multi-level and multi-criteria spatial analysis on the Nicaragua

- Development Gateway. *8th Int. Conf. on the Global Spatial Data Infrastructure*, Egypte, 2005.
- Inmon, W.H. (1996). *Building the Data Warehouse*. 2nd Ed. New York: Wiley.
- Jensen, C., Kligys, A., Pedersen, T., and Timko, I. (2004). Multidimensional data modeling for location-based services. *International Journal on Very Large Data Bases*, 13(1):1-21.
- Kouba, Z., Matousek, K., and Miksovský, P. (2000). On Data Warehouse and GIS Integration. *11th International Conference on Database and Expert Systems Applications*, 04-08 September, 2000, London, UK. London: Springer, 604-613.
- Longley, P., Goodchild, M., Maguire, D., and Rhind, D. (2001). *Geographic Information Systems and Science*. New York: John Wiley & Sons.
- Lujan-Mora, S., Trujillo, J., and Song, I. (2002). Multidimensional modeling with UML package diagrams. *21th International Conference on Conceptual Modeling*, 7-11 October, 2002, Tampere, Finlande. Berlin: Springer, 199-213.
- MacEachren, A. M. and Kraak, M. J. (2001). Research challenges in geovisualization. *Cartography and Geographic Information Systems*, 1:3-12.
- Malinowski, E., and Zimányi, E. (2005). Spatial Hierarchies and Topological Relationships in the Spatial MultiDimER model. *British National Conf. on Databases*. Berlin: Springer, 7-28.
- Matias, R., and Moura-Pires, J. (2007). Revisiting the OLAP Interaction to Cope with Spatial Data and Spatial Analysis. *9th International Conference on Enterprise Information Systems*, 12-16 June, 2007, Funchal, Madeira, Portugal. Portugal: INSTICC.
- Mitchell, A (2005). *The ESRI Guide to GIS Analysis: Volume 2: Spatial Measurements and Statistics*. Redlands, Californy: ESRI Press.
- Regnauld, N. and McMaster, R (2007). A synoptic View of Generalization Operators. *Generalisation of Geographic Information Cartographic Modelling and Applications*. Amsterdam: Elsevier Science.
- Rivest, S., Bédard, Y., Proulx, M.-J., Nadeau, M., Hubert F., and Pastor, J. (2005). SOLAP: Merging Business Intelligence with Geospatial Technology for Interactive Spatio-Temporal Exploration and Analysis of Data. *Journal of Int. Society for Photogrammetry and Remote Sensing*, 60(1):17-33.
- Sampaio, M., C., Sousa, A., and Baptista, C. (2006). Towards a Logical Multidimensional Model for Spatial Data Warehousing and OLAP. *ACM 9th International Workshop on Data Warehousing and OLAP*, Washington.USA: ACM Press.
- Scotch, M., and Parmanto, B. (2005). SOVAT: Spatial OLAP Visualization and Analysis Tool. *38th Hawaii Int. Conf. on System Sciences*. IEEE, 42-149.
- Silva, J., Times, V., and Salgado, A (2006). An open source and web based framework for geographic and multidimensional processing. *ACM Symposium on Applied Computing*, 23-27 April, 2006, Dijon, France. New York: ACM Press, 63-67.

- Spaccapietra, S., Parent, C., and Zimanyi, E (2007). Spatio-Temporal and Multi-Representation Modeling: A Contribution to Active Conceptual Modeling. *Active Conceptual Modeling for Learning*. Berlin Heidelberg: Springer.
- Stolte, C., Tang, D., and Hanrahan, P. (2003). Multiscale visualization using data cubes. *IEEE Transaction on Visualization and Computer Graphics*, 9(2):176–187.
- Weibel, R., and dutton, G. (2001). Generalizing Spatial Data and Dealing with Multiple Representations. *Geographic Information Systems and Science*. New York: John Wiley & Sons, 125-155.

Summary

The introduction of spatial data into multidimensional models leads to the concept of Spatial OLAP (SOLAP). In this paper, we show that SOLAP models hardly take into account the semantic component of geographic information and the specificities of spatial analysis. To overcome these limitations, we define the concept of Geographical Dimension and we describe its different types of hierarchies. We propose the introduction of some new operators which permit to adapt the spatial analysis operators to the multidimensional paradigm. Moreover, we present our prototype which permits to integrate these new concepts and provides a web-based spatial and multidimensional interface.

