

Sommaire

Liste des figures	IX
Liste des tableaux	XI
Liste des abréviations	XII
Introduction Générale	1
I Synthèse de l'état de l'art	8
1 Entrepôts de Données et OLAP	9
1.1 Introduction	9
1.2 Systèmes d'information (SI)	10
1.3 Architecture d'un système d'information décisionnel	10
1.3.1 Sources de données	11
1.3.2 Entrepôts et magasins de données	11
1.3.3 Serveurs OLAP	12
1.3.4 Les outils d'analyse	13
1.4 Modélisation multidimensionnelle	14
1.4.1 Modélisation conceptuelle	15
1.4.2 Modélisation logique	18
1.4.3 Modélisation physique	19
1.5 Modélisation des systèmes OLAP : Etat de l'art	19
1.6 Limite des systèmes OLAP	21
1.7 Conclusion	22
2 Systèmes d'Information Géographique et Spatial OLAP	23
2.1 Introduction	24
2.2 L'aide à la décision spatiale	24
2.2.1 Types de décisions spatiales	24
2.2.2 Processus de prise de décision spatiale	25
2.3 L'Information Géographique : Concepts de base	26
2.3.1 Composantes de l'information géographique	26
2.3.2 L'objet géographique	27

2.3.3	Systèmes de coordonnées	27
2.3.4	Notion de l'échelle	29
2.4	Systèmes d'Information Géographique (SIG)	30
2.4.1	Définition	30
2.4.2	Structure de données dans un SIG	31
2.4.3	Les principales fonctions d'un SIG	33
2.4.4	Logiciels SIG	36
2.5	Limite des systèmes d'information géographique	37
2.6	Travaux utilisant les SIG dans la foresterie	37
2.7	ED Spatial et SOLAP	39
2.7.1	Modèle spatio-multidimensionnel	39
2.7.2	Opérateurs de navigation SOLAP	41
2.7.3	Fonctions d'Agrégation dans le SOLAP	41
2.7.4	Architecture d'un système décisionnel basée sur SOLAP	42
2.7.5	Modes d'intégration SIG-OLAP	43
2.7.6	Modélisation spatio-multidimensionnelle : Etat de l'art	44
2.8	Limite des systèmes SOLAP	45
2.9	Conclusion	46
3	Méthodologie Multicritères d'Aide à la Décision	48
3.1	Introduction	49
3.2	L'aide à la décision	49
3.2.1	Typologie des décisions	49
3.2.2	Acteurs de la décision	50
3.2.3	Systèmes Interactifs d'Aide à la Décision (SIAD)	50
3.3	Notions relatives à l'Analyse MultiCritères (AMC)	51
3.3.1	Définition de L'AMC	51
3.3.2	Le concept d'action	51
3.3.3	Critère et famille de critères	52
3.3.4	Matrice des performances	52
3.3.5	Les paramètres subjectifs	53
3.3.6	Relation de surclassement	53
3.3.7	Typologie des problèmes décisionnels	53
3.4	Agrégation multicritères	54
3.4.1	Définition de l'agrégation	54
3.4.2	Type d'agrégation	55
3.5	Méthodes d'AMC par agrégation partielle	55
3.5.1	La famille des méthodes ELECTRE	56
3.5.2	La famille de méthodes PROMETHEE	60
3.6	Synthèse des méthodes d'analyse multicritères	61
3.7	Travaux connexes	62
3.8	Conclusion	63

4	Qualité de l'analyse SOLAP	64
4.1	Introduction	65
4.2	Qualité des données géographiques	65
4.2.1	Critères quantitatifs	66
4.2.2	Critères qualitatifs	67
4.3	Contraintes d'intégrité dans les bases de données spatio-temporelles	67
4.4	Qualité d'analyse spatio-multidimensionnelle (SOLAP)	68
4.4.1	Selon la qualité de données	68
4.4.2	Selon la façon d'agrégation des mesures	68
4.4.3	Selon l'exploration des données	69
4.5	Contraintes d'intégrité dans les cubes de données spatiales	69
4.6	Classification des langages de spécification des CI	71
4.6.1	Langages naturels	71
4.6.2	Langages visuels	71
4.6.3	Langages logiques	72
4.6.4	Langages hybrides	72
4.7	OCL et Spatial OCL	72
4.8	Travaux effectués sur la qualité d'analyse dans les structures MD	74
4.8.1	Travaux sur la qualité d'agrégation	75
4.8.2	Travaux sur la qualité d'exploration	78
4.9	Conclusion	79
II	Contributions	81
5	Approche intégrée (SOLAP-AMC) pour l'aide à la décision. Application dans la gestion agroforestière	82
5.1	Introduction	83
5.2	Description du système d'aide à la décision spatiale proposé "Silvicultura"	84
5.2.1	Sous-système de « base de données »	85
5.2.2	Sous-système de « surveillance »	85
5.2.3	Sous-système de « traitement des problèmes »	85
5.2.4	Sous-système « Interface utilisateur »	86
5.3	Le modèle décisionnel adopté	86
5.3.1	Structuration du Modèle	87
5.4	Exploitation du modèle	88
5.4.1	Proposition de critères de gestion	88
5.4.2	Les méthodes d'analyse multicritères utilisées	88
5.4.3	Déduction des indicateurs et des vérificateurs	89
5.4.4	Analyse spatio-multidimensionnelle (SOLAP)	89
5.4.5	Proposition d'actions candidates	89
5.5	La démarche décisionnelle adoptée par " <i>silvicultura</i> "	90
5.6	Modélisation UML pour " <i>Silvicultura</i> "	91
5.6.1	Diagramme d'activité de " <i>Silvicultura</i> "	91

5.6.2	Diagramme des cas d'utilisation de " <i>Silvicultura</i> "	91
5.6.3	Diagramme de séquence de " <i>Silvicultura</i> "	92
5.7	Modélisation MD de l'Entrepôt de Données Spatiales pour l'Agroforesterie .	93
5.7.1	Profil UML pour les cubes de données spatiales	93
5.7.2	Schéma en constellation pour l'EDS de l'agroforesterie	96
5.8	Etude de Cas : Résultats et Discussion	99
5.8.1	Données utilisées	99
5.8.2	« ELECTRE I » pour le choix des critères de gestion	100
5.8.3	Traitement analytique en ligne spatial	101
5.8.4	Analyse Multicritères (ELECTRE III) pour classer les actions à exé- cuter	104
5.8.5	Agrégation	105
5.8.6	Exploitation	105
5.9	Conclusion	105
6	Un système "UIC-SOLAP" pour le contrôle d'utilisation dans SOLAP	108
6.1	Introduction	108
6.2	Nouvelle CI pour les faits dans le schéma en constellation	109
6.3	Contraintes d'intégrité pour les utilisateurs d'un SOLAP	109
6.3.1	CI-Utilisateurs liés aux données spatiales	110
6.3.2	CI-Utilisateurs liées à la politique d'exploitation du SOLAP	111
6.3.3	CI-Utilisateurs liées aux requêtes SOLAP	111
6.3.4	CI-Utilisateurs liées aux agrégations	112
6.4	Le système " <i>UIC-SOLAP</i> " pour les contraintes d'intégrité utilisateur dans SOLAP	113
6.5	Le processus adopté par " <i>UIC-SOLAP</i> "	113
6.6	Etude de Cas	115
6.7	Discussion des Résultats	117
6.8	Conclusion	119
	Conclusion Générale et Perspectives	121
	Publications issues de la présente thèse	125
	Bibliographie	128

Liste des figures

1.1	Architecture d'un système décisionnel	11
1.2	Un cube de données à trois dimensions.	16
1.3	Modèle en étoile	17
1.4	Modèle en flocon de neige	17
1.5	Modèle en constellation	18
1.6	Architecture ROLAP.	18
2.1	La forme théorique de la terre (géoïde)	27
2.2	Notion de l'ellipsoïde	28
2.3	Types de projection a : cylindrique, b : conique, c : azimutale	29
2.4	Echelle graphique	29
2.5	Représentation vectorielle et matricielle	31
2.6	Organisation en couches thématiques	33
2.7	Modèle Numérique de Terrain (MNT)	35
2.8	Exemple de SIG 3D	36
2.9	Exemple de mesure spatiale	40
2.10	Exemple d'instance d'hypercube spatiale	41
2.11	Architecture d'un système décisionnel basé sur SOLAP.	42
2.12	Exemple de Client SOLAP.	43
4.1	Classification des CI en fonction des éléments MD	70
4.2	CI exprimée en langage visuel	72
4.3	CI spatiale avec le générateur de code Spatial OCL2SQL.	74
4.4	Types des CI d'agrégabilité	78
4.5	Représentation d'un patron de requêtes OLAP	79
5.1	Le système d'aide à la décision spatiale "Silvicultura". Une vue d'ensemble	84
5.2	Modèle décisionnel de "Silvicultura"	86
5.3	Diagramme d'activité de "Silvicultura"	92
5.4	Diagramme des cas d'utilisation de "Silvicultura"	93
5.5	Diagramme de séquence du processus décisionnel	94
5.6	Stéréotypes du profil SDW pour "Silvicultura"	95
5.7	Contrainte OCL dans le contexte du stéréotype hypercube	96
5.8	Dimensions et hiérarchies	97

5.9	Diagramme de classe de l'hypercube "Agro-forest"	98
5.10	Espaces forestiers de la wilaya de Mostaganem	100
5.11	Résultat d'ELECTRE I	102
5.12	Exemple de visualisation avec le client SOLAP "Sylvicole"	103
5.13	Résultats de l'agrégation de ELECTRE III	106
5.14	Résultats de l'exploitation d'ELECTRE III	106
6.1	Les catégories des CI de 'Fait'	109
6.2	CI inter-fact dans un hypercube.	110
6.3	CI-utilisateurs dans un SOLAP.	110
6.4	Exemple de CI-Utilisateur.	111
6.5	Exemple de CI relative aux agrégations.	112
6.6	"UIC-SOLAP" Contraintes d'intégrité Utilisateurs de SOLAP.	114
6.7	Diagramme d'activité pour "UIC-SOLAP"	116
6.8	Stereotype et Class UML pour UserProfile.	117
6.9	Contraintes d'intégrité Inter-facts.	117
6.10	Notification de violation d'une CI de Requête.	118
6.11	Notification de violation d'une CI de Requête.	119

Liste des tableaux

1.1	Différences entre systèmes OLTP et OLAP	14
3.1	Table de performances (Actions-Critères)	52
3.2	Les principales caractéristiques des méthodes ELECTRE	59
3.3	Avantages et inconvénients des méthodes ELECTRE	60
5.1	Critères de gestion, Indicateurs et Vérificateurs	102
5.2	La matrice de Performances et paramètres de ELECTRE III	105

Rapport-Gratuit.com

Liste des abréviations

AGL : Atelier de génie logiciel

AMC : Analyse Multicritères

BI : business intelligence

CI (IC) : Contrainte d'Intégrité (Integrity Constraint)

ED (DW) : Entrepôts de données (Data Warehouse)

EDS (SDW) : Entrepôts de données Spatiales (Spatial Data Warehouse)

ELECTRE : Elimination Et Choix Traduisant la REalité

MD : MultiDimensionnel

OLAP : On-Line Analytical Processing

PROMETHEE : Preference Ranking Organisation METHod for Enrichement Evaluation

SAD (DSS) : Système d'Aide à la Décision (Decision Support System)

SDSS : Spatial Decision Support System

SID : Système d'Information Décisionnel

SIG : Système d'Information Géographique

SOLAP : Spatial On-Line Analytical Processing

UML : Unified Modeling Language

UTM : Universal Transverse Mercator

Introduction Générale

Mise en contexte

Les organisations de différents secteurs privés ou étatiques, à caractère économique ou social, ont un besoin croissant d'outils pour l'analyse de données leur permettant d'obtenir des indicateurs d'aide à la décision.

En effet, ces outils doivent être capables de fournir une vision claire concernant les activités d'une organisation ainsi que son environnement. Ceci est dans le but de développer sa stratégie de production, de conservation du patrimoine existant et le développement de son fonctionnement général.

Les systèmes d'information décisionnels (SID) sont typiquement utilisés pour fournir des mécanismes d'analyse et de stockage de données. Ils sont, principalement, adaptés aux requêtes analytiques (Kimball et Ross, 2003).

Dans ce contexte, les systèmes d'entrepôts de données permettent de fournir un espace de stockage et d'intégration de toutes les données requises pour l'analyse décisionnelle d'une part; et d'une autre part, les systèmes OLAP (On-Line Analytical Processing) définissent des outils intuitifs pour une exploration simple et rapide de ces données. En outre, un système d'aide à la décision spatiale (SAD spatial) est un outil spécifique pour la résolution de problèmes dans les processus de gestion. Basé sur une infrastructure de base des données spatiales, ce système est utilisé pour faciliter la gestion de l'évolution des actions de nature spatiale.

En Algérie, à l'effet du changement croissant de l'activité humaine et des événements naturels qui influent sur le territoire et les ressources naturelles, l'usage des systèmes décisionnels est incontournable dans divers domaines, notamment la gestion des zones agroforestières. Ceci, comprend les forêts, les zones à vocation forestière, les zones humides, etc. Ces espaces sont définis dans le régime général des forêts en Algérie¹.

Quant à la politique forestière de notre pays, les éléments de réflexion en matière de gestion et de préservation du patrimoine devront permettre de :

- Préserver, régénérer et développer le patrimoine existant à travers des interventions sylvicoles et des plans d'aménagement des forêts.
- Conserver et consolider les aires protégées par l'élaboration et la mise en œuvre de plan de gestion.
- Préserver et développer les zones humides par leur identification et classement et la définition des zones prioritaires présentant un intérêt économique et écologique.

1. Loi N° 84-12 du 23 juin 1984 modifié et complété

- Réhabiliter l'espace naturel par la restauration des forêts récréatives et les espaces verts.
- etc.

En outre, en matière de lutte contre l'érosion et d'agriculture de montagne, la politique forestière vise à :

- Préserver la frange littorale par la stabilisation à base d'une formation végétale forestière.
- Maintenir l'équilibre naturel des écosystèmes par des travaux de fixation des dunes et l'amélioration du couvert dunaire.
- Protéger les bassins versants, menacés par l'érosion, à travers l'aménagement intégré de périmètre de mise en valeur.
- etc.

À ce titre, les services chargés de la gestion du domaine forestier possèdent un volume important de données complexes et multisources, d'où la nécessité de systèmes efficaces pour consulter et analyser les tendances dans l'objectif d'une gestion durable de ressources forestières. Cette nécessité se traduit par son rôle fondateur dans les statistiques qui influent d'une manière ou d'une autre sur la courbe de développement économique et social du pays.

Dans cette optique, l'élaboration d'un système d'aide à la décision (SAD) pour la gestion des espaces agroforestiers à plusieurs axes d'observation (propriétés cadastrales, gestion de la faune et la flore, travaux sylvicoles, etc.), implique la manipulation de données multisources. Dans ce contexte, la gestion des données spatiales est la tâche du Système d'Information Géographique (SIG) qui permet l'analyse spatiale et l'exploration des données à référence spatiale.

Un SIG peut être défini comme un «*système de gestion de données conçu pour saisir, stocker, extraire, manipuler, analyser et afficher des données spatiales à des fins de recherche et de prise de décision*» (Pourabbas, 2014). Pour cela, le SIG est largement utilisé dans l'aménagement du territoire et la gestion des ressources naturelles.

Cependant, le SIG est basé sur des processus transactionnels et est incapable d'effectuer une analyse multidimensionnelle offerte par les technologies de traitement analytique en ligne (OLAP) et qui est disponible au sein des SID. C'est pourquoi une nouvelle technologie a vu le jour sous le nom de Spatial-OLAP (SOLAP). En fait, cette dernière permet d'étudier la distribution géographique des phénomènes et d'effectuer une comparaison à travers la granularité géographique par la visualisation des résultats d'analyse sur des tableaux, des diagrammes et des cartes.

Néanmoins, la technologie SOLAP à elle seule, souffre encore de plusieurs inconvénients dus en grande partie au manque de capacité à supporter les problèmes spatiaux.

Afin d'évoluer vers un véritable outil d'aide à la décision, la solution proposée, dans cette thèse, est de l'associer à l'analyse multicritères d'aide à la décision (AMC) qui offre plusieurs avantages lorsqu'il s'agit de prendre en compte des intérêts conflictuels.

En effet, l'AMC est un outil d'aide à la décision qui permet d'intégrer des méthodes afin d'apporter des solutions à des problèmes complexes avec une diversité d'intérêts. Il

évalue l'importance relative des critères et reflète leur importance dans le résultat final (Belton et Stewart, 2002). L'AMC permet donc, le traitement de données quantitatives et qualitatives et peut être utilisée lorsque le problème à traiter est bien défini (Kilgour et Eden, 2010).

Dans le contexte de la politique agroforestière, un système d'aide à la décision spatiale doit prendre en compte des critères économiques, environnementaux et sociaux. L'AMC est l'une des approches fréquemment utilisées dans la planification agroforestière, elle est particulièrement utilisée dans des situations décisionnelles de groupe où de nombreux intervenants sont impliqués (Belton et Stewart, 2002).

Problématique

En plus des outils OLAP déployés pour accéder, visualiser et analyser des données intégrées et agrégées pour faciliter les processus analytiques, une combinaison d'OLAP et de concepts SIG a donné naissance à l'OLAP Spatial (SOLAP) qui peut être définie comme étant «une plate-forme visuelle spécialement conçue pour supporter une analyse spatio-temporelle rapide et facile et une exploration des données selon une approche multidimensionnelle» (Rivest et al, 2003).

Dans le contexte des travaux de recherche concernant le SOLAP, nous retrouvons quelques auteurs qui abordent les problématiques liées à l'utilisation des systèmes SOLAP dans le domaine forestier. Leurs études peuvent être classifiées en trois catégories, selon le facteur influant la mise en place et l'utilisation d'un tel système, à savoir : le développeur, l'utilisateur final et les données manipulées. Cette classification peut se résumer comme suit (Abdallah Bensalloua et Hamdadou, 2018a) :

1. *Des études traitant le 1er facteur relatif aux développeurs* : Cette catégorie concerne les méthodes de modélisation et le développement de la structure du cube de données spatiales.

A ce titre, et dans le domaine forestier, dans (Bakillah et al, 2006), les auteurs traitent la problématique de l'évolution de la structure du cube de données au niveau sémantique et géométrique. En effet, le modèle proposé est basé sur le rétablissement des liens entre des versions du cube. Sur le plan sémantique, leur approche rétablit les liens en employant une fonction de similarité sémantique basée sur l'ontologie et qui tient compte du plus fin niveau de définition des concepts. Au niveau géométrique, cette approche est basée sur une méthode d'indexation Quadtree pour constituer une matrice de correspondances spatiales entre les géométries des différentes époques.

2. *Des études traitant le 2ème facteur relatif aux données manipulées en entrée et en sortie du système* : Dans certains cas, les utilisateurs arrivent à utiliser plus d'un cube de données spatiales pour répondre à des besoins décisionnels.

Dans (Sbouï et al, 2008), les auteurs prennent un exemple dans l'analyse du risque de feu de forêt où les utilisateurs disposent de deux cubes de données spatiales ; et vu que ces cubes sont généralement modélisés différemment d'une organisation à une autre, ou même d'un concepteur à un autre, les auteurs de cet article ont proposé une catégorisation des problèmes d'hétérogénéité en tenant compte des différents éléments

des cubes de données spatiales. Par conséquent, ils ont défini quatre catégories, notamment : l'hétérogénéité Cube-à-Cube, l'hétérogénéité Dimension-à-Dimension, l'hétérogénéité Niveau-à-Niveau et l'hétérogénéité Mesure-à-Mesure. Pour chaque catégorie, ils ont considéré les différentes composantes des modèles de cubes de données, notamment le schéma et les métadonnées.

Une autre étude présentée dans (Taher et Maryvonne, 2005) concerne les dimensions dans les hypercubes ; en effet, l'approche multidimensionnelle proposée par les auteurs, est basée sur le concept des faits vus en différentes dimensions. Ces dimensions sont considérées comme des axes d'analyse formant un espace pour repérer les faits par un ensemble de coordonnées. Les dimensions donc, ont des valeurs discrètes. Néanmoins, ces valeurs ne sont pas suffisantes lors de l'analyse des phénomènes naturels continus, telle que la météorologie ou la pollution. Dans cet article, les auteurs ont introduit des mécanismes basés sur l'interpolation dans les dimensions spatiales et temporelles, qui donnera à l'utilisateur l'impression de naviguer dans un hypercube continu. Dans (Maryvonne et al, 2002), les auteurs proposent des solutions pour concevoir des structures multidimensionnelles lorsque les sources de données sont hétérogènes du point de vue temporel, spatial et sémantique. L'objectif est de permettre l'extraction des connaissances géographiques par l'exploration des données détaillées associées à une époque.

3. *Des études traitant le 3ème facteur relatif aux utilisateurs finaux (décideurs)* : Afin de proposer des solutions pour le traitement interactif, où l'utilisateur pose une action spatiale afin de visualiser rapidement les conséquences, l'expérience de l'utilisateur joue un rôle important dans le processus de décision. En effet, Julien (Julien, 2004) propose une solution pour concevoir une légende interactive pour le système SOLAP. Dans ce contexte, il a étudié la sémiotique graphique et son applicabilité à l'analyse multidimensionnelle.

Une autre question concernant l'utilisation du système SOLAP est celle traitée dans (Guimond, 2005). En effet, l'auteur pose le problème que certains utilisateurs connaissent leurs besoins en termes d'aide à la décision après une période d'utilisation. Pour cette raison, l'auteur a développé des scénarios pour intégrer la conception de la mise en page dans un processus de développement d'applications SOLAP.

En agroforesterie, diverses études ont traité les problèmes de conception, d'intégration de données ou d'utilisation d'un système SOLAP. Ils rapportent la complexité d'une gestion efficace des zones naturelles à des fins écologiques, économiques et sociales. Cependant, dans la plupart des études, les problèmes discutés sont liés à l'hétérogénéité des données manipulées, ou aux méthodes de développement. Ainsi, peu d'études ont abordé des questions liées aux utilisateurs finaux.

Dans le domaine de l'agroforesterie, le même territoire est généralement, sous le contrôle de plusieurs acteurs associés à différentes disciplines dont : les activités sylvicoles, la gestion cadastrale, l'exploitation des produits forestiers et même l'agriculture de montagne. Cela nécessite une gestion intégrée des zones forestières.

Cependant, l'utilisation de l'outil SOLAP par différents acteurs leur permet de manière indépendante d'extraire des indicateurs décisionnels qui sont hétérogènes et conduisent à

des conflits. Cela ne permet pas une gestion intégrée prenant en compte les considérations écologiques, économiques et sociales. En conséquence, le processus décisionnel ne bénéficiera d'aucun partage d'expérience d'une part ; et d'autre part, l'existence de plusieurs intervenants implique souvent des situations conflictuelles, d'où les exigences de chacun d'eux, doivent être prises en considération.

De plus, l'analyse multidimensionnelle offerte pour différents utilisateurs du SOLAP peut engendrer des problèmes de consistance des résultats. Ceci exige de mettre en évidence des mécanismes de contrôle de qualité de manipulation aux différents stades du SID.

Contribution

L'objectif principal de la présente étude concerne la conception et la mise en œuvre d'un système d'aide à la décision spatiale et son application dans l'agroforesterie. Le système devrait prendre en compte toutes les exigences complexes de ce domaine liées à la dimension spatiale des données d'une part, et aux différents intérêts économiques, écologiques et sociaux d'autre part.

Pour cela, l'intégration "SOLAP-AMC" peut aider les décideurs à gérer efficacement les zones agroforestières en tirant profit de l'analyse spatiale des SIG, de l'analyse multidimensionnelle des systèmes OLAP et l'AMC qui permet l'analyse de problèmes complexes avec différentes exigences.

En effet, notre proposition est basée sur l'évaluation des critères de gestion, définis par les intervenants dans le domaine de l'agroforesterie, en utilisant l'analyse multicritères (AMC) d'une part, et l'extraction d'indicateurs pour l'analyse de la situation par l'intermédiaire de SOLAP d'une autre part. Les résultats de cette analyse devraient permettre aux acteurs de la gestion agroforestière de définir l'ensemble des opérations (*Actions*) pouvant être mises en œuvre. L'AMC permet alors de classer ces opérations. La liste finale est donc, la solution proposée en cas de présence d'un risque potentiel, qui peut menacer le fonctionnement des écosystèmes agroforestiers, ou une solution de mesure préventive dans le cas contraire.

A travers le présent travail, d'autres objectifs sont visés. En effet, les tâches suivantes sont à effectuer :

- La proposition d'une catégorisation des études dans le domaine de la foresterie et de l'OLAP spatial, en fonction des facteurs interférant à savoir le concepteur, l'utilisateur final et les données ;
- La proposition d'un système décisionnel Silvicultura fondée sur une approche intégrée SOLAP-AMC pour la gestion agroforestière.
- La proposition d'un système UIC-SOLAP pour le contrôle de qualité de l'exploitation du système Silvicultura, à travers les contraintes d'intégrité utilisateurs, au niveau de différentes phases depuis la modélisation jusqu'à l'exploitation de l'application SOLAP.
- L'adoption de la modélisation UML (Unified Modeling Language) qui fournit des mécanismes puissants pour représenter les aspects statiques et dynamiques des systèmes

- proposés, tels que les diagrammes d'activités, des cas d'utilisation, de séquences et de classes ;
- La proposition d'un profil UML, qui permet d'étendre les concepts de l'UML afin de considérer les propriétés principales de la modélisation multidimensionnelle (MD). Ceci permet de construire des cubes de données incluant plusieurs tables de faits ;
 - La mise en œuvre du profil UML proposé, pour l'implémentation du schéma en constellation de l'entrepôt de données spatiales de l'agroforesterie ;
 - La proposition d'un outil SOLAP "*Sylvicole*" pour l'analyse multidimensionnelle avec l'usage du langage MDX (Multi Dimensional eXpressions) à travers l'architecture de l'entrepôt des données spatiales. Cela permet une meilleure exploration temporelle de données et une analyse spatiale avec des mécanismes SIG intégrés.
 - L'utilisation des méthodes de l'AMC à savoir ELECTRE I et ELECTRE III. Cela permet la prise en compte de toutes les considérations ainsi que les points de vue relatifs aux différents intervenants dans le domaine de l'agroforesterie.
 - Le développement de l'application "*Forests*" qui permet le contrôle d'analyse effectuée par les utilisateurs du système SOLAP.

Organisation de la thèse

La suite de cette thèse s'organise en six chapitres répartis en deux parties comme suit :
Partie 1- Synthèse de l'état de l'art. Cette partie présente, principalement, les concepts liés au cadre général de la thèse. Elle comprend trois chapitres :

Dans le premier chapitre, nous présentons les concepts de base des systèmes d'information décisionnels, relatifs en particulier aux entrepôts de données (ED) et On-Line Analytical Processing (OLAP). Nous présentons, également, les principes de la modélisation multidimensionnelle. Enfin, nous présentons les principales solutions proposées dans la littérature afin de modéliser les cubes de données et leurs éléments multidimensionnels.

Dans le deuxième chapitre, nous présentons les problématiques d'aide à la décision spatiale. Nous présentons également les concepts liés à l'information spatiale et les systèmes d'information géographique. Puis, sont présentés les concepts se rapportant aux systèmes d'information décisionnels spatiaux (SDSS), relatifs en particulier aux EDS et SOLAP. Ensuite, quelques solutions proposées pour la modélisation spatio-multidimensionnelle sont présentées.

Dans le troisième chapitre, nous présentons les concepts de la méthodologie multicritères d'aide à la décision. Les notions relatives à l'aide à la décision sont présentées avant de présenter les différentes familles et méthodes d'analyse multicritères. Enfin, nous présentons quelques solutions basées sur l'AMC pour résoudre des problèmes en liaison avec la présente thèse.

Dans le quatrième chapitre, nous présentons les concepts de la qualité d'analyse dans les systèmes SOLAP. Il s'agit de la présentation de la qualité des données dans les

EDS, la présentation de la qualité d'agrégation relative à l'application des opérateurs ; et la qualité d'exploration des données dans le but d'extraire les indicateurs d'aide à la décision. Quelques travaux dans la littérature sont présentés à la fin du chapitre.

Partie 2- Contributions. Cette partie décrit nos contributions relatives au cadre général de cette thèse. Elle comprend deux chapitres :

Dans le cinquième chapitre, nous présentons notre première contribution relative à la proposition d'une approche intégrée (SOLAP-AMC) pour l'aide à la décision et son application dans la gestion agroforestière. Il s'agit de la réalisation d'une analyse du domaine d'étude notamment l'agroforesterie ; ensuite, la proposition d'un système d'aide à la décision en décrivant les différents éléments du processus décisionnel. Nous passons, par la suite, à sa modélisation UML à travers les différents diagrammes afin de modéliser les aspects statiques et dynamiques du processus décisionnel. La modélisation de l'ED pour l'agroforesterie grâce au profil-UML, l'implémentation d'un outil SOLAP *Silvicole*, puis, l'agrégation multicritères sont présentés dans ce chapitre. Enfin, une étude de cas est présentée pour appliquer l'approche proposée à un jeu de données relatives aux espaces forestiers de la wilaya de Mostaganem.

Dans le sixième chapitre, nous présentons notre deuxième contribution qui concerne la mise en œuvre du système UIC-SOLAP dédié au contrôle de la qualité d'utilisation du système décisionnel. Ce contrôle est assuré par l'implémentation des contraintes d'intégrité relatives aux utilisateurs du SOLAP *Silvicole*, dont le développement est décrit dans le chapitre précédent à travers l'application *Forest*.

Le présent document s'achève par une conclusion où nous récapitulons les apports de la présente étude tout en ouvrant des perspectives de recherche.

Première partie

Synthèse de l'état de l'art

Chapitre 1

Entrepôts de Données et OLAP

Dans ce chapitre, nous présentons les concepts de base se rapportant aux systèmes d'information décisionnels, notamment les Entrepôts de Données et les systèmes OLAP ainsi que leurs architectures et modélisation.

Sommaire

1.1	Introduction	9
1.2	Systèmes d'information (SI)	10
1.3	Architecture d'un système d'information décisionnel	10
1.3.1	Sources de données	11
1.3.2	Entrepôts et magasins de données	11
1.3.3	Serveurs OLAP	12
1.3.4	Les outils d'analyse	13
1.4	Modélisation multidimensionnelle	14
1.4.1	Modélisation conceptuelle	15
1.4.2	Modélisation logique	18
1.4.3	Modélisation physique	19
1.5	Modélisation des systèmes OLAP : Etat de l'art	19
1.6	Limite des systèmes OLAP	21
1.7	Conclusion	22

1.1 Introduction

Dans le but d'aider les décideurs pour atteindre leurs objectifs et élaborer des plans de gestion, l'utilisation des technologies de l'information est très nécessaire en particulier celles liées à l'informatique décisionnelle (BI : Business Intelligence). Ceci est devenu incontournable dans un environnement complexe caractérisé par l'incertitude de l'information et de la décision. Pour ce fait, l'information appropriée doit être offerte au moment opportun.

Cette information est, actuellement, disponible en grande quantité grâce au développement des solutions logicielles et matérielles, permettant sa récolte, traitement et stockage dans des structures de données décisionnelles dédiées à la prise de décisions, tout en prenant en compte le caractère multidimensionnel de la vision du gestionnaire.

La mise en place d'une architecture de l'informatique décisionnelle, est donc d'une importance croissante, permettant ainsi, la gestion de la performance d'une organisation et de l'amélioration de sa courbe de production. Cette architecture est basée généralement sur les concepts d'entrepôt de données (ED) (DW : Data Warehouse). Dans ce qui suit de ce chapitre, nous allons présenter l'ensemble des concepts dont les entrepôts de données sont basés pour servir les systèmes d'information à caractère d'aide à la décision.

1.2 Systèmes d'information (SI)

Un système d'information (SI) signifie un ensemble de moyens techniques, administratifs, et humains réunis pour collecter et transmettre les informations entre les membres d'une organisation. Il existe deux grandes familles de SI (Ferragu, 2013) :

a) **Les SI opérationnels** : qui sont utilisés pour la gestion des transactions quotidiennes.

(Reix, 2004) a donné une définition pour le système d'information qui est la suivante : «un système d'information est un ensemble organisé de ressources : matériel, logiciel, personnel, données, procédures... permettant d'acquérir, de traiter, de stocker des informations (sous forme de données, textes, images, sons, etc.) dans et entre des organisations». À partir de cette définition, nous constatons qu'un SI se base sur les moyens matériels et logiciels pour la gestion quotidienne de toutes les opérations de production ou des moyens de production. Les données stockées peuvent être sources d'information pour restituer des connaissances nécessaires pour le contrôle périodique du fonctionnement et de la stratégie générale d'une entreprise. Ceci est la mission de la deuxième famille des SI.

b) **Les SI décisionnels (SID)** : qui ont une finalité d'aide à la décision. Dans ce deuxième cas, les organisations ont généralement recours aux systèmes d'information qui offrent les mécanismes de stockage et d'analyse de données et qui sont adaptés aux requêtes d'analyse décisionnelle (Kimball et Ross, 2003 ; Malinowski et Zimányi, 2008). Ils ont pour but d'aider les décideurs dans leur processus de prise de décision (Golfarelli et Rizzi, 2009).

Dans ce contexte, le concept d'entrepôt de données (ED) est utilisé pour la gestion des données volumineuses d'une part ; d'autre part, les techniques d'analyse en ligne (On-Line Analytical Processing (OLAP)) constituent des outils intuitifs pour l'exploration simple et interactive permettant une analyse approfondie des données dans des structures multidimensionnelles.

1.3 Architecture d'un système d'information décisionnel

Il n'existe pas de consensus sur une typologie d'architectures pour les SID. L'architecture type d'un système décisionnel basé sur l'entrepôt de données et l'OLAP peut être représentée par la Figure 1.1. Elle est divisée en quatre niveaux :(Naoum, 2006)

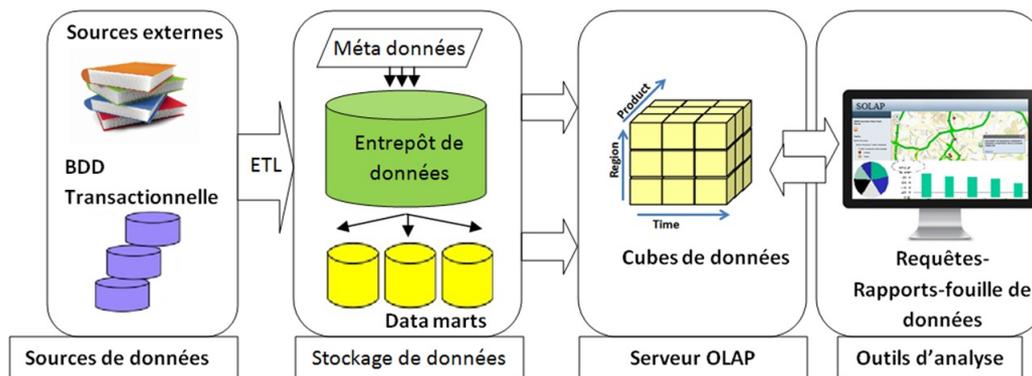


FIGURE 1.1 – Architecture d'un système décisionnel

1.3.1 Sources de données

Les systèmes décisionnels peuvent avoir besoin de données issues de différentes sources et de différents formats de stockage. Ceci regroupe les fichiers texte, les rapports, les fichiers de base de données issus de différents systèmes de gestion de bases de données (SGBD), etc.

Pour ce fait, ces données doivent passer par le processus d'Extraction, Transformation et chargement (ETL) pour qu'elles soient exploitables. Ce processus peut être résumé comme suit (Naoum, 2006) :

- a) **Extract (E)** : Dans la phase d'extraction, seules les données destinées à l'exploitation pour l'analyse qui sont gardées en se connectant aux différentes applications ou bases en production.
- b) **Transform (T)** : Dans la phase de transformation, la mise au format des données, la fusion ou l'éclatement des informations et l'agrégation des données peuvent être effectuées.
- c) **Load (L)** : Enfin, dans la phase de chargement, les informations sont stockées dans les entrepôts de données.

1.3.2 Entrepôts et magasins de données

Nous distinguons deux types de structures de données :

1. *Entrepôts de données*

Selon Immon (Immon, 2005), un Entrepôt de Données (ED) est défini comme étant «une collection de données orientées sujet, intégrées, non volatiles, historisées disponibles pour le support du processus de prise de décision». Cette définition peut être expliquée comme suit :

- *Orientées sujet* : les données sont organisées par sujet d'analyse selon les besoins analytiques des entreprises.
- *Intégrées* : les données hétérogènes issues de différentes sources feront l'objet d'une intégration dans un espace de stockage.

- *Non volatiles* : les données ne sont ni modifiables, ni supprimables.
- *Historisées* : du fait que les données sont non volatiles, des intervalles de temps leurs sont associés.

Pour gérer l'ensemble des données entreposées, l'ED doit disposer des "informations sur les données", à savoir les « *métadonnées* ». Celles-ci doivent permettre de répondre aux questions telles que : *Comment extraire les données ? Quelles transformations effectuer ? etc.*

Les métadonnées doivent également spécifier les droits d'accès et d'utilisation associés à ces données.

Les données entreposées dans les entrepôts de données sont souvent stockées sous forme de vues matérialisées. Il s'agit de tables contenant les résultats de requêtes. Elles améliorent l'exécution des requêtes en pré-calculant les opérations les plus coûteuses comme la jointure et l'agrégation (Boulil, 2012).

2. Magasins de données

Un magasin de données (data mart) peut être considéré comme un sous-ensemble de l'ED concernant un secteur particulier de l'entreprise ou d'une organisation quelconque. Il s'agit de données extraites, adaptées à une classe particulière de décideurs. Un magasin de données peut être considéré comme un petit entrepôt correspondant à un sujet précis. Ceci permet de réduire le temps de réponse aux requêtes (Naoum, 2006).

1.3.3 Serveurs OLAP

Dans cette section, nous allons définir le concept de système OLAP et les opérateurs assurés par son serveur.

1. Système OLAP

Sur la base de ce type de structure (ED), les systèmes OLAP sont utilisés pour définir un modèle analytique transformant les données entreposées en informations d'aide à la décision.

Selon les auteurs dans (Codd et al., 1993) et (Kimball et Ross, 2003), «un système OLAP est une catégorie d'outils qui permet exploration interactive suivant une approche multidimensionnelle à plusieurs niveaux d'agrégation». Ceci peut se réaliser en exploitant un ED. Il a pour but d'évaluer l'activité et aide à la décision au sein d'une organisation.

Dans (Codd et al., 1993), un cahier des charges comprenant douze règles que doivent satisfaire les modèles OLAP est défini :

- (a) la structure multidimensionnelle : le modèle OLAP est multidimensionnel,
- (b) la transparence : le serveur OLAP est transparent pour l'utilisateur,
- (c) l'accessibilité : l'utilisateur OLAP dispose de l'accessibilité à toutes les données nécessaires à ses analyses,
- (d) la stabilité : le système reste stable quelque soit le nombre de dimensions,

- (e) l'architecture client-serveur : le serveur OLAP s'intègre dans une architecture client serveur,
- (f) le dimensionnement : le dimensionnement est générique pour assurer les analyses,
- (g) la gestion complète : le serveur OLAP assure la gestion des données clairessemées,
- (h) les multiutilisateurs : le serveur OLAP offre un support multiutilisateur (gestion des mises à jour, intégrité, sécurité),
- (i) l'inter-dimension : le serveur OLAP permet la réalisation d'opérations inter dimensions sans restriction,
- (j) l'aspect intuitif : le serveur OLAP permet une manipulation intuitive des données,
- (k) la flexibilité : la souplesse de l'édition des rapports est intrinsèque au modèle,
- (l) l'analyse sans limites : le nombre de dimensions et de niveaux d'agrégation possibles est suffisant pour autoriser les analyses les plus poussées.

A la différence des systèmes transactionnels (OLTP : On-Line Transactional Processing) qui permettent la gestion des activités opérationnelles quotidiennes de l'entreprise, la technologie OLAP est proposée pour fournir un support pour la prise de décision.

La Table 1.1 résume quelques différences entre ces deux types de systèmes (Boullil, 2012).

2. *Opérateurs OLAP*

Le serveur OLAP doit assurer un ensemble d'opérateurs permettant la navigation à travers les données entreposées. Nous citons, dans ce qui suit, les opérateurs typiques (Rivest et al., 2003) :

- *Roll-up* : cet opérateur permet d'agréger les valeurs de mesure en montant dans une hiérarchie de dimension.
- *Drill-down* : cet opérateur augmente le niveau de détail de la mesure en descendant dans la hiérarchie de dimension inversement à *Roll-up*,
- *Slice* : cet opérateur effectue une sélection des cellules de l'hypercube en utilisant une condition définie sur les membres d'une dimension. En résultat nous obtiendrons un sous hypercube.
- *Dice* : cet opérateur est utilisé pour réaliser une sélection en utilisant une condition définie sur deux dimensions ou plus.
- *Projection* : cet opérateur permet de sélectionner un sous-ensemble de mesures de l'hypercube.
- *Pivot* : cet opérateur effectue un pivotage des axes de l'hypercube.

1.3.4 Les outils d'analyse

L'outil d'analyse (client OLAP) représente l'élément le plus important pour l'utilisateur final qui permet d'exploiter les données stockées. C'est l'élément qui correspond à la partie visible du système par rapport au décideur.

TABLE 1.1 – Différences entre systèmes OLTP et OLAP

	OLTP	OLAP
Vocation (objectif)	Gestion des activités opérationnelles quotidiennes	évaluation de l'activité et aide à la décision
Utilisateurs	nombreux (des milliers), agents opérationnels	moins nombreux (des centaines), analystes et décideurs
Pattern d'utilisation	régulier, prédictible, et fréquent	irrégulier, non prédictible, et moins fréquent
Modèle de données	Normalisé (3FN) et optimisé pour les exigences de performance des traitements transactionnels	dénormalisé et optimisé pour les performances des traitements analytiques
Données	Opérationnelles orientées transaction, détaillées, courantes (un horizon temporel moyen de 60 à 90 jours), non redondantes et moins volumineuses	Orientées analyse, moins détaillées, historiques (un horizon temporel moyen de 5 à 10 ans), redondantes et volumineuses
Mode d'accès	lecture, rajout, modification et suppression	lecture et rajout
Type de traitements	Transactionnel : accès à des centaines d'enregistrements	analytique : agrégation et accès à des millions d'enregistrements
Type technologies	Optimisé pour le traitement transactionnel (temps de réponse et gestion d'accès concurrents)	Optimisé pour le traitement analytique (historisation et analyse en ligne)

L'analyse de données dans un système décisionnel repose sur des outils d'analyse statistiques dont le concept d'OLAP se base. Il représente la technologie qui offre une analyse multidimensionnelle avec un affichage des résultats sous forme de graphiques ou des tableaux (Naoum, 2006).

1.4 Modélisation multidimensionnelle

La Modélisation multidimensionnelle d'après Teste (Teste, 2000), consiste à « *considérer un sujet analysé comme un point dans un espace à plusieurs dimensions. Les données sont organisées de manière à mettre en évidence le sujet analysé et les différentes perspectives de l'analyse* ».

L'objectif d'une modélisation multidimensionnelle est de permettre aux analystes d'avoir une vision des données qui supporte et aide leur processus de prise de décision.

Les données sont alors, stockées dans une structure multidimensionnelle et à l'intersection de plusieurs dimensions se trouvent des valeurs nommées indicateurs ou variables. Ces valeurs sont calculées par le moteur OLAP à travers des opérations mathématiques ou statistiques plus ou moins complexes.

Nous distinguons trois types de modélisation multidimensionnelle à savoir conceptuelle, logique et physique décrites ci-dessous (Boulil, 2012).

1.4.1 Modélisation conceptuelle

La modélisation conceptuelle fait référence aux concepts définis par les auteurs dans (Ravat et al., 2005) et qui sont utilisés dans différents types de schémas de données (Naoum, 2006 ; Kimball, 1996).

1. Concepts de base

Un ensemble de concepts sont utilisés pour concevoir un modèle multidimensionnel de données. Dans la littérature, certains auteurs proposent d'autres définitions (Bédard et Han, 2009).

- a) **Fait** : c'est le concept qui modélise le sujet de l'analyse. Il est formé des informations de l'activité analysée.
- b) **Mesure** : C'est la valeur d'attribut quantitatif ou qualitatif qui évalue un fait. Nous distinguons trois types de mesures (Boulil, 2012) :
 - *Mesures de type « flux »* : mesures qui peuvent être sommées selon toutes les dimensions (e.g. "montant de vente") ;
 - *Mesures de type « stock »* : mesures qui ne peuvent pas être sommées selon certaines dimensions (e.g. "population"),
 - *Mesures de type « valeur par unité »* : mesures qui ne peuvent jamais être sommées (e.g. "température").
- c) **Dimension** : ensemble de paramètres qui peuvent faire varier les mesures. Elle modélise une perspective de l'analyse
- d) **Agrégation des mesures** : c'est une opération qui permet de calculer différents indicateurs d'analyse à différents niveaux de détail. Elle utilise trois éléments clefs du modèle multidimensionnel à savoir la mesure, la fonction d'agrégation et la hiérarchie de dimension. Elle dépend des conditions d'agrégabilité. Dans la littérature, nous distinguons trois conditions d'agrégabilité (Lenz et Shoshani, 1997),
 - *Disjonction* : pour permettre d'éviter le comptage en double des valeurs de mesure.
 - *Complétude* : pour le but d'éviter le problème d'agrégats incomplets.
 - *Compatibilité de type* : cette condition vérifie que les natures des trois éléments, mesure, dimension et fonction d'agrégation, sont compatibles.
- e) **Hiérarchie** : une hiérarchie permet d'organiser les membres d'une dimension selon leur niveau de détail. Les hiérarchies peuvent être classées en deux catégories (Mazón et al., 2009) :
 - *Hiérarchies régulières* : qui satisfont les conditions de la complétude et la disjonction. Nous distinguons :

- *Hiérarchies strictes* : spécifient, pour chaque membre feuille, un seul chemin d'agrégation vers le membre racine de la hiérarchie.
- *Hiérarchies onto* : tout membre non feuille possède au moins un membre fils et tous les membres feuille se trouvent au même niveau qui est le niveau d'agrégation feuille de la hiérarchie.
- *Hiérarchies covering* : une hiérarchie est dite covering si elle ne présente pas de raccourcis ou de sauts dans les liens d'agrégation.
- *Hiérarchies irrégulières* : ces hiérarchies ne satisfont pas l'une ou les deux conditions d'agrégabilité citées ci-dessus (disjonction et complétude).

f) **Cube de données** : En plus de ces concepts, la notion du cube de données (hypercube) a été proposée par (Gray et al., 1997) et est défini comme « *un ensemble de données organisées selon des dimensions. On appelle mesure la valeur contenue dans une cellule du cube, associée aux valeurs prises sur les dimensions composant le cube* ».

La Figure 1.2, représente un cube à trois dimensions : produits, localisations et temps (Bimonte et al., 2007) . Dans ce cas, la mesure du cube est constituée des résultats des ventes des produits pour différentes villes à différents mois.

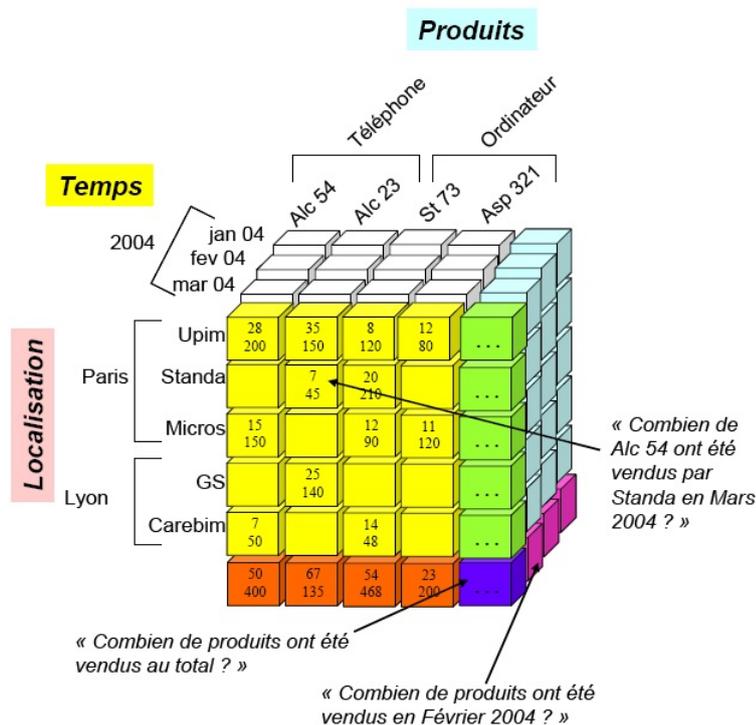


FIGURE 1.2 – Un cube de données à trois dimensions.

2. Schémas de données

Les concepts définis dans la section précédente peuvent être représentés en utilisant différents schémas à savoir : en étoile, en flocon de neige, en constellation et mixte (Teste,

2000).

a) **Modèle en étoile** (*star schema*) : comme illustré dans la Figure 1.3, le schéma est constitué du fait central et des dimensions représentées de manière dénormalisée (Teste, 2000).

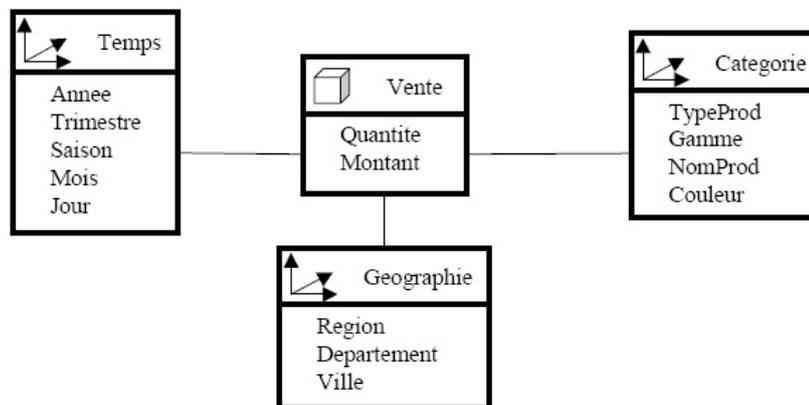


FIGURE 1.3 – Modèle en étoile

b) **Modèle en flocon** (*snowflake*) : dans ce schéma, la table de faits est conservée et les dimensions sont éclatées conformément à sa hiérarchie. Ce modèle permet d'éviter le problème de redondance qu'on peut trouver dans le modèle en étoile, voir Figure 1.4) (Teste, 2000).

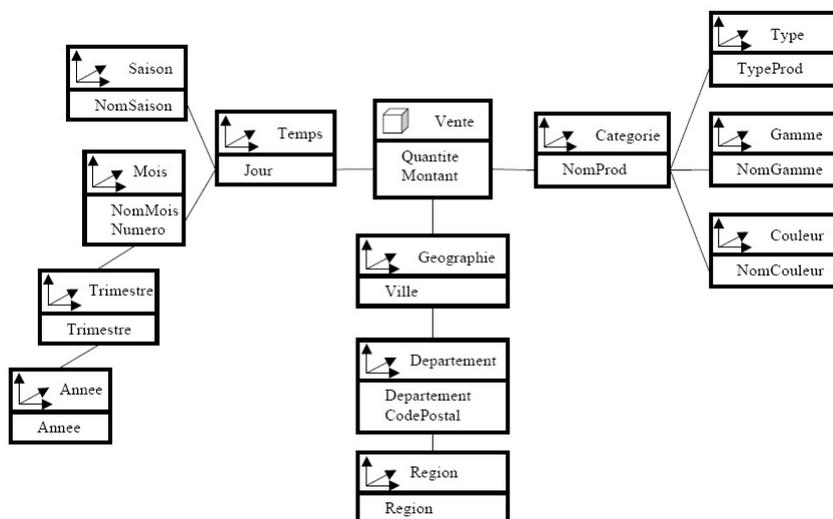


FIGURE 1.4 – Modèle en flocon de neige

c) **Modèle en constellation** : cette technique consiste à fusionner plusieurs modèles en étoile. Il correspond donc à plusieurs tables de faits qui partagent des dimensions communes, voir Figure 1.5) (Teste, 2000).

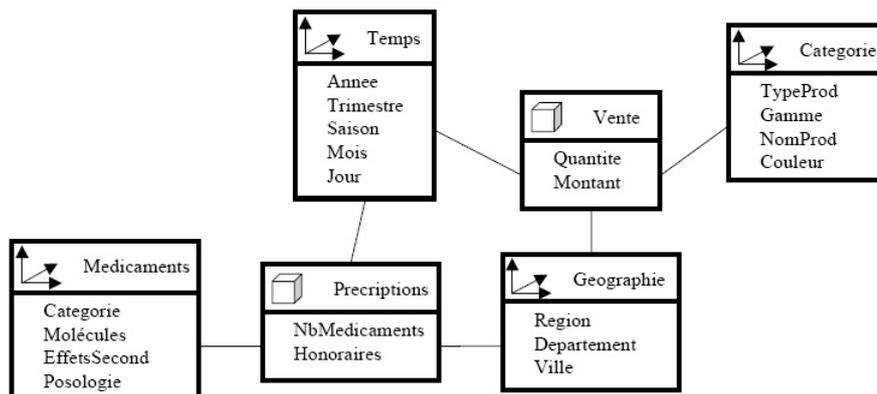


FIGURE 1.5 – Modèle en constellation

d) *Modèle mixte* : cette technique consiste à fusionner plusieurs modèles en étoile et en flocon de neige.

1.4.2 Modélisation logique

La modélisation logique représente la technique du stockage physique des données selon le système de gestion de base de données (SGBD) utilisé ce qui donne naissance à trois modèles logiques (Rivest, 2000) :

a) *OLAP Relationnel (ROLAP)*

Dans ce type d’implémentation un Système de Gestion de Bases de Données Relationnelles (SGBDR) est utilisé pour le stockage des données. Le serveur ROLAP permet de simuler la vue multidimensionnelle et de communiquer entre le SGBDR et les outils client OLAP. L’architecture ROLAP permet une grande souplesse dans la gestion de gros volumes de données et une meilleure administration des données. La Figure 1.6 représente l’architecture ROLAP (Rivest,2000).

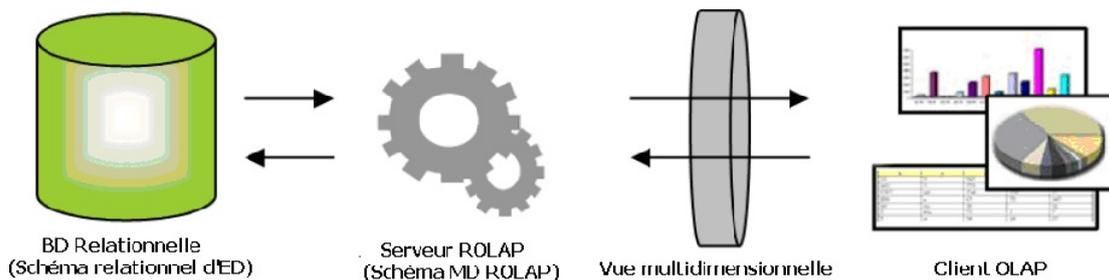


FIGURE 1.6 – Architecture ROLAP.

b) *OLAP Multidimensionnel (MOLAP)*

Dans ce type d’implémentation un Système de Gestion de Base de Données Multidimensionnelles (SGBDM) natifs est utilisé. Dans ce cas les données décisionnelles sont

maintenues dans des structures multidimensionnelles. Les cellules des différents hypercubes sont pré-calculées à partir de l'ED ou des bases sources et stockées dans ces structures. Par conséquent, Les opérateurs OLAP permettent d'explorer ces cellules d'une façon simple et rapide.

L'architecture MOLAP offre moins de performances que la précédente concernant l'administration de données à cause du pré-calcul des cellules.

b) OLAP hybride (HOLAP)

Ce type d'implémentation combine les deux technologies ROLAP et MOLAP. En effet, une partie des données dont les utilisateurs accèdent le plus fréquemment est stockée dans des structures multidimensionnelles gérées par un SGBDM. Une autre partie est stockée dans des structures relationnelles gérées par un SGBDR.

Les systèmes HOLAP offrent un compromis des avantages et inconvénients des systèmes ROLAP et MOLAP.

1.4.3 Modélisation physique

Après avoir effectué les deux premières modélisations conceptuelle et logique, nous devons passer à la modélisation physique de l'ED. Cette phase a pour objet d'appliquer techniques adéquates pour stocker efficacement les données afin d'en assurer un accès rapide. Ceci correspond à la traduction du modèle logique dans le langage de la plateforme d'implémentation (e.g. Oracle, MS SQLServer, ... etc), (Mazon et Trujillo, 2008). Ensuite, des techniques d'optimisation peuvent être appliquées en vue d'optimiser les performances (Bouil, 2012) :

- *La matérialisation de vues* : certaines requêtes les plus utilisées sont calculées et stockées physiquement pour améliorer le temps de réponse. Néanmoins, certains problèmes peuvent se présenter tel que le choix des requêtes et leurs mise à jour.
- *L'indexation* : c'est une technique qui vise à assurer un accès direct aux données. Nous distinguons les indexes binaires dans des tables simples ; et les indexes de jointure entre deux tables.
- *La fragmentation* : c'est une technique qui consiste à diviser les tables volumineuses en plusieurs fragments pour accélérer l'accès aux données.

1.5 Modélisation des systèmes OLAP : Etat de l'art

Les recherches dans le domaine de la modélisation des EDs sont nombreuses. Leur but est principalement de pouvoir fournir des modèles claires et indépendants de l'implémentation (Turlone, 2003).

En ce qui concerne la modélisation conceptuelle, les études dans ce domaine se basent généralement sur l'extension des standards tels qu'UML et Entité-Relation, ou bien des modèles ad hoc. Toutefois, dans la plupart des travaux récents, l'UML est le langage le plus utilisé pour la modélisation conceptuelle des aspects statiques et dynamiques de systèmes d'ED. Nous soulignons aussi l'importance des mécanismes d'extension qu'il offre

à travers les profils-UML. Nous notons, aussi, son interopérabilité avec tous les langages de programmation orienté-objets, ainsi que le langage d'expression des contraintes d'intégrité OCL. Aussi, les règles de passage des outils AGL (Atelier de Génie Logiciel) vers les différents systèmes de gestion de bases de données (SGBD), sont claires.

Dans la littérature, nous constatons les contributions remarquables de quelques travaux tels que (Abelló et al., 2006) où les auteurs ont défini un modèle multidimensionnel contenant les différents concepts permettant la conception des cubes de données de type constellation, en se basant sur les packages UML pour organiser les niveaux de détail. Les auteurs dans (Lujan-Mora et al., 2006) ont proposé une amélioration de la solution précédente par l'extension de l'UML en utilisant les profils qui définissent une spécialisation d'UML pour la modélisation multidimensionnelle. Pour éviter une mauvaise conception, un ensemble de contraintes d'intégrité sont définis dans le profil. Afin d'implémenter cette modélisation, les auteurs ont utilisé l'AGL Rational Rose¹.

Dans (Pinet et Schneider, 2009), les auteurs définissent une modélisation UML avec quelques spécialisations telles que : Classe identifiée destinée à la modélisation des faits et les niveaux d'agrégation. Pour la modélisation des relations d'agrégation ils définissent le concept Association d'agrégation. Ils ont aussi, utilisé OCL pour exprimer les contraintes d'intégrité.

Les auteurs dans (Prat et al., 2010), ont défini une organisations entre les concepts de l'entrepôt (la table de faits et les tables de dimension) d'une part, et ceux du cube de données (hypercube, hiérarchie, etc.). Pour l'expression des contraintes d'agrégation, les auteurs utilisent PPR (Production Rule Representation language).

En outre, en se basant sur le standard ER, les auteurs dans (Malinowski et Zimányi, 2008) ont défini un modèle conceptuel multidimensionnel pour les EDs. Pour la représentation des hiérarchies les auteurs classifient ces hiérarchies et leurs attribuent des notations graphiques.

Dans (Tryfona et al., 1999) et (Sapia et al., 1998), les auteurs ont également proposé des extensions du modèle ER pour la modélisation conceptuelle des concepts multidimensionnels de l'ED.

Pour les modèles ad hoc, nous pouvons citer les études effectuées par les auteurs dans (Torlone, 2003) et (Husemann et al., 2000). En effet, pour la modélisation de différents concepts multidimensionnels, ils ont défini une notation graphique. Ce type de modèle est caractérisé par sa difficulté de développement.

En ce qui concerne la modélisation logique relative à l'implémentation ROLAP de l'entrepôt de données, les auteurs dans (Malinowski et Zimányi, 2008) ont proposé un schéma en étoile qui ne contient aucune normalisation. En effet, pour chaque dimension, il regroupe tous ses niveaux d'agrégation dans une seule table. Ceci a pour but d'optimiser le temps de recherche de données en réponse de requêtes analytiques.

Ce modèle présente l'inconvénient de produire des tables de dimension très volumineuses contenant des valeurs vides. D'autres travaux ont proposé des normalisations maximales. Cette solution présente aussi des inconvénients en générant des modèles très compliqués ce qui alourdi la qualité de temps de réponse au requêtes.

1. <https://www-01.ibm.com/software/rational/uml/products/> (dernière visite le 24/05/2018)

Une solution hybride est proposée par la plupart des concepteurs, appelé *starflake*, tente de combiner les deux types de schémas précédents c'est-à-dire uniquement certaines tables représentant des niveaux d'agrégation sont normalisées. Ici, c'est l'expérience du concepteur qui définit les choix entre le coût de stockage et le temps de réponse aux requêtes.

La modélisation physique correspond à trouver les meilleures solutions techniques pour optimiser le stockage et le temps de réponse aux requêtes analytiques. Les auteurs dans (Bellatreche et al., 2004) proposent des indexes de jointure qui permettent d'accélérer l'exécution des requêtes. Dans (Bellatreche, 2000), l'auteur a proposé la technique de fragmentation qui consiste à diviser les tables volumineuses en plusieurs fragments. Néanmoins, cette technique nécessite de choisir des méthodes pour la réécriture et l'exécution de requêtes. Les auteurs dans (Barr et Bellatreche, 2012), ont proposé une technique basée sur les colonies de fourmis pour obtenir le schéma de fragmentation horizontale qui minimise le coût global de la charge des requêtes.

Dans (Ziyati, 2010), l'auteur a proposé une technique d'optimisation inspirée des algorithmes génétiques pour trouver des solutions à des problèmes d'optimisation globale des requêtes OLAP en Entrepôts de Données.

1.6 Limite des systèmes OLAP

Les systèmes OLAP ont montré leur efficacité notamment pour assurer l'analyse multidimensionnelle des données entreposées, non volatiles et à différents niveaux de granularité. Ils sont considérés comme des outils d'aide à la décision. Néanmoins, comme nous l'avons indiqué dans la section précédente, ils présentent aussi des difficultés remarquables aux différents stades de modélisation notamment conceptuelle, logique et physique.

En effet, les systèmes OLAP ne font pas la mise à jour des données, mais seulement l'archivage des versions. Nous nous retrouvons parfois devant des situations de gestion de *Big Data*. Ceci est relatif à la périodicité de l'application du processus ETL.

Les techniques et méthodes proposées pour résoudre ces problèmes sont efficaces dans quelques cas. Dans les autres cas, ces propositions font complexifier encore ces outils destinés au départ pour l'analyse rapide et facile des données.

En plus, d'autres problèmes se posent relatives à la qualité des analyses effectuées par les utilisateurs. Cette qualité dépend de plusieurs facteurs notamment la consistance des données entreposées, la consistance des agrégations et la consistance de l'exploitation c'est-à-dire la formulation des requêtes. Ces problèmes sont traités en détail dans le chapitre 4 notamment avec les systèmes spatial OLAP.

En outre, les EDs traditionnels ne permettent pas la gestion des données géographiques. Or, la grande partie des données sont géo-référencées portant des attributs géométriques, en particulier dans les problématiques d'aide à la décision territoriales et la gestion des ressources naturelles notamment l'agroforesterie.

A cet effet, le recours aux outils permettant la manipulation de ce type de données est incontournable pour la mise en place de systèmes d'information décisionnels spatiaux. Dans cette situation, nous parlons des systèmes d'information géographiques et le spatial OLAP qui feront l'objet du chapitre suivant.

A ces limites, s'ajoutent aussi le fait que les systèmes OLAP ne peuvent pas résoudre les problématiques avec des exigences conflictuelles. A cet effet, ils peuvent être couplés à des outils d'analyse multicritères.

1.7 Conclusion

Dans ce chapitre, nous avons décrit les notions concernant les SID, les ED et l'OLAP. Puis, nous avons présenté les niveaux de l'architecture d'un système d'information décisionnel basé sur les ED. A ce titre, l'importance de la phase d'intégration de données ainsi que les règles et opérateurs du serveur OLAP sont soulignés.

Nous avons ensuite présenté le modèle multidimensionnel sur lequel se base ces systèmes et les différentes modélisations le concernant à savoir : la modélisation conceptuelle, la modélisation logique et la modélisation physique. En effet, les structures de données (dimension, mesure, hiérarchie, etc.) ainsi que les techniques d'agrégation et types d'additivité des mesures sont présentées.

Nous avons, ensuite, présenté les différents schémas pour les cubes de données à savoir les modèles en étoile, en flocon de neige, en constellation et mixte. Enfin, nous avons présenté les différentes architectures logicielles avec les trois types d'implémentations possibles de cette architecture (ROLAP, MOLAP et HOLAP). Les techniques d'optimisation des performances de modélisation physiques sont aussi présentées.

Par la suite, nous avons présenté un état de l'art relatif aux travaux effectués pour la modélisation conceptuelle, logique et physique des EDs.

Enfin, nous avons exploré quelques limites des systèmes OLAP relatives en particulier aux problèmes d'optimisation de stockage et du temps de réponse aux requêtes en présence des entrepôts de données très volumineuses d'une part, et aux limites de gérer les données géographiques d'une autre part.

Dans le chapitre suivant, nous allons présenter les Systèmes d'Information Géographique (SIG), les entrepôts de données spatiales et SOLAP. Ces systèmes, sur lesquels portent nos travaux de thèse, représentent des extensions des systèmes OLAP pour la prise en compte de l'information spatiale dans l'analyse multidimensionnelle.

Chapitre 2

Systemes d'Information Géographique et Spatial OLAP

Dans ce chapitre, nous allons présenter les principales caractéristiques de l'information géographique. Nous allons décrire ensuite, les notions se rapportant aux systèmes d'information géographique à savoir les concepts de base, les structures de données, l'analyse, la visualisation et les solutions logicielles disponibles. Par la suite, nous allons décrire l'extension spatiale des entrepôts de données (ED) et des systèmes OLAP donnant naissance aux EDS et SOLAP.

Sommaire

2.1	Introduction	24
2.2	L'aide à la décision spatiale	24
2.2.1	Types de décisions spatiales	24
2.2.2	Processus de prise de décision spatiale	25
2.3	L'Information Géographique : Concepts de base	26
2.3.1	Composantes de l'information géographique	26
2.3.2	L'objet géographique	27
2.3.3	Systèmes de coordonnées	27
2.3.4	Notion de l'échelle	29
2.4	Systemes d'Information Géographique (SIG)	30
2.4.1	Définition	30
2.4.2	Structure de données dans un SIG	31
2.4.3	Les principales fonctions d'un SIG	33
2.4.4	Logiciels SIG	36
2.5	Limite des systèmes d'information géographique	37
2.6	Travaux utilisant les SIG dans la foresterie	37
2.7	ED Spatial et SOLAP	39
2.7.1	Modèle spatio-multidimensionnel	39
2.7.2	Opérateurs de navigation SOLAP	41
2.7.3	Fonctions d'Agrégation dans le SOLAP	41
2.7.4	Architecture d'un système décisionnel basée sur SOLAP	42
2.7.5	Modes d'intégration SIG-OLAP	43

2.7.6	Modélisation spatio-multidimensionnelle : Etat de l'art	44
2.8	Limite des systèmes SOLAP	45
2.9	Conclusion	46

2.1 Introduction

Dans le but de résoudre des problèmes spatiaux, les décideurs utilisent les systèmes d'aide à la décision spatiale (SDSS : Spatial Decisional Support System) dans l'objectif d'intégrer les capacités de la modélisation analytique et spatiale. Les données à caractère spatial et non spatial sont donc à gérer correctement pour enfin bénéficier des capacités de l'exploration géographique des données.

L'utilisation de l'information à référence spatiale pour soutenir les processus décisionnels est de plus en plus importante grâce à l'évolution de nouvelles technologies. En particulier dans les problématiques telles que la gestion environnementale, le système d'information géographique (SIG) offre une contribution consistante aux domaines de recherche relatives à l'analyse spatiale et l'aide à la décision spatiale. Pour cela, une variété de perspectives théoriques et méthodologiques a été suggérée au cours des dernières années.

2.2 L'aide à la décision spatiale

Prendre une décision peut être interprété par le fait de choisir une solution parmi plusieurs alternatives. Les choix potentiels sont faits pour répondre à un ensemble d'objectifs définis au préalable.

Les décisions qui se portent sur l'espace peuvent être liées aux problèmes de localisation des entités spatiales tels que l'emplacement d'une ressource naturelle ou bien les sites sujets de différentes actions. Elles peuvent aussi concerner l'établissement de plans d'aménagement ou de gestion de risques, etc.

Dans ce cas, l'information géographique est cruciale pour la prise de décision, suite à l'importance de la présentation cartographique de la solution du problème d'une part, et à la quantité d'informations spatiales disponible grâce au développement de matériel et de logiciels liés aux applications spatiales d'une autre part.

2.2.1 Types de décisions spatiales

Les décisions liées à l'information spatiale peuvent se présenter dans la majorité des problématiques qui portent sur le territoire. Il existe une multitude de types des problèmes dans ce sens. Dans la littérature, nous retrouvons plusieurs études qui se sont intéressées à la catégorisation des décisions spatiales.

En effet, selon le Comité sur la Fondation géographique pour l'Agenda 21¹, les décisions spatiales peuvent se classées en 3 principales catégories (Jensen et al., 2002 ; Sugumaran et De Groote, 2011) :

1. Plan d'action de la conférence des Nations unies sur l'environnement et le développement, connue sous le nom de sommet de la Terre de Rio, tenue à Rio de Janeiro au Brésil du 5 au 30 juin 1992

1. les décisions d'allocation des ressources, qui exigent des considérations techniques. Par exemple dans le but de placer un poste de contrôle dans la forêt.
2. les décisions sur l'état des ressources, qui nécessitent des informations spatiales opportunes. Par exemple dans le but de savoir l'état de santé de la végétation ou le stade de maturité.
3. les décisions politiques, par exemple dans le but d'aider la population des montagnes, les gouvernements cherchent convenables à implanter les installations nécessaires à la vie.

Dans l'étude présentée par (Kemp, 2008), les décisions spatiales sont organisées en quatre types :

1. la sélection du site : qui nécessite la prise en compte de différents facteurs à référence spatiale.
2. l'allocation de site, d'où l'objectif principal de la décision est de placer une ressource dans un emplacement optimale pour faciliter et optimiser son utilisation.
3. la sélection de l'utilisation des terres : pour une parcelle donnée, quelle serait l'utilisation idéale.
4. l'affectation des sols, pour planifier à un ensemble de parcelles de différentes caractéristiques, quelles pratiques à appliquer pour certaines fins.

La décision spatiale nécessite des informations multi-sources. Simon (Simon, 1960) classifie les décisions comme étant structurées, semi-structurées ou non structurées.

Les décisions spatiales peuvent être classifiées comme semi-structurées, ce qui signifie qu'elles peuvent présenter des connaissances insuffisantes et des difficultés à quantifier les données. Elles sont aussi caractérisées par l'incertitude et les conflits entre les différents intervenants (Wang et Cheng, 2006). Dans la plupart des cas, la solution de ce type de décisions n'est pas unique.

Par exemple, les décisions potentielles prises pour améliorer la qualité de l'eau des lacs nécessitent une prise en compte des informations spatiales ; en plus des considérations impliquant de nombreux intervenants et approches scientifiques multidisciplinaires. Ce qui peut rendre incertain l'information analysée ainsi que la décision prise.

2.2.2 Processus de prise de décision spatiale

Le processus décisionnel correspond aux différents étapes dont lesquelles le(s) décideur(s) doivent en passer en partant de la définition du problème jusqu'à arriver à la solution finale.

Simon (Simon, 1960) a suggéré que ce processus soit itératif et présente quelques fois des retours en arrière. Il passe par trois phases :

- **Formulation du problème** : qui correspond à la recherche d'informations pertinentes pour trouver des solutions au problème.
- **Exploitation** : l'analyse des données et des informations pour arriver à la solution.
- **Choix** : dans la dernière phase, la sélection des alternatives est faite

A la fin de ces phases, peut s'ajouter une phase d'évaluation. Il s'agit d'une étape de communication permettant l'explication et la motivation du choix ainsi que le suivi de son application (Feeney et Williamson, 2002).

Le processus de prise de décision est modélisé par Huber (Huber, 1989) comme un processus constitué d'un ensemble d'étapes, telles que l'identification d'un problème, la génération d'actions, l'identification des conséquences des actions, et la sélection d'une alternative.

Ayant des attributs particuliers à ce type de processus, les problèmes qui nécessitent des considérations spatiales sont complexes, multidimensionnels, caractérisés par des aspects d'incertitude et impliquent souvent de nombreux intervenants.

Pour ce fait, il est nécessaire de suivre un processus itératif avec diverses interactions entre les intervenants. Dans le cas des questions décisionnelles sur le territoire, l'objectif global étant de réduire des conséquences environnementales et économiques négatives.

Keller (Keller, 1989) a proposé un autre processus traitant la décision spatiale. Il est composé de cinq étapes notamment : 1) identifier le problème, 2) rassembler les données nécessaires, 3) définir le problème, y compris les objectifs, les hypothèses et les contraintes, 4) trouver des procédures de la solution, et 5) résoudre le problème en trouvant une solution optimale.

Selon, Keller (Keller, 1989), la minimisation des objectifs en un seul objectif global, ainsi que la définition des procédures, constituent les étapes les plus importantes du processus décisionnel.

Dans certains cas, il est nécessaire de définir plusieurs scénarios de solution. Par conséquent, une combinaison de modèles sera nécessaire pour aider à évaluer efficacement ces différents scénarios avec la participation de tous les intervenants.

Hamdadou (Hamdadou, 2008), a proposé un modèle d'aide à la décision collective spatiale dédié pour la résolution de la problématique de localisation en aménagement du territoire. Le modèle proposé est basé sur l'analyse multicritères et l'utilisation des protocoles de négociation.

2.3 L'Information Géographique : Concepts de base

L'information géographique peut être définie comme « *la représentation d'un objet ou d'un phénomène réel ou imaginaire, localisé dans l'espace à un moment donné* » (Degréne et Salgé, 1997). Dans cette section nous présentons quelques concepts de base de ce type d'information.

2.3.1 Composantes de l'information géographique

L'information géographique est caractérisée par une composante purement spatiale et une composante sémantique :

a) *La composante spatiale*

La composante spatiale de l'information géographique correspond à la position sur la surface terrestre et la forme de l'objet ou du phénomène représenté. La position

est décrite dans un système de référence (système de coordonnées). La forme décrit la morphologie ou le périmètre de l'information représentée (Bimonte, 2007).

b) *La composante sémantique*

La composante sémantique de l'information géographique correspond à la nature, l'aspect et les propriétés descriptives de l'objet ou phénomène représenté. Cette représentation peut être multiple et à différentes échelles. Elle peut contenir des relations avec d'autres objets ou phénomènes (Weibel et Dutton, 2001).

Nous devons souligner, par ailleurs, que l'espace géographique est un espace qui évolue. A cet effet, la représentation de l'information géographique ne doit pas négliger la composante temporelle car une description statique de l'espace géographique ne devient plus suffisante.

2.3.2 L'objet géographique

Un objet géographique peut être défini comme le support de l'information géographique ou le référentiel permettant de positionner d'autres objets. Il est considéré géographique lorsqu'il est localisable directement par des coordonnées, ou indirectement par des données de type adresse, numéro de commune, numéro de borne kilométrique, code postal, numéro de parcelle cadastrale, etc.

2.3.3 Systèmes de coordonnées

Afin de localiser un objet sur la terre, l'utilisation d'un système géodésique² est nécessaire, ce qui permet d'obtenir les coordonnées géographiques : longitude, latitude et altitude (λ , φ , h).

Les coordonnées terrestres

Le Géoïde est la forme théorique qui se rapproche le plus de la surface réelle de la terre c'est-à-dire le niveau moyen des mers. Il sert de référence pour déterminer les altitudes (h), comme le montre la Figure 2.1.

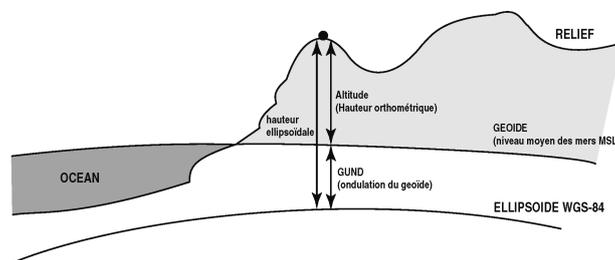


FIGURE 2.1 – La forme théorique de la terre (géoïde)

2. <http://eduscol.education.fr/localisation/pedago/geologie/coordonnees.htm> (dernière visite le 24/05/2018)

L'ellipsoïde est la surface mathématique qui se rapproche le plus de la forme du géoïde ; grâce à ce dernier, nous pouvons calculer les coordonnées géographiques en LONGITUDE (λ) et en LATITUDE (φ) comme illustré dans la Figure 2.2

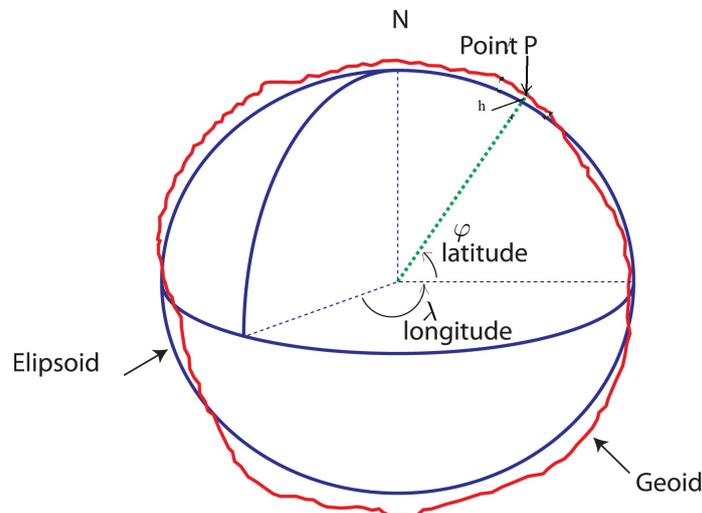


FIGURE 2.2 – Notion de l'ellipsoïde

Coordonnées planes et projection cartographique

La représentation plane utilise les coordonnées dites en "*projection*", issues du passage de l'ellipsoïde au plan. Ceci a pour but d'obtenir des valeurs métriques plus exploitables que les unités angulaires, en particulier pour la mesure des surfaces. Néanmoins, cette transformation va engendrer des déformations, et nous aurons donc trois choix (Bedel, 2009) :

- a) *conserver les surfaces* (projections équivalentes), utilisée souvent en cartographie à petite échelle.
- b) *conserver localement les angles* (projections conformes), utilisée en géodésie³ et topographie.
- c) *ne conserver ni les angles ni les surfaces* (projections dites "aphylactiques"), utilisée généralement en cartographie thématique.

La projection passe généralement par la représentation de la totalité ou une partie de l'ellipsoïde sur une surface développable. Les trois formes mathématiques qui répondent à ce critère (à savoir le plan, le cylindre et le cône) donnent lieu aux trois types principaux de projections illustrées par la Figure 2.3 :

- a) **Projection cylindrique** : l'ellipsoïde est projeté sur un cylindre qui l'englobe. Celui-ci peut être tangent au grand cercle, ou sécant en deux cercles. Puis on déroule le cylindre pour obtenir la carte.

3. la science de la forme et de la dimension de la terre et du champs de pesanteur

Dans la littérature, diverses projections cylindriques sont utilisées notamment : Projection de *Mercator* (conforme), Projection de *Peters* (équivalente), Projection de Robinson (pseudo-cylindrique, aphyllactique), Projection UTM⁴ (conforme),... etc.

- b) **Projection conique** : l'ellipsoïde est projeté sur une surface conique tangente à une ellipse ou sécant en deux ellipses. Puis on déroule le cône pour obtenir la carte.
Différentes projections coniques sont utilisées dans les travaux de recherche à savoir : Projection conique conforme de *Lambert*, Projection d'*Albers*,... etc.
- c) **Projection azimutale** : l'ellipsoïde est projeté sur un plan tangent en un point ou sécant en un cercle.

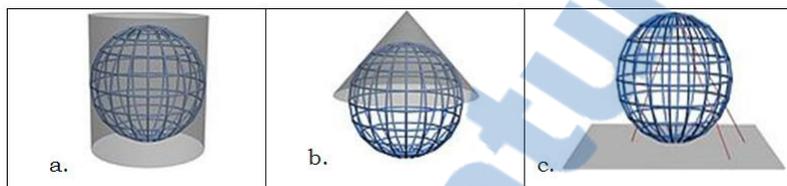


FIGURE 2.3 – Types de projection a : cylindrique, b : conique, c : azimutale

2.3.4 Notion de l'échelle

L'échelle d'une carte est le rapport constant qui existe entre les lignes de la carte et les dimensions réelles du terrain représenté.

- a) **Echelle numérique** : L'échelle s'exprime par une fraction $1/N$. Par exemple l'échelle $1/10000$ signifie 1 cm sur carte correspond à 10000 cm sur la terre.
- b) **Echelle graphique** : La carte ou le plan est accompagné d'une représentation graphique correspondant à une ligne divisée en parties égales. Voir Figure 2.4



FIGURE 2.4 – Echelle graphique

Pour résumer, nous constatons que la représentation d'une portion de la terre sur un plan selon une échelle, est une problématique qui nécessite le passage par une modélisation mathématique. Cette dernière est basée sur la projection de la réalité sur une surface dite développable, permettant ainsi le passage vers le plan. A cet effet, la forme utilisée pour la projection est donc de type conique, cylindrique ou plan.

chaque point de la portion de la terre, connu en longitude (λ) et latitude (ϕ), est représenté en coordonnées (X , Y). la fonction mathématique f qui permet ce passage

4. <https://gisgeography.com/utm-universal-transverse-mercator-projection/> (dernière visite le 24/05/2018)

$((X, Y) = f(\lambda, \varphi))$, est choisie de telle sorte à minimiser les déformations résultant du passage de 3 dimensions à 2 dimensions.

Pour ce fait, il y a des projections dites conformes qui conservent les angles ; des projections dites équivalentes qui conservent le rapport des surfaces ; et celles qui ne conservent ni les angles ni les surfaces dites aphyllactiques. Le choix entre ces modèles dépend généralement de l'objectif de la cartographie et de la surface à représenter.

2.4 Systèmes d'Information Géographique (SIG)

Les systèmes d'information géographique (SIG) sont des systèmes d'information qui permettent l'organisation et la manipulation des données à référence spatiale telles que des coordonnées géographiques, un code postal, ... etc. Ils permettent aussi l'interrogation et la présentation cartographique des résultats d'analyse spatiale de différents phénomènes réels ou imaginaires. Dans la suite de cette section, nous allons décrire quelques notions relatives à ce type de système d'information.

2.4.1 Définition

Il existe de nombreuses définitions pour les systèmes d'information géographique. Voici quelques unes :

— Une définition américaine, donnée par FICCDC-1988 (le comité fédéral de coordination inter-agences pour la cartographie numérique) :

« Système informatique de matériels, de logiciels et de processus conçus pour permettre la collecte, la gestion, la manipulation, l'analyse, la modélisation et l'affichage de données à référence spatiale afin de résoudre des problèmes complexes d'aménagement et de gestion »

— Une définition donnée par l'économiste Michel Didier pour le conseil national de l'information géographique (CNIG) (français) :

« Ensemble de données repérées dans l'espace, structuré de façon à pouvoir en extraire commodément des synthèses utiles à la décision. »

D'une autre façon, nous pouvons dire que les systèmes d'information géographique sont une famille de systèmes d'information conçus spécifiquement pour la manipulation et la visualisation des données géoréférencées.

Le SIG est un système pour produire des cartes, intégrer les informations, visualiser des scénarios. Les rôles d'un SIG peuvent se résumer dans les 5A qui permettent les opérations suivantes (Pourabbas, 2014) :

- *Abstraction* : la conception du monde réel
- *Acquisition* : la donnée brute et l'information géographique,
- *Archivage* : gestion de base des données spatiales,
- *Analyse* : manipulations spatiales, fonctions élémentaires et complexes,
- *Affichage* : mise en forme et visualisation cartographique,

L'utilité première d'un SIG est sa capacité de répondre à des questions d'ordre spatial. Nombre de ces questions se résument à deux types d'interrogation : (Paegelow, 2004)

- Qu'est-ce qui se trouve à un endroit ?
- Quelle est la position d'un objet ?

2.4.2 Structure de données dans un SIG

Un SIG stocke généralement deux types d'informations. La première est une information géographique ou spatiale alors que la seconde est descriptive portant des données d'attribut sur ces objets. Afin d'utiliser le SIG nous devons organiser cette information de manière efficace. Pour ce faire, nous devons utiliser des structures de données qui peuvent être définies comme des moyens d'organiser l'information.

Il existe deux structures de données utilisées dans les SIG : l'approche vectorielle et l'approche matricielle. Voir Figure 2.5 retirée du site d'ESRI⁵.

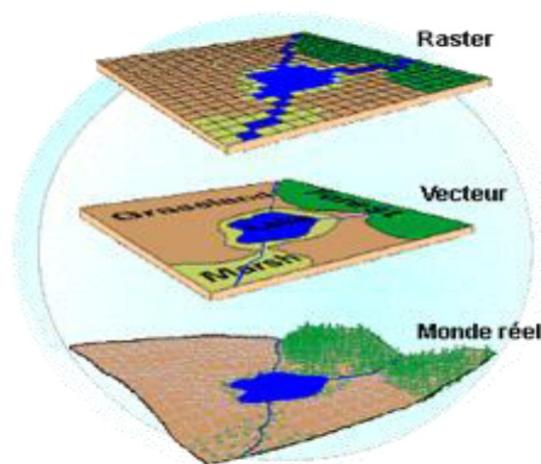


FIGURE 2.5 – Représentation vectorielle et matricielle

Modèle vectoriel

Dans l'approche vectorielle, les objets au sol sont représentés à l'aide de trois types géométriques : les points, les lignes et les zones. Cette représentation concerne la composante thématique qui regroupe des attributs non spatiaux ; et la composante spatiale qui comprend la position ainsi que les limites de l'objet.

La description géométrique dans le modèle vectoriel peut se faire grâce à deux modèles qui prennent en compte ou non la topologie du voisinage notamment le modèle spaghetti et le modèle topologique.

Dans le modèle spaghetti dit aussi « métrique », les géométries (lignes ou surfaces) sont décrites indépendamment. Ce modèle est simple à mettre en œuvre ; néanmoins, il présente quelques inconvénients relatifs aux chevauchements ou des parties non connectées entre les objets de même type. (Bedel, 2009).

5. <https://www.esrifrance.fr/sig3.aspx> (visité le 27/01/2018)



Dans le modèle topologique, nous pouvons considérer les connexions entre les géométries grâce à :

- *La topologie de réseau* : dans laquelle les connexions peuvent se matérialiser entre les lignes. En effet, chaque ligne est orientée et délimitée par un nœud initial et un nœud final. Elle est utilisée par exemple pour modéliser un réseau routier.
- *La topologie zonale* : dans laquelle les connexions peuvent se matérialiser entre les des surfaces. Ceci peut se réaliser par un graphe reliant des faces, des arcs et des nœuds. Elle est utilisée par exemple dans la décomposition cadastrale. (Bedel, 2009).

Modèle matriciel (raster)

Dans l'approche matricielle, la zone d'intérêt est divisée en cellules de même étendue géographique. Les cellules sont ensuite codées en fonction des propriétés de la zone qu'elles couvrent. Ce type de modélisation permet de considérer la continuité d'un phénomène propre à une zone géographique. Elle peut par exemple modéliser l'altitude moyenne dans un modèle numérique de terrain (MNT) ou la réponse spectrale dans une image satellitaire.

Dans les systèmes d'information, l'organisation des informations de différents types passe par leur regroupement structuration en couches d'information. Ces couches décrivent l'ensemble des éléments relatifs aux différents thèmes étudiés d'une zone géographique déterminée. Par exemple comme illustré dans la Figure 2.6 retirée du site de ESRI⁶, pour représenter les données du monde réel, nous pouvons utiliser différentes couches relatives à : l'occupation du sol, altitude, parcelles, rues et les adresses des clients

Comparaison entre les deux modèles (raster et vector)

Les modèles de données matricielles et vectorielles présentent tous deux des avantages et des inconvénients selon le phénomène traité, (Bedel, 2009).

En effet, le modèle de données vectorielles correspond mieux à certains cas où les entités traitées sont bien différenciées de leur voisinage comme le réseau routiers, zone bâtie... etc.

En revanche, la représentation avec le modèle de données matricielles est plus adaptée pour décrire les phénomènes qui varient continuellement dans l'espace. Ceci correspond par exemple au modèle numérique du terrain (élévation) la température, ... etc.

Dans certains cas, nous sommes arrivés à modéliser des phénomènes qui peuvent être classifiés dans les deux possibilités par exemple le cas de l'occupation de sol. Dans ce cas le choix dépend de la nature des sources de données collectées. Par exemple la représentation des limites issues de l'image satellitaire sera mieux adaptée au modèle raster ; par contre le vectoriel sera mieux si la source était la photogrammétrie. (Sugumaran et DeGroot, 2011).

Pour cela, une application SIG doit bénéficier des atouts des deux possibilités comme la plupart des logiciels SIG permettent l'affichage de données vectoriels et raster, et offrent des fonctionnalités de conversion entre les deux modèles de représentation.

6. <https://www.esrifrance.fr/sig3.aspx> (visité le 27/01/2018)

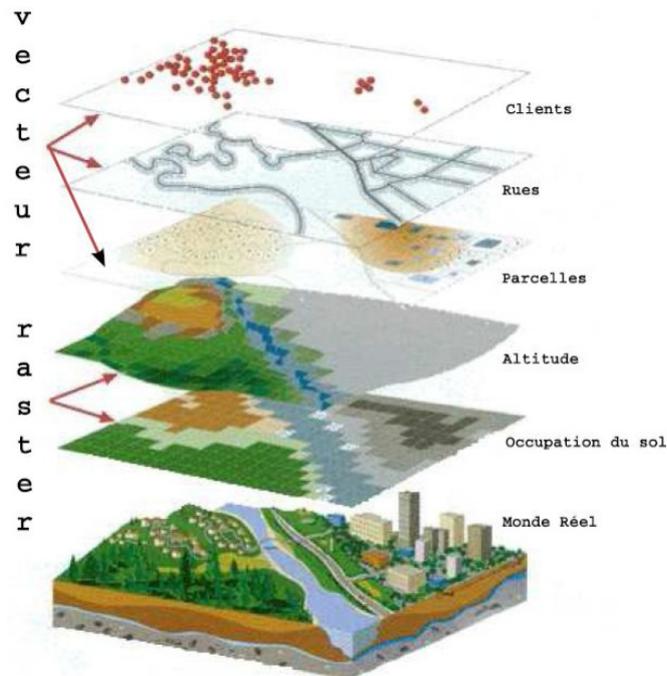


FIGURE 2.6 – Organisation en couches thématiques

2.4.3 Les principales fonctions d'un SIG

Dans cette section, nous allons décrire quelques fonctions clés d'un SIG.

Intégration de données

Les données relatives à des lieux ou entités géographiques bien définis (points, lignes ou zones) peuvent être intégrées au SIG, quelle que soit leur origine. En effet, lorsque les enregistrements sont géo-référencés dans un format accepté, ils peuvent être importés dans un SIG. Cependant, il n'est pas toujours nécessaire que les données aient eux-mêmes un identifiant spatial à intégrer dans le SIG. En effet, certains ensembles de données peuvent être intégrés, sur la base d'un champ commun, à une référence relative (code postal). Cette jonction de tables est extrêmement significative dans le contexte d'une opération de création et d'intégration de données (Fediaevsky, 2002).

Analyse de données dans un SIG

L'analyse de données peut se faire en différentes techniques.

a) Requêtes

Les données peuvent être interrogées de deux façons principales dans un SIG :

1. *Sur la base de l'emplacement* : les enregistrements peuvent être sélectionnés en fonction de l'endroit où ils se trouvent, ce qui est unique pour le SIG.

2. *Sur la base de l'attribut* : comme tout système de gestion de base de données, des requêtes structurées peuvent être appliquées dans un SIG à un ensemble de données pour extraire des enregistrements sur la base de leurs caractéristiques, indépendamment de leur emplacement.

b) Analyse spatiale

L'analyse spatiale porte sur les modèles et les associations qui en existent et entre les couches d'information spatiales, (Fediaevsky, 2002).

1. *Superposition de couches d'information* : Il s'agit de superposer une ou plusieurs couches thématiques à une autre. À son niveau le plus simple, cela peut être un processus visuel, pour voir comment la distribution d'un ensemble de caractéristiques se rapporte à une autre.
2. *Analyse booléenne* : Il s'agit de requêtes plus complexes utilisant ce qu'on appelle des opérateurs booléens tels que 'AND, OR et EXCLUSION'. Une approche combinée peut également être appliquée pour identifier les enregistrements qui répondent à certains critères d'attribut ainsi qu'à des critères spatiaux.
3. *Zone de buffer (tampon)* : La mise en zone de tampon est la création d'une entité polygonale autour d'une entité déjà existante (points, lignes ou polygones). Une fois créés, les tampons peuvent être utilisés dans toutes les autres opérations décrites ci-dessus.
4. *Analyse de voisinage* : L'analyse de voisinage permet une approche plus interactive pour définir les entités en fonction de leur relation spatiale avec d'autres entités.
5. *Analyse thématique* : Il s'agit dans ce cas de réaliser une analyse sur les attributs des entités spatiales. Elle est aussi appelé 'statistique spatiale'. Elle peut générer différentes cartes en se basant sur les variables visuelles.
6. *Analyse de réseau* : L'analyse de réseau se concentre sur le mouvement des objets le long d'un réseau, de routes ou de chemins reliés entre eux. L'analyse comprend un certain nombre de paramètres, notamment : la connectivité, la vitesse et le sens.

c) Modélisation des données

La modélisation est le processus de génération des éléments additionnels à partir de jeux de données existants. Par exemple, pour concevoir un modèle de dispersion de pollution atmosphérique, nous devons prendre en considération l'ensemble des paramètres tels que l'emplacement du rejet, la force et la direction du vent et les caractéristiques du terrain ... etc. Cette capacité à exécuter différents scénarios et à évaluer leurs implications probables est essentielle aux activités associées à la préparation aux situations d'urgence. Donc, si les données et les modèles sont robustes, les résultats sont potentiellement très utiles aux gestionnaires.

d) Analyse du terrain

L'analyse de terrain est l'ensemble des techniques qui analysent la surface de la terre pour identifier les cartes de pente ou d'inter-visibilité qui sont par exemple pertinentes pour la modélisation hydrologique.

Cette analyse est généralement effectuée sur un Modèle Numérique de Terrain (MNT) qui est une couche de données matricielles dans laquelle chaque pixel reçoit l'altitude moyenne du terrain par rapport à référentiel des altitudes (ellipsoïde). Ceci peut être enrichi par une autre information relative à la hauteur des objets de sursol, connue sous le nom de 'Modèle Numérique d'Élévation' (MNE), voir Figure 2.7 retirée du site d'ArcGis⁷.

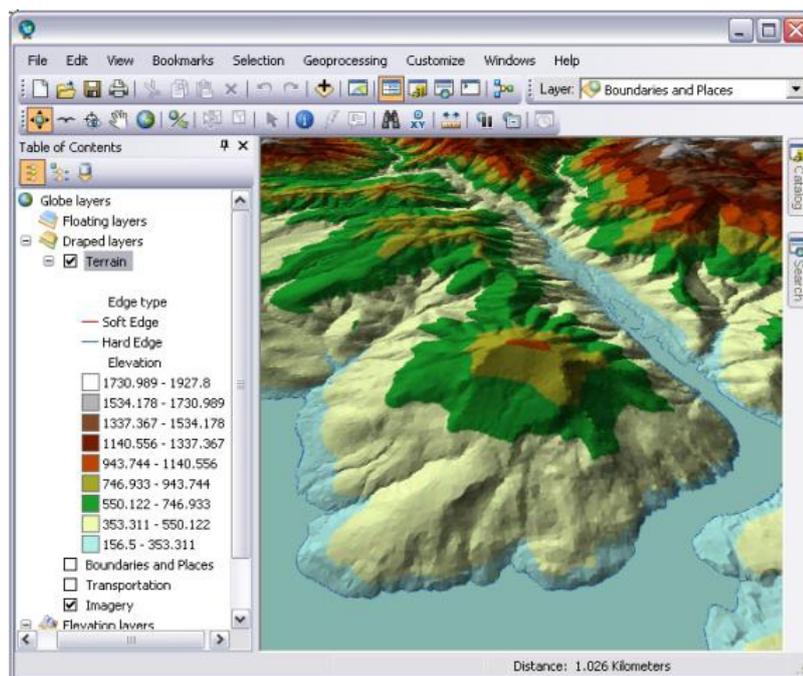


FIGURE 2.7 – Modèle Numérique de Terrain (MNT)

Nous devons aussi noter les nouvelles tendances des SIG 3D. Il s'agit de l'exploitation de la 3ème dimension dans l'analyse spatiale. Ils sont beaucoup plus destinés au domaine de la planification urbaine. Néanmoins, quelques applications SIG 3D sont élaborées pour la gestion forestière comme le montre la Figure 2.8 retirée du site d'ESRI⁸.

En résumé, nous pouvons constater qu'un SIG se base en particulier sur l'organisation des données géographiques selon leurs thématiques, en une multitude de couches d'information vectorielles ou matricielles, similaires aux tables de bases de données. En effet, cette organisation des données permet la production de différentes cartes résultant de la superposition des couches ou bien l'analyse spatiale offerte par les logiciels SIG.

7. <http://desktop.arcgis.com/fr/arcmap/10.3/manage-data/terrains/about-displaying-terrain-datasets-in-arcgis.htm>, (visité le 27/01/2018)

8. <https://www.esrifrance.fr/sig3.aspx> (visité le 27/01/2018)

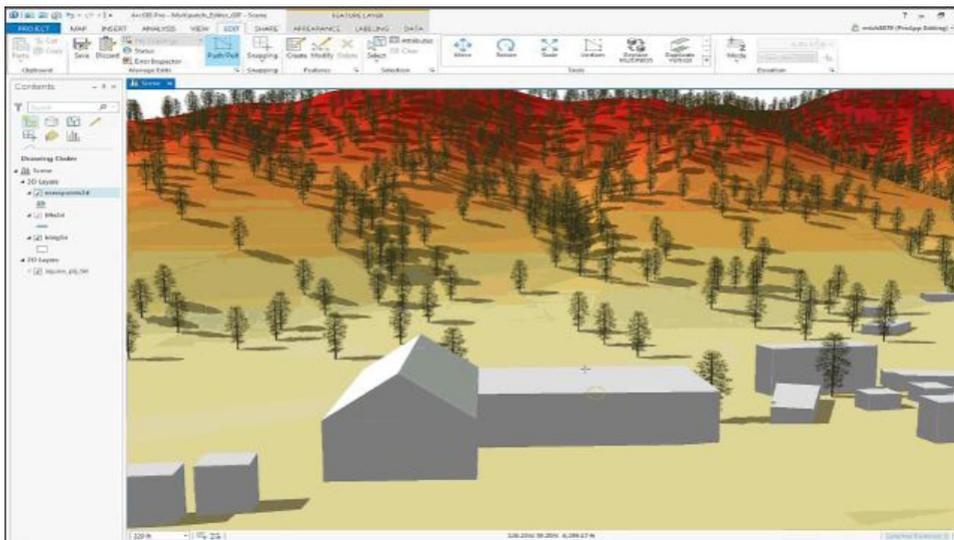


FIGURE 2.8 – Exemple de SIG 3D

Cependant, cette organisation de données qui est prédéfinie par le producteur, présente quelques limites relatives en particulier à la description des objets d'une couche en même niveau de détail. Ceci peut être corrigé par les techniques de généralisation cartographique. Aussi, l'utilisateur doit connaître la structuration des données pour pouvoir exprimer les relations dans les tables de bases de données (jointure par exemple). Des approches de représentation de connaissances orientées-objet telles que 'Géodatabase' proposée par ESRI, permettent de répondre à certains limites.

En plus, la plupart des logiciels SIG proposent le couplage avec les SGBD (exp : Oracle, SQL Server), pour bénéficier des atouts de ces derniers, en matière de gestion des accès concurrents et la vérification des requêtes.

Pour pouvoir être qualifié d'un outil d'aide à la décision, un SIG doit assister l'utilisateur dans sa recherche d'information. A ce titre, nous devons souligner l'importance de la navigation interactive des cartes qui peut faciliter la recherche d'information, à travers les différents opérateurs géométriques disponibles dans certains logiciels SIG.

2.4.4 Logiciels SIG

Les systèmes d'information géographique (SIG) ont largement évolués depuis leur naissance. Cette évolution leur permettait de passer d'une simple architecture basée sur les stations de travail, à des architectures plus complexes (orientée service).

Un SIG doit généralement se baser sur des logiciels permettant la manipulation, le stockage et la diffusion des données. Pour cela, Il existe une grande variété de programmes SIG qui peut être classée en deux grandes familles notamment : SIG commerciaux et SIG libres (open source) disponibles avec les différentes fonctionnalités.

A titre d'exemple, nous citons quelques solutions SIG commerciales les plus utilisées, à savoir : ESRI ArcGIS, GeoMedia, MapInfo, IDRISI et autres.

Le logiciel ArcGIS d'ESRI est le produit le plus dominant sur le marché depuis plusieurs années. Il fournit un ensemble très riche de fonctionnalités de stockage, gestion et du traitement des données à référence spatiale.

MapInfo étant un logiciel simple et facile mais son utilisation n'est pas aussi large que le précédent.

GeoMedia est un autre grand acteur du marché des SIG. Il fournit un ensemble important de fonctionnalités pour créer, gérer et analyser les données spatiales.

D'autre part, les logiciels SIG open source sont en forte croissance bien qu'ils ne disposent pas de toutes les fonctionnalités offertes par la solution commerciale.

Les solutions open source les plus populaires incluent GRASS (Geographic Resource Analysis Support System) qui est le plus ancien et le plus complet, Quantum GIS (QGIS), SAGA, et d'autres (Sugumaran et DeGroot, 2011).

2.5 Limite des systèmes d'information géographique

L'importance des SIG est irréfutable dans la gestion et la production des données géo-référencées. Son utilisation est généralisée, en plus de la gestion du territoire et les phénomènes environnementaux, pour atteindre les problématiques de géomarketing et sociaux.

Néanmoins, ils présentent quelques limites concernant plusieurs niveaux. En effet, au niveau des applications SIG mise à la disponibilité des utilisateurs, un problème sérieux qui se pose dans certains cas est lié à des difficultés d'interprétation des résultats. Ceci rend la construction des indicateurs d'aide à la décision, très compliquée et parfois conduit à des conclusions erronées.

En plus, s'ajoutent encore des limites techniques relatives dans certains cas, à l'absence de données, ou bien le coût d'acquisition. Les problèmes de cohérence relative à la précision et la validité spatiotemporelle se posent aussi vu la multitude et l'hétérogénéité des sources de données.

Nous pouvons souligner aussi, que les SIG sont des systèmes de gestion transactionnels. En effet, ils ne permettent pas l'analyse des données à plusieurs axes de réflexion et plusieurs granularités. Pour cela, les SIG peuvent être améliorés par l'intégration des techniques d'analyse multidimensionnelle notamment les systèmes OLAP. Leur utilisation peut être aussi enrichie par l'introduction des techniques d'analyse multicritères dans des situations nécessitant, la prise en compte des exigences conflictuelles.

2.6 Travaux utilisant les SIG dans la foresterie

Les systèmes d'information géographique sont utilisés dans diverses disciplines pour l'exploitation et la production des données spatiales, notamment en superposition des couches d'information, l'analyse des réseaux, les zones de buffer, etc. Le fait que les caractéristiques spatiales de l'agroforesterie soient largement utilisées, plusieurs études SIG spécifiques à l'agroforesterie ont été élaborées.

En effet, plusieurs axes de l'agroforesterie ont été couverts par la contribution des SIG dans le domaine de gestion des ressources naturelles. Nous avons classé ces contributions selon ces axes comme suit :

- Dans le domaine de l'inventaire forestier, les auteurs dans (Warneck et al., 2002) ont effectué une analyse basée sur le SIG pour l'évaluation écologique, socioéconomique et l'amélioration des activités de gestion des ressources forestières.
- Dans le domaine de défense contre les feux de forêts, le SIG a été utilisé par les auteurs dans (Akay et Erdogan, 2017), pour la proposition d'une étude d'analyse décisionnelle basée sur le SIG et l'AMC pour la cartographie des risques de feux de forêts. Les auteurs dans (Belhadj-aissa et al., 2003) ont également proposé une étude basée sur le SIG et la télédétection pour la gestion des feux de forêts en Algérie.
- Dans le domaine de la conservation et biodiversité des forêts, les auteurs dans (Wilson et Lowe, 2003) ont proposé une modélisation basée-SIG pour la planification de scénarios pour la conservation de la biodiversité native. Dans (Wang et Moskovits, 2001), les auteurs ont proposé une étude pour la dérivation de carte d'occupation de sol en utilisant les données multi-temporelles à travers les modèles de SIG et télédétection. Dans (Bentrup et Leininger, 2002), le SIG est utilisé pour faciliter la planification de conservation des forêts, dans le centre des Etats-Unis.
- Dans le domaine de la gestion de l'infrastructure forestière, nous pouvons citer à titre d'exemple, l'étude effectuée par les auteurs dans (Abdi et al., 2009), où une étude basée sur le SIG pour la gestion de l'infrastructure routière au sein des forêts a été proposée.
- Dans le domaine de la gestion de la faune forestière, les auteurs dans (Kilgo et al., 2002) ont effectué une étude pour tester l'exactitude du modèle d'habitat faunique basé sur le SIG BIRDHAB pour prédire la qualité relative de l'habitat des oiseaux au niveau des peuplements forestiers. Dans (Danks et Klein, 2001), les auteurs ont proposé l'utilisation des cartes de la végétation pour déterminer la faisabilité du SIG en tant qu'outil pratique pour l'évaluation et la prévision de l'habitat faunique.

Il faut noter, également, que certains de ces études ont vu une réalisation concrète. Les autres études n'ont pas été réalisées à cause de l'une des limites notées plus haut. Ceci était généralement à cause du coût élevé d'acquisition de données ou autres difficultés techniques.

Au sein du Laboratoire d'Informatique d'Oran (LIO), les travaux effectués sur les systèmes d'aide à la décision spatiale sont nombreux, en particulier ceux qui traitent les problématiques de l'aménagement du territoire.

En effet, quant à l'intégration des systèmes d'information géographique, un modèle décisionnel en aménagement du territoire par utilisation des SIG et les réseaux de neurones a été proposé dans (Younsi et al., 2007). Aussi, dans (Oufella et al., 2009), les auteurs ont proposé la conception d'un système interactif d'aide à la décision collective pour les données spatialement localisées.

2.7 ED Spatial et SOLAP

L'utilisation intensive de nouvelles technologies et applications spatiales, a conduit à une production de gros volume de données qui ont un caractère spatial d'une importance remarquable. Ceci a conduit à l'intégration de ces données dans les processus d'aide à la décision en particulier dans l'analyse multidimensionnelle qui profite du concept de l'Entrepôts de Données Spatiales (EDS) (Abdallah Bensalloua et al., 2015).

Un Entrepôt de Données Spatiales est considéré comme « *une collection de données spatiales et non spatiales orientées sujet, intégrées, variables dans le temps et non volatiles, utilisée pour supporter le processus décisionnel* » (Stefanovic et al., 2000).

Dans le même contexte, les systèmes OLAP sont étendus en présence de la dimension spatiale, pour donner naissance au concept de Spatial OLAP (SOLAP). Ceci permet une exploration interactive des EDS en se basant sur un modèle spatio-multidimensionnelle (Bédard et Han, 2009).

2.7.1 Modèle spatio-multidimensionnel

Le modèle spatio-multidimensionnelle est basé sur un ensemble de concepts étendus du modèle multidimensionnelle.

— **Mesure spatiale :** Une mesure spatiale est une mesure dont le type de données est géométrique et/ou numérique. Par exemple la surface d'une zone géographique de forêt incendiée, la circonférence d'une zone de forêt protégée, la distance entre régions spatiales, etc. Trois types de mesures spatiales sont reconnus dans les différents travaux liés à la géomatique⁹.

1. *mesure spatiale géométrique* : qui est définie par les entités géométriques représentant des objets spatiaux, tels que des "foyers d'incendie".
2. *mesure spatiales numériques* : qui sont des valeurs numériques résultant des opérateurs spatiaux telles que : distance ou superficie.
3. *mesure spatiale complète* : introduite par (Bédard et Han, 2009). Elle est spécifique aux cubes de données raster. Il s'agit d'une combinaison d'une mesure spatiale numérique et géométrique relative à une position de cellule raster et de sa valeur associée. (Rivest et al., 2003)

Exemple de valeur de mesure spatiale : un exemple de mesure spatiale pour une "zone géographique incendié" est montré dans la Figure 2.9 (Abdallah Bensalloua et Hamdadou, 2018b).

— **Niveau d'agrégation spatial**

Un niveau d'agrégation spatial est un niveau d'agrégat : 4 représentant un ensemble de membres. Ceux-ci possèdent des attributs spatiaux qui leur permettent de représenter les données sous forme cartographique. (Bimonte, 2007; Malinowski et Zimányi, 2008).

9. Géomatique : l'ensemble des méthodes qui permettent l'acquisition, la représentation et l'analyse des données à référence spatiale

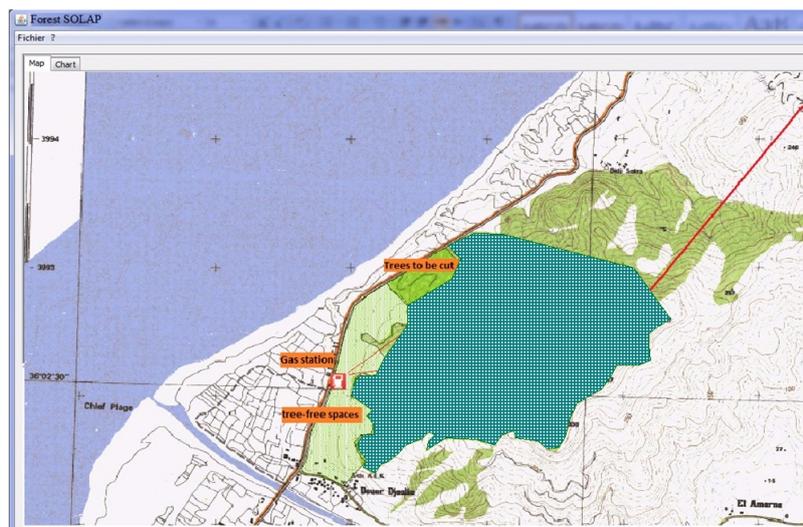


FIGURE 2.9 – Exemple de mesure spatiale

Un niveau d'agrégation spatial peut être géométrique ou non géométrique relatif à une description textuelle.

— **Dimension spatiale**

La dimension spatiale peut être définie comme une dimension avec au moins un niveau d'agrégation spatial. Elle permet de représenter l'information spatiale des faits en axes d'analyse. (Bédard et Han, 2009).

— **Hierarchie de dimension spatiale**

Une hiérarchie de dimension spatiale est définie comme une hiérarchie qui contient au moins un niveau d'agrégation spatial (Boulil, 2012).

— **Fait spatial**

Un fait spatial est défini comme étant un fait contenant au moins une mesure spatiale ou un niveau d'agrégation spatial. (Salehi et al., 2010). Divers autres définitions peuvent exister dans la littérature notamment celles présentées par (Malinowski et Zimányi, 2008) et celle de (Boulil, 2012).

— **Hypercube spatial**

Un hypercube (ou data cube) spatial peut être défini comme un hypercube possédant au moins une mesure spatiale ou une dimension spatiale. (Salehi et al., 2010). Il existe aussi d'autres définitions dans la littérature. Dans cette thèse nous allons considérer qu'un hypercube doit posséder au moins une mesure spatiale et une dimension spatiale pour être qualifié de spatial.

— **Entrepôt de données Spatial (EDS)**

Un Entrepôt de données Spatial est un modèle multidimensionnel contenant au moins un hypercube spatial, (Boulil, 2012); Voir Figure 2.10 (Pinet et Schneider, 2010).

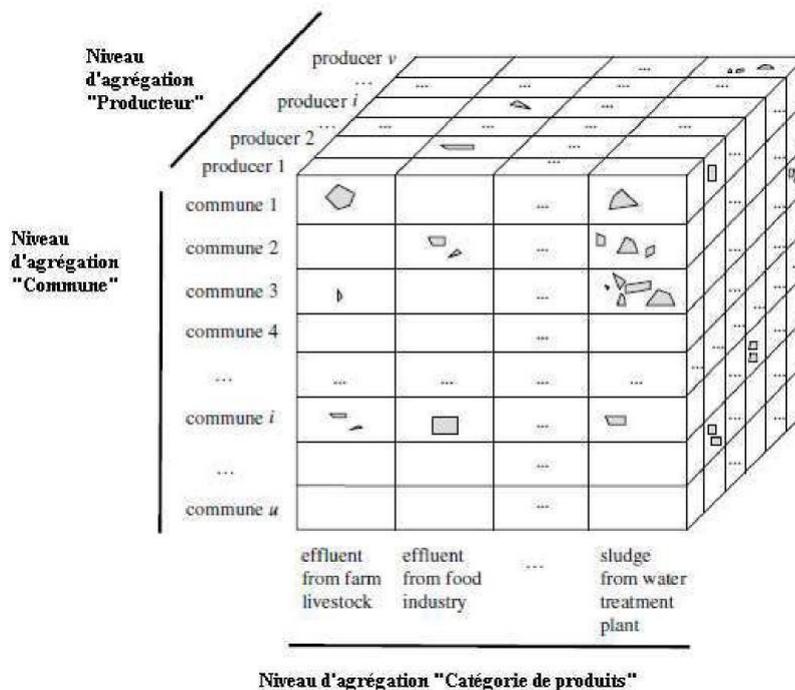


FIGURE 2.10 – Exemple d'instance d'hypercube spatiale

2.7.2 Opérateurs de navigation SOLAP

Pour permettre l'exploration des hypercubes de données spatiales, un ensemble d'opérateurs de navigation doivent être définis. Ces opérateurs sont l'extension des opérateurs OLAP. (Bédard et Han, 2009) :

- **Spatial Roll-up** : Cet opérateur est utilisé pour la navigation dans une hiérarchie de dimension d'un niveau d'agrégation spatial vers un autre moins détaillé.
- **Spatial Drill-down** : Cet opérateur est utilisé pour la navigation dans une hiérarchie de dimension d'un niveau d'agrégation spatial vers un autre moins détaillé.
- **Spatial Slice** : cet opérateur est utilisé pour la sélection d'un sous-ensemble des cellules de l'hypercube spatial par l'application d'un prédicat spatial sur les membres d'une dimension spatiale.
- **spatial Dice** : cet opérateur est utilisé pour la sélection d'un sous-ensemble des données par l'application des prédicats spatiaux à des membres spatiaux de plusieurs dimensions spatiales. (Boulil, 2012)

2.7.3 Fonctions d'Agrégation dans le SOLAP

L'application des fonctions d'agrégation spatiale sur des mesures spatiales est une opération fondamentale pour l'analyse spatio-multidimensionnelle. Ce qui permet de retourner un ensemble d'objets géométriques simple ou complexes.

Pour garantir la pertinence de cette application, nous devons vérifier les paramètres suivant (Pedersen et Tryfona, 2001) :



- L'additivité de la mesure (stock, flux ou valeur par unité),
- La structuration de la hiérarchie (stricte ou non stricte),
- Le type de fonction d'agrégation,
- La distributivité de la fonction d'agrégation,
- Les relations spatiales topologiques entre les membres spatiaux.

Au sein du SOLAP, nous pouvons appliquer différentes fonctions d'agrégation applicables dans OLAP traditionnel. En plus, concernant les données géométriques, des fonctions d'agrégation spatiale de type géométrique peuvent être appliquées telles que l'intersection, l'union, centroïde . . . etc.

2.7.4 Architecture d'un système décisionnel basée sur SOLAP

En plus des données sources, l'architecture d'un système décisionnel basée sur SOLAP peut être résumée en quatre couches logicielles : ETL spatial, ED spatiales, Serveur SOLAP et Client SOLAP, cette architecture est illustrée dans la Figure 2.11, (Bouilil, 2012).

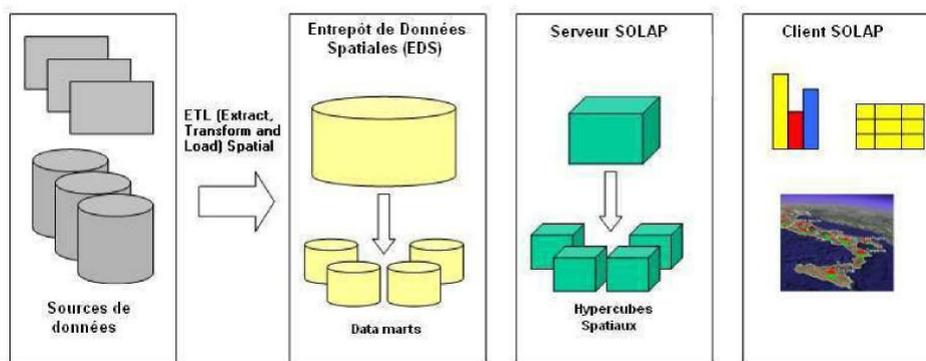


FIGURE 2.11 – Architecture d'un système décisionnel basé sur SOLAP.

a) Couche ETL spatial

L'intégration de données spatiales dans l'EDS nécessite des outils ETL (Extract-Transform-Load) spatiaux (par exemple *Spatial Data Integrator*¹⁰), qui permettent le traitement et la transformation des données spatiales pour résoudre des problèmes d'hétérogénéité de données obtenues de différents systèmes de référence spatiale à différentes échelles géographiques, etc. (Xi-Qian et al., 2004; Bédard et Han, 2009). Cette opération peut être effectuée par changement de projection cartographique, application du géocodage, . . . etc.

b) Couche EDS

Cette couche se base sur un modèle spatio-multidimensionnel. Elle englobe l'entrepôt de données spatiales et non spatiales, les métadonnées et éventuellement magasins de données. Elle permet d'historiser et de gérer de très gros volumes de données nécessaires à l'analyse SOLAP tels qu'Oracle Spatial et SQLServer.

10. <http://www.spatialdataintegrator.com> (dernière visite le 24/05/2018)

c) Couche Serveur SOLAP

Grace à cette couche nous pouvons calculer l'hypercube et implémenter les opérateurs de navigation pour permettre l'exploration spatio-multidimensionnelle dans le SOLAP.

d) Couche Client SOLAP

Cette couche permet de fournir aux utilisateurs les interfaces de navigation interactive facile et rapide au sein de l'hypercube de données spatiales et non spatiales.

Ces interfaces doivent fournir différents modes d'affichages interactifs (histogrammes, tables de pivot et cartes) d'une part ; et d'appliquer facilement les différents opérateurs de navigation d'une autre part. Cette couche permet l'exploitation du produit final de l'analyse spatio-multidimensionnelle. Voir Figure 2.12. (Abdallah bensalloua et al., 2014).

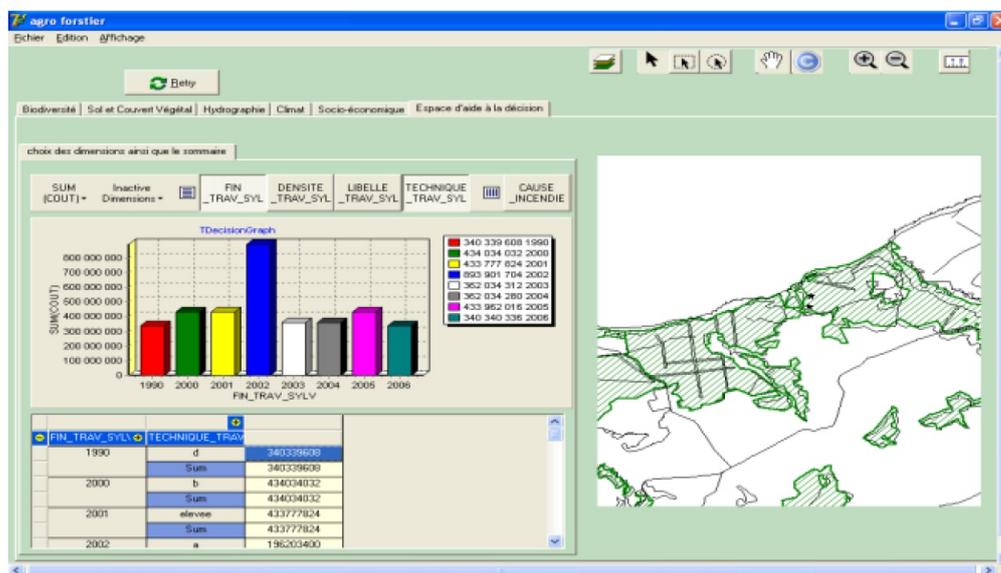


FIGURE 2.12 – Exemple de Client SOLAP.

2.7.5 Modes d'intégration SIG-OLAP

Dans un système SOLAP, nous cherchons à utiliser les techniques d'analyse multidimensionnelle des systèmes OLAP en intégration des fonctionnalités de représentation cartographique et l'analyse spatiale des Systèmes d'Information Géographique.

Il existe trois familles de solutions technologiques pour le développement et l'implantation d'une application SOLAP (Zaamoune, 2015) :

a) OLAP dominant

Ce type d'application est développé autour d'un serveur OLAP. Les fonctions OLAP sont dominantes alors que les fonctions SIG sont minimales. Ex. zoom, déplacement, sélection, gestion des couches actives. Parfois, forage spatial minimal.

Il a l'avantage de supporter l'exploration et la visualisation OLAP en utilisant toutes les capacités d'un serveur OLAP. Néanmoins, il fournit des fonctionnalités d'exploration et de visualisation cartographiques simples.

b) SIG dominant

Ce type d'application est développé autour d'un SIG et d'un SGBD. Il fournit toutes les fonctions de la cartographie et de l'analyse spatiale des SIG. Néanmoins il ne permet pas toutes les capacités d'un serveur OLAP. (Abdallah Bensalloua et al., 2012)

c) SOLAP intégré

Ce type d'application permet un haut niveau de fonctionnalités pour les vues et données spatiales et non-spatiales. Il intègre et synchronise toutes les fonctions OLAP et SIG. (Abdallah Bensalloua et al., 2014).

2.7.6 Modélisation spatio-multidimensionnelle : Etat de l'art

Dans la littérature, beaucoup de travaux ont entamé la problématique liée à la modélisation spatio-multidimensionnelle. En effet, ils visent définir de nouveaux concepts par l'extension des concepts classiques de la modélisation multidimensionnelle traditionnelle.

Ces travaux peuvent être classés en deux catégories notamment : i) celle basée sur des standards tels que Unified Modeling Language (UML) et Entité-Relation (ER) ; et ii) celle basée sur des formalismes ad hoc.

Pour plus de détails sur ces aspects nous pouvons nous référer à (Kimball et Ross, 2002 ; Malinowski et Zimányi, 2008 ; Salehi, 2009).

Modèles basés sur des standards

a) Modèles UML

Un profil UML pour la modélisation multidimensionnelle est présenté dans (Lujan-Mora et al., 2006). Ce profil est étendu par les auteurs dans (Glorio et Trujillo, 2008), en définissant de nouveaux stéréotypes notamment « *SpatialMeasure* » et « *SpatialLevel* » pour représenter les mesures spatiales et les niveaux d'agrégation spatiale consécutive.

Dans ce modèle, les hypercubes peuvent partager des dimensions, des hiérarchies et des niveaux d'agrégation. En effet, l'utilisation des paquetages UML permet l'organisation des éléments multidimensionnels selon des niveaux de détail, ce qui permet une lecture simple des modèles complexes.

Dans (Pinet et Schneider, 2010), les auteurs étendent le modèle UML présenté dans (Pinet et Schneider, 2009) par le concept de « *mesure spatiale* » qui signifie un attribut UML dont le type de données est géométrique ; et le concept de « *classe identifiée spatiale* » qui signifie une classe ayant un attribut géométrique. De plus les auteurs unifient les représentations des faits et des niveaux d'agrégation sous en « *classes identifiées spatiales* » ; et les représentations des relations d'agrégation et des relations (fait-dimension) en « *associations d'agrégation* ».

Dans le travail présenté dans (Pinet et al., 2010), les auteurs proposent un profil UML pour SOLAP, dans lequel ils représentent les hypercubes par des paquets UML, avec les niveaux d'agrégation représentés par classes reliés par des relations d'agrégation UML. Les mesures sont regroupées dans un stéréotypée « *Measure* » ; alors que les dimensions sont spécialisées en thématiques, spatiales et temporelles. Pour le meilleur de nos connaissances, Boulil (Boulil, 2012) a fait une transition remarquable dans le domaine de la modélisation spatio-multidimensionnelle ces dernières années. En effet, dans (Boulil et al., 2015), les auteurs proposent un nouveau profil UML pour modéliser des EDS et des agrégations spatiales complexes. Ils proposent aussi un outil pour la mise en œuvre automatique de cubes de données spatiales en utilisant ce profil. Cette solution permet de générer différentes représentations logiques du modèle de l'EDS (schéma en étoile et schéma en flocon de neige) et d'implémenter des indicateurs d'analyse SOLAP complexes en utilisant le langage MDX (MultiDimensional eXpression).

b) Modèles ER

Dans la littérature, nous trouvons peu de travaux qui ont entamé la modélisation spatio-multidimensionnelle de type (Entité-Relation). Nous pouvons citer à titre d'exemple le modèle présenté dans (Malinowski et Zimányi, 2008) qui étend le modèle multidimensionnel présenté (Malinowski et Zimányi, 2006) pour proposer un modèle spatio-multidimensionnel ER. Ce dernier fournit un support spatial pour les éléments multidimensionnels (mesures, faits, niveaux d'agrégation, hiérarchies et dimensions). En effet, la mesure spatiale est définie comme une mesure dont le type de données est géométrique.

Modèles ad hoc

La plupart des propositions de la modélisation spatio-multidimensionnelle traitent des problématiques particulières. Par exemple dans (Jensen et al., 2004) les auteurs traitent le problème de comptage en double des valeurs de mesure dans les hiérarchies de type non strictes. Il s'agit d'une extension du modèle multidimensionnel proposé dans (Pedersen et al., 2001). Dans (Bimonte et al., 2005), les auteurs traitent le problème de prise en compte de la dépendance entre les agrégations numérique et spatiale. Les auteurs dans (Damiani et Spaccapietra, 2006), proposent une solution pour représenter les mesures spatiales à différentes échelles géographiques.

Dans la plus part des modèles de type ad hoc, ne fournissent pas de notations graphiques ce qui rend leur compréhension et utilisation très difficile. (Boulil, 2012).

2.8 Limite des systèmes SOLAP

Comme tout type d'outils, les systèmes SOLAP présentent quelques limites. Ces limites sont des difficultés qui résident dans plusieurs niveaux. En effet, lors de l'exploitation des systèmes SOLAP, quelques problèmes liés à la consistance des résultats, sont posés. Il

s'agit de problèmes liés à la qualité des données entreposées, la qualité d'agrégation et à la qualité de l'exploration effectuée par les utilisateurs qui formulent des requêtes invalides.

Des études ont proposé de résoudre ces problèmes à un niveau bien avancé notamment au stade de modélisation.

Néanmoins, ces efforts, ont contribué à la complexification de la conception des entrepôts de données spatiales. En plus, l'application de ces techniques, ajoute beaucoup d'informations artificielles, qui rendent les résultats d'analyse loin de la réalité.

En outre, les données géographiques sont par nature volumineuses. Ici, face au caractère incrémental des entrepôts de données spatiales, se posent des problèmes d'optimisation de la gestion d'espace de stockage d'une part, et d'optimisation des techniques de recherche d'information au sein des EDS très volumineux d'une autre part.

Bien que l'intégration OLAP-SIG dans un système d'information décisionnel, a permis de résoudre beaucoup de problématiques liées à la gestion du territoire, avec un grand volume de données géo-référencés; cette solution est devenue incapable de fournir des mécanismes d'aide à la décision dans des situations complexes, où nous devons tenir en considération plusieurs exigences parfois conflictuelles. Cette situation se pose clairement dans la gestion des ressources naturelles et en particulier en agroforesterie. A cet effet, le recours aux techniques de l'analyse multicritères devient une nécessité.

2.9 Conclusion

Dans ce chapitre, nous avons abordé les problématiques d'aide à la décision spatiale. Pour ce fait, nous avons exploré les types de décision spatiale. Ensuite, le processus de prise de décision spatiale a été entamé. Les systèmes d'information décisionnels (SID) spatiaux (SDSS), sont utilisés pour fournir aux décideurs des solutions leur permettant la prise de décision qui porte sur le territoire en se basant sur des données géo-référencées. Les SID spatiaux, se basent donc, sur des mécanismes qui permettent l'analyse des données historisées non volatiles à plusieurs niveaux de granularité à travers l'EDS et l'OLAP et des techniques qui permettent la gestion, l'exploration et l'analyse des données à référence spatiale par les SIG.

A cet effet, nous avons présenté premièrement, les concepts de l'information géographique, les systèmes de coordonnées et les systèmes de projection. Puis nous avons abordé la présentation des systèmes d'information géographique. Le SIG, étant un système permettant de rassembler et organiser, de gérer, d'analyser et d'archiver et de présenter des données géographiques; il est considéré comme un outil d'aide à la décision pour le décideur dans différents domaines. Nous avons présenté par ailleurs, quelques travaux relatifs à l'utilisation des SIG dans le domaine de l'agroforesterie.

Ensuite, nous avons entamé la présentation des systèmes d'information décisionnels basés sur les entrepôts de données spatiales.

Dans ce contexte, un système SOLAP bénéficie à la fois des concepts de OLAP qui fournit des composants de modélisation et d'analyse multidimensionnelle d'une part; et des SIG qui permettent l'analyse spatiale et la présentation cartographique des données géo-localisées d'autre part. Un état de l'art dans le domaine de la modélisation spatio-

multidimensionnelle est ainsi présenté. À la fin, nous avons présenté quelques limites des systèmes SOLAP.

Dans le chapitre suivant, nous allons aborder les méthodologies d'analyse multicritères. Ces méthodologies qui offrent des mécanismes de résolution des problèmes avec plusieurs exigences, peuvent couvrir les limites des solutions offertes par les systèmes SIG et SOLAP.

Chapitre 3

Méthodologie Multicritères d'Aide à la Décision

Dans ce chapitre, nous allons décrire les principales notions relatives à l'aide multicritères à la décision. Les principes de l'analyse multicritères ainsi que les méthodes d'AMC par agrégation partielle, sont également, présentés. Par la suite, nous allons effectuer une synthèse concernant ces méthodes avant de présenter quelques travaux connexes.

Sommaire

3.1	Introduction	49
3.2	L'aide à la décision	49
3.2.1	Typologie des décisions	49
3.2.2	Acteurs de la décision	50
3.2.3	Systèmes Interactifs d'Aide à la Décision (SIAD)	50
3.3	Notions relatives à l'Analyse MultiCritères (AMC)	51
3.3.1	Définition de L'AMC	51
3.3.2	Le concept d'action	51
3.3.3	Critère et famille de critères	52
3.3.4	Matrice des performances	52
3.3.5	Les paramètres subjectifs	53
3.3.6	Relation de surclassement	53
3.3.7	Typologie des problèmes décisionnels	53
3.4	Agrégation multicritères	54
3.4.1	Définition de l'agrégation	54
3.4.2	Type d'agrégation	55
3.5	Méthodes d'AMC par agrégation partielle	55
3.5.1	La famille des méthodes ELECTRE	56
3.5.2	La famille de méthodes PROMETHEE	60
3.6	Synthèse des méthodes d'analyse multicritères	61
3.7	Travaux connexes	62
3.8	Conclusion	63

3.1 Introduction

Prendre une décision est une tâche avec une importance considérable de toute personne responsable de ce qui va résulter de cette décision. Pour remédier aux mauvaises suites, l'objectif consiste à améliorer le processus de prise de décision qui correspond à des interactions entre différents systèmes, acteurs, et dans certains cas plusieurs décideurs.

Le processus de décision peut atteindre un niveau de complexité très avancé en présence de certains paramètres complexes. Ceci est lié en particulier à des problèmes d'éclaircissement de la vision des acteurs et de quantification de leurs préférences. Le problème se multiplie en présence d'une problématique de décision sur le territoire urbain ou naturel.

Dans le processus d'aide à la décision, nous cherchons donc à modéliser une réalité présente qui est souvent difficilement mesurable, et une réalité future pour prédire les conséquences du futur choix (Roussat, 2007).

Dans le cadre de cette thèse, la décision concerne des espaces agroforestiers d'où l'intervention de différents acteurs qui affectent le processus d'aide à la décision. Ce dernier cherche à leur fournir l'analyse rigoureuse des différents paramètres du problème à résoudre.

3.2 L'aide à la décision

Bernard Roy (Roy, 1985) définit l'aide à la décision comme étant «*l'activité de celui qui, en prenant appui sur des modèles, aide à obtenir des éléments de réponse aux questions que se pose un intervenant dans un processus de décision, éléments concourant à éclairer la décision et à recommander un comportement de nature à accroître la cohérence entre l'évolution du processus et les objectifs de cet intervenant*».

A ce titre, une décision représente une action pouvant être prise pour modifier un état quelconque en vue d'atteindre un certain objectif ou de résoudre un problème (Trahand, 1999). Prendre une décision correspond donc, à choisir une action entre plusieurs alternatives possibles.

3.2.1 Typologie des décisions

Igor Ansoff (Ansoff H.I, 1989) propose une classification des décisions pouvant être prises dans une entreprise. A cet effet, une décision peut être :

- a) **Stratégique** : portant sur les affaires extérieures de l'entreprise. Elle présente le positionnement de l'entreprise et cherche à répondre aux questions de type : quels sont les objectifs de l'entreprise ? Comment doit-elle exploiter sa position actuelle sur le marché ? Comment diversifier ses activités ? ...
- b) **Administrative** : portant sur la gestion des ressources. Elle cherche à organiser les structures de l'entreprise (responsabilités, flux du travail et de l'information), et le développement des ressources.
- c) **Opérationnelle** : portant sur l'exploitation efficace des ressources. Elles concernent principalement la répartition budgétaire, la planification des opérations, la direction

des activités, etc.

Dans le même cadre, et selon le processus à appliquer, Simon (Simon, 1983) distingue deux types de décisions :

- a) *les décisions programmées*, appliquant un processus prédéfini.
- b) *les décisions non programmées*, appliquant un des processus de décision partiellement maîtrisés ou non.

Selon les auteurs dans (Bressy et Konkuyt, 2004), et sur la base du degré d'incertitude, une décision peut être d'un avenir :

- a) *certain* : il s'agit de décision de gestion courante caractérisée par un risque presque nul dans la mesure où l'on connaît le résultat de la prise de décision dans la majorité des cas. Toutefois un risque totalement nul n'existe pas. Les décisions certaines sont celles qui portent le moins de conséquences (décision de gestion courante).
- b) *incertain* : caractérisée par la complexité de l'environnement, où, l'entreprise ne maîtrise pas toutes les variables.
- b) *aléatoire* : dans ce cas, nous ne pouvons pas connaître le résultat à l'avance. Des techniques telles que l'arbre de décision peuvent être utilisées pour rationaliser la décision.

3.2.2 Acteurs de la décision

Les acteurs de la décision peuvent être multiples. Nous considérons la définition présentée par Bernard Roy dans (Roy, 1985) : *"Un individu ou un groupe d'individus est acteur d'un processus de décision si, par son système de valeur, que ce soit au premier degré du fait des intentions de cet individu ou groupe d'individus, ou au second degré par la manière dont il fait intervenir ceux d'autres individus, il influence directement ou indirectement la décision. De plus, pour qu'un groupe d'individus (corps constitué ou collectivité) soit identifié comme un seul et même acteur, il faut que, relativement au processus, les systèmes de valeurs, systèmes informationnels et réseaux relationnels des divers membres du groupe n'aient pas à être différenciés."*

Les principaux acteurs d'un processus décisionnel peuvent être :

- **Décideur (groupe de décideurs)** : fixe l'objectif à atteindre, exprime les préférences.
- **L'homme d'étude** : prend en charge le processus d'aide à la décision
- **Intervenants** : ils peuvent être des experts qui influencent la décision en essayant d'imposer leurs propres points de vue.

Dans certaines situations, d'autres acteurs peuvent se présenter tels que le négociateur ou le médiateur, etc. (Roussat, 2007).

3.2.3 Systèmes Interactifs d'Aide à la Décision (SIAD)

Dans le but de faciliter le processus d'aide à la décision, les entreprises (ou différentes organisations) utilisent des systèmes interactifs d'aide à la décision (SIAD). Il s'agit de

systèmes d'information ayant la capacité d'accéder aux différentes bases de données. Ils développent des modèles pour représenter et évaluer différentes situations afin de générer des scénarios de solution. Tel qu'il est noté au chapitre précédent, et selon Simon (1960), le processus décisionnel passe par les trois phases relatives à l'identification du problème, l'élaboration de scénarios et le choix de solution. Par conséquent, la mission fondamentale d'un SIAD est de fournir aux décideurs, l'information permettant l'identification d'une situation où une décision doit être prise (Rivard et Talbot, 2001).

Dans la section suivante, nous allons présenter l'analyse multicritères qui englobe des techniques permettant à aider les décideurs dans la prise décision.

3.3 Notions relatives à l'Analyse MultiCritères (AMC)

L'analyse Multicritères (AMC) désigne un ensemble d'outils d'aide à la décision développés dans le but de résoudre des problèmes en présence de plusieurs critères souvent conflictuels. La solution peut se présenter en plusieurs alternatives.

3.3.1 Définition de L'AMC

Dans la littérature, différentes définitions pour l'AMC sont présentées. En effet, dans (Vincke, 1989), l'analyse multicritères est définie comme étant « *une approche constructiviste visant à fournir des outils permettant de progresser dans la résolution d'un problème où plusieurs points de vue, souvent contradictoires, doivent être pris en compte* ».

L'analyse multicritères vise à orienter un choix sur la base de plusieurs critères pour porter un jugement comparatif entre des alternatives différentes. Ceci en impliquant tous les acteurs et en simplifiant le problème initial pour aboutir à des recommandations pour la solution finale. Ici, l'homme d'étude prend en charge le processus d'aide à la décision en utilisant les méthodes d'AMC.

3.3.2 Le concept d'action

Bernard Roy (Roy, 1985) définit une action comme : *La représentation d'une éventuelle contribution à la décision globale susceptible, en égard à l'état d'avancement du processus de décision, d'être envisagée de façon autonome et de servir de point d'application à l'aide à la décision*". Elle désigne une des solutions proposée pour le problème à résoudre. Selon les auteurs dans (Roy et al., 1993), une action peut être :

- *Réelle* : qui peut être exécutée.
- *Fictive* : correspond à un objet de référence dans les problèmes de classification multicritères.
- *Réaliste* : dont l'exécution peut être envisagée.
- *Irréaliste* : correspond à la satisfaction d'objectifs incompatibles.

Une action est dite potentielle si elle est jugée réaliste par au moins un acteur. Il faut souligner par ailleurs, que la solution finale d'un processus décisionnel n'est pas forcément une action unique, mais elle peut être une combinaison de plusieurs actions.

3.3.3 Critère et famille de critères

Un critère est considéré comme une fonction 'g' à valeurs définie sur l'ensemble des actions potentielles de telle sorte qu'il soit possible de comparer deux actions 'a' et 'b' à partir des deux nombres 'g(a)' et 'g(b)' (Roy et al., 1993). Il est défini aussi dans (Roy, 1985) comme "Tout effet ou attribut de l'action susceptible d'interférer avec les objectifs ou avec le système de valeurs d'un acteur du processus de décision, en tant qu'élément primaire à partir duquel il élabore, justifie ou transforme ses préférences" .

Une échelle de préférence est associée à chaque critère pour permettre le classement des actions selon leur performance par rapport à ce critère. (Roussat, 2007).

3.3.4 Matrice des performances

La matrice de performance est un tableau dans lequel chaque ligne représente une action et chaque colonne un critère. Les cellules de ce tableau représentent le jugement de l'action 'i' par rapport au critère 'j'. La Table 3.1 illustre un exemple d'une matrice de performances. (Hamdadou et al., 2016).

TABLE 3.1 – Table de performances (Actions-Critères)

	Surface	Distance élect.	Bruit	Impacts	Equipement	Climat	Nuisance	Distance école
ZONE 1	1230	123	12	11	14	0.3	1	152
ZONE 2	231	1453	15	15	11	0.6	0.6	52
ZONE 3	458	452	0.6	0	150	0.8	8.5	360
ZONE 4	456.36	78	0.8	13	135	1	2.3	450
ZONE 5	478.2	325.5	1.2	12	145	1.2	12.3	230
ZONE 6	1456.36	74	13	14	600	0.8	8	25
ZONE 7	745	769	10.5	2	320	1.3	0	36
ZONE 8	789.78	15	11	6	300	0.5	1.6	453
ZONE 9	478	2	0.9	5	258	0.5	0	123
ZONE 10	1023	56.79	14	9	2	0.5	0	14

3.3.5 Les paramètres subjectifs

Les paramètres subjectifs sont des valeurs fixées par le décideur, et doivent être respectées durant le processus décisionnel. Ils peuvent être inter-critères ou intra-critères (Hamdadou, 2008).

a) Paramètres inter-critères

Dans cette catégorie, nous parlons souvent de poids d'un critère, et son importance vis à vis des autres critères. Il existe différentes méthodes qui aident à définir le poids des critères notamment : « la méthode de cartes de SIMOS » (Simos, 1990) et « la méthode de l'échelle de Saaty » (Saaty, 1984).

b) Paramètres intra-critères

Ils représentent des seuils qui expriment l'appréciation subjective. Ils peuvent être : (Hamdadou et al., 2016)

- **Seuil d'indifférence** : Représente l'écart dans lequel aucune préférence ne peut être établie sur un critère.
- **Seuil de préférence** : Représente l'écart à partir duquel une préférence nette peut être établie entre deux évaluations.
- **Seuil de veto** : Représente une valeur si elle dépassée sur un critère, alors l'action ne peut être prise en considération.

3.3.6 Relation de surclassement

La Relation de surclassement entre deux actions a et b pour dire que « a surclasse b » notée aSb est défini si « a est au moins aussi bonne que b relativement à une majorité de critères, sans être trop nettement plus mauvaise relativement aux autres critères. » (Hamdadou et al., 2016). Ceci correspond à l'existence de raisons claires qui justifient soit une préférence stricte, soit une préférence faible en faveur de l'une des deux actions ou bien une indifférence entre les deux.

3.3.7 Typologie des problèmes décisionnels

Dans la littérature, sont identifiées différents types de décisions de différentes portées et qui ne sont pas traitées de la même façon.

Selon Roy (Roy, 1985), les problèmes décisionnels peuvent être classés en quatre catégories principales notamment : problématiques de choix, de tri, de rangement et de description.

- a) **Problématiques de choix ($P.\alpha$)** : dans ce type, nous cherchons à sélectionner avec une précision maximale, un sous ensemble A' qui contient les meilleures actions à partir de l'ensemble initial des actions A . vu que les critères sont souvent conflictuels, il est nécessaire de fournir plusieurs actions représentant des variantes de l'action idéale.

Exemple : pour le choix d'un site d'une forêt récréative, plusieurs sites sont fournis et qui représentent des actions choisies.

b) **Problématiques de tri ($P.\beta$)** : dans ce type nous cherchons à affecter plusieurs actions à un ensemble de catégories qui sont prédéfinies. Ici, nous examinons chaque action individuellement en faisant trier les actions potentielles à un ensemble de catégories représentant des normes prédéfinies.

Exemple : affecter différentes parcelles forestières à une pratique silvicole particulière.

c) **Problématique de rangement ($P.\gamma$)** : dans ce type nous cherchons à ranger l'ensemble initial des actions en allant de la meilleure action à la moins bonne avec éventuellement des cas ex-aequo.

Ceci est intéressant si nous avons besoin de différencier les actions selon leur importance. En pratique, le rangement peut être nécessaire pour un sous ensemble d'actions jugées les plus intéressantes.

Exemple : ranger les différentes actions d'intervention sur un ensemble de parcelles en présence de dégradation écologique, en commençant par la parcelle la plus sensible du risque.

d) **Problématique de description ($P.\delta$)** : dans ce type nous cherchons à décrire un ensemble d'actions avec leurs conséquences selon un procédé d'investigation cognitive.

Exemple : identification et description des différentes actions pour lutter contre les feux de forêts.

Cette catégorisation est valable pour les problématiques décisionnelles sur le territoire. Néanmoins, le dernier type relatif aux 'problématiques de description' n'est pas retenu par certains auteurs vu qu'elles ne disposent pas de méthodes particulières pour les résoudre. En revanche, les problèmes spatiaux peuvent se servir de cette catégorie pour évaluer certaines situation et pas forcément pour prendre des décisions (Chakhar, 2006).

3.4 Agrégation multicritères

L'agrégation multicritères correspond à l'établissement d'un modèle des préférences globales (Hamdadou, 2008).

3.4.1 Définition de l'agrégation

Il s'agit de trouver une manière pour comparer les actions potentielles en tenant compte de tous les critères. Par exemple si nous avons un ensemble d'actions $\{a, b, c, \dots\}$ avec leurs performance par rapport à un ensemble de critères. Nous allons donc voir que l'action 'a' soit meilleure que les autres actions pour certains critères, et pas pour les autres ; et de même pour 'b', 'c', ... etc. Le problème de l'agrégation correspond donc à prendre en compte, d'une manière compréhensive, les performances de 'a', les performances de 'b', etc.

3.4.2 Type d'agrégation

Dans le contexte des méthodes d'aide à la décision multicritères, nous pouvons distinguer trois types d'agrégation, notamment : Agrégation Complète, Agrégation Partielle et Agrégation Locale Itérative. (Hamdadou et al., 2016)

Agrégation Complète (critère unique)

Dans ce type d'agrégation, on procède à introduire toutes les performances dans une seule fonction d'agrégation avec d'éventuels poids. Ceci implique une compensation totale entre les critères. L'inconvénient majeur de cette approche correspond aux difficultés de choisir la fonction d'agrégation avec l'intégration de toutes les préférences de(s) décideur(s). Dans ce cas nous pouvons citer à titre d'exemple : MAUT (Multiple Attribute Utility Theory), UTA (Utilité Additives), AHP (Analytic Hierarchy Process), WPM (Weight Product Method), WSM (Weight Sum Method).

Agrégation partielle (surclassement de synthèse)

Dans cette approche, une comparaison des actions deux à deux est établie. Puis on procède à la synthèse des résultats de ces comparaisons.

Ce type d'agrégation est utilisé s'il existe au moins un critère quantitatif; et pour les autres les unités sont très hétérogènes ne permettant pas un codage en une échelle commune. Dans ce cas des seuils de préférences ou de veto doivent être pris en compte.

Cette approche présente en particulier l'inconvénient de de la complexité et l'imprécision des résultats du au nombre important de comparaisons. Néanmoins, Elle prend la forme d'un ensemble de conditions permettant d'accepter ou de rejeter un surclassement au niveau global. Ce qui permet une réponse synthétique et définitive au problème de l'agrégation des performances (Hamdadou et al., 2016).

Agrégation locale (Approche du jugement local interactif)

Dans cette approche une solution de départ est choisie. Ensuite, un groupe de variantes relativement proches à celle-ci est sélectionné. Par la suite, on cherche une meilleure variante par rapport à celle sélectionnée. Ce choix est considéré comme une solution initiale pour une nouvelle itération. Cette méthode est applicable lorsqu'il y a un nombre important de variantes

Parmi les méthodes s'appuyant sur l'approche d'agrégation locale nous pouvons citer PLM (Programmation Linéaire Multicritères), UTA interactive, Prefcalc (Hamdadou, 2008).

3.5 Méthodes d'AMC par agrégation partielle

Les méthodes dites par agrégation partielle, ont l'avantage majeurs, d'accepter les situations d'incomparabilité. Elles suivent un procédé en deux étapes : (Roussat, 2007)

- *Etape 1* : Effectuer la comparaison par paire des actions,

- *Etape 2* : Chercher à synthétiser les relations de surclassement, d'incomparabilité ou d'indifférence pour obtenir un classement final.

Dans ce qui suit, Nous allons présenter, les principales méthodes de d'AMC d'agrégation partielle notamment ELECTRE (I, II, III,... etc) d'une part ; et PROMETHEE d'une autre part.

3.5.1 La famille des méthodes ELECTRE

La Méthode ELECTRE I

a) *Principe de la méthode*

La méthode ELECTRE I traite les problématiques d'aide à la décision de type α (Choix). Elle consiste à chercher un sous ensemble contenant les "meilleures" actions parmi l'ensemble des actions envisagées. Cette méthode utilise des critères dits francs. Donc les relations entre deux actions a_i et a_k pour un critère c_j sont de trois types : *aiPak ou aiIak ou akPai* (Roussat, 2007).

b) *Démarche d'utilisation*

La méthode ELECTRE I est basée sur les étapes suivantes

- Réalisation du tableau des performances
- Calcul de la matrice de concordance
- Calcul de la matrice de discordance
- Test de concordance
- Test de discordance
- Relation de surclassement S

La Méthode ELECTRE II

Cette méthode relève de la problématique gamma γ (le classement), en utilisant les relations d'ordre sur chacun des critères.

a) *Principe de la méthode*

Cette méthode vise à munir l'ensemble des actions potentielles d'une structure de pré ordre total pour rendre facile le choix. Elle consiste donc à assurer le classement des actions potentielles, depuis la "meilleure" jusqu'à la "moins bonne". Elle utilise aussi la relation de surclassement S, mais les tests de concordance et de non discordance sont imbriqués, à la différence d'ELECTRE I.

b) *Démarche d'utilisation*

Pour appliquer la méthode ELECTRE II, nous devons suivre les étapes suivantes : (Hamdadou et al., 2016) :

- Réaliser le tableau des performances
- Vérifier la concordance
- Vérifier la discordance
- Établir des relations de surclassement
- Exploiter les relations de surclassement

La nouveauté de cette méthode est l'introduction de deux niveaux des surclassements notamment fort et faible. Mais malgré les changements qu'elle apporte, elle reste équivalente à la précédente. (Hamdadou et al., 2016)

La Méthode ELECTRE III

a) Principe de la méthode

Comme la précédente, cette méthode de surclassement, cherche à résoudre les problématiques de type gamma γ (le classement). Ceci en traitant une matrice composée des actions et des pseudo-critères. Un préordre final partiel est établi suite au traitement de surclassement muni sur cette matrice.

ELECTRE III apporte des nouveautés par rapport aux autres méthodes de surclassement. Ceci est traduit par l'introduction du flou dans la relation de surclassement, et de la notion de préférence faible à travers l'utilisation des seuils d'indifférence et de préférence stricte.

b) Démarche d'utilisation

La méthode ELECTRE III opère en deux étapes : (Hamdadou et al., 2016)

1. **Agrégation** : relative à l'introduction des paramètres subjectifs à savoir : Poids des critères, Seuil de préférence, Seuil d'indifférence et Seuil de veto.
2. **Exploitation** : relative à l'établissement de deux distillations (descendante et ascendante). Ensuite, le graphe de surclassement illustrant le préordre final partiel est établi à travers l'application d'un algorithme comparant les deux préordres.

Les atouts de cette méthode peuvent se manifester du fait qu'elle considère les critères qualitatifs et quantitatifs. En plus, les notions de seuils d'indifférence et de préférence permettent de considérer l'imprécision dans l'évaluation des critères. La notion du seuil de veto permet aussi de mettre en évidence des différences inacceptables entre deux actions sur un critère.

En revanche, les paramètres techniques devant être fixés rendent cette méthode difficile à être appliquée par l'homme d'étude et ses conséquences difficile à comprendre.

La méthode ELECTRE IV

La méthode ELECTRE IV relève de la problématique de classement γ .

a) Principe de la méthode

La méthode ELECTRE IV simplifie l'utilisation de ELECTRE III en éliminant les poids des critères (Tille, 2000). Elle est généralement applicable lorsqu'il devient extrêmement difficile d'élaborer une pondération.

b) Démarche d'utilisation

Cette méthode se base sur les étapes suivantes : (Hamdadou et al., 2016)

- Réaliser le tableau des performances
- Comparer les actions deux à deux

- Construire la relation de surclassement
- Exploiter la relation de surclassement

La méthode ELECTRE IS

La méthode ELECTRE IS traite la problématique de choix α .

- a) **Principe de la méthode** La méthode ELECTRE IS adapte la méthode ELECTRE I pour l'utilisation des critères flous. Elle utilise aussi les seuils d'indifférence, de préférence et de veto. Elle divise les actions potentielles en deux groupes : des actions non surclassées constituant le noyau et des actions surclassées. La meilleure action se trouve dans la premier groupe.
- b) **Démarche d'utilisation**
- La méthode ELECTRE IS comporte les étapes suivantes : (Hamdadou et al., 2016)
- Calcul des indices de concordance
 - Calcul des indices de discordance
 - Etablissement de la relation de surclassement
 - Recherche du noyau

La méthode ELECTRE TRI

La méthode ELECTRE TRI traite la problématique de tri ou d'affectation β .

- a) **Principe de la méthode**
- Après l'établissement des catégories indépendamment de l'ensemble des actions, cette méthode affecte chaque action potentielle à l'une de ces catégories.
- b) **Démarche d'utilisation**
- Pour résoudre les problèmes de tri nous devons passer par les deux phases suivantes : (Hamdadou et al., 2016)
- Phase I : Modélisation des catégories : dans cette phase, deux étapes sont suivies :**
1. **Structuration** : comporte la définition des actions de référence ainsi que leurs paramètres
 2. **Validation** : consiste à valider les paramètres précédents en appliquant une technique directe avec l'intervention du décideur ou indirecte sans intervention du décideur

Phase II : Affectation

Après avoir déterminé les normes d'affectation, l'affectation des nouvelles actions aux différentes catégories est entamée. Dans le cas de tri ordinal, l'affectation d'une action à une catégorie lorsqu'elle est entre les frontières de cette catégorie. Alors que dans le cas d'un tri nominal, l'affectation d'une action se fait lorsqu'elle est similaire à au moins une action de référence de cette catégorie.

L'application de la méthode ELECTRE TRI passe par les étapes suivantes :

- Réaliser le tableau des performances
- Calculer les indices de concordance par critère
- Calculer les indices de concordance globale
- Calculer les indices de discordance par critère
- Calculer le degré de crédibilité
- Etablir la relation de surclassement
- Affecter les actions à des catégories selon une procédure d'affectation

La méthode ELECTRE TRI a l'avantage de permettre juger chaque variante individuellement sur sa valeur relative vis-à-vis des variantes de références. Ainsi, le temps de calcul est inférieur par rapport aux autres méthodes quelque soit le nombre de variantes.

Néanmoins, la méthode ELECTRE TRI ne peut pas détecter la présence de «clones» qui sont des variantes proches l'une de l'autre et qui donnent des résultats faussés.

La Table 3.2 représente une récapitulation des caractéristiques des différentes méthodes ELECTRE, (Hamdadou et al., 2016).

TABLE 3.2 – Les principales caractéristiques des méthodes ELECTRE

Méthode ÉLECTRE	I	II	III	IV	IS	Tri
problématique	α	γ	γ	Γ	α	β
Critères Francs	x	x				
Critères Flous			x	x	x	x
Seuils Concordance S_c	X	X				
Seuils Discordance S_d	X	X				
Seuils Indifférence S_i			X	X	X	X
Seuils Préférence S_p			X	X	X	X
Seuils Veto S_v			X	X	X	X
Indice de concordance par critère $c_j (a_i, a_k)$			X		X	X
Indice de concordance Globale C_{ik}	X	X	X		X	X
Indice de discordance par critère $d_j (a_i, a_k)$			X		X	X
Indice de discordance Globale D_{ik}	X	X				
ponds	X	X	X		X	X
Degrés de crédibilité δ_{ik}			X	X		X
Exploitation du surclassement	X				X	
Recherche du noyau		X				
Classement direct et inverse			X			
Distillation				X		
ascendante et descendante						

La Table 3.3 présente quelques avantages et inconvénients des méthodes ELECTRE adapté de l'étude présentée par (Hamdadou et al., 2016).

TABLE 3.3 – Avantages et inconvénients des méthodes ELECTRE

Méthode ÉLECTRE	Avantages	Inconvénients
I	<ul style="list-style-type: none"> - Simplicité d'utilisation - Compréhension aisée du processus 	<ul style="list-style-type: none"> - Rigidité des jugements - Résultats difficiles à interpréter - Echelle des notes ~ poids - Noyau : «le brillant second » est hors du noyau et circuits possibles
II	<ul style="list-style-type: none"> - Apparition de la nuance par surclassement - floue ou nette - Résultat plus tranché 	<ul style="list-style-type: none"> - Solution intermédiaire entre ELECTRE I et ELECTRE II - Critère francs - Entre les deux préordres totaux, risque d'action baladeuse
III	<ul style="list-style-type: none"> - Logique floue décrite par le degré de crédibilité du surclassement - seuils sont utilisés - Information totalement nuancée 	<ul style="list-style-type: none"> - Complexité du processus - Algorithme complexe - Nécessite un logiciel spécifique - Compréhension plus difficile pour le décideur (boite noire)
IV	<ul style="list-style-type: none"> - Absence de poids 	<ul style="list-style-type: none"> - Complexité du processus

3.5.2 La famille de méthodes PROMETHEE

Les méthodes PROMETHEE sont aussi des méthodes basées sur le principe de surclassement. Elles sont utilisées pour résoudre les problèmes de rangement. L'objectif de ces méthodes est de construire un classement des alternatives de la meilleure à la moins bonne, à travers un système de préférences floues. (Ayadi, 2010)

Elles sont utilisées récemment dans de nombreux cas de décision sur l'environnement. (Simon et al., 2004).

Principe des méthodes PROMETHEE

Ces méthodes se basent sur l'établissement d'un processus de comparaison de chaque action par rapport aux autres. Ceci permet le classement ordonné des actions. Le procédé peut se résumer comme suit : (Hamdadou et al., 2016)

- Attribution d'un poids et d'un type à chaque critère. Le type est une fonction de préférence qui a pour but d'avoir une différence de performance normée pour une différence de performance brute entre deux actions.
- Attribution d'un indice de préférence global $\Pi(A,B)$ entre deux actions A et B. cet indice est donné par :

$$\Pi(A, B) = \frac{\sum_{j \in k} w_j^* P_j(A, B)}{\sum_{j \in k} w_j} \quad (3.1)$$

Avec :

- w_j : poids attribué au critère j

- k : l'ensemble des critères, et
 - $P_j(A, B)$: la valeur de retour de la différence de notation entre A et B pour le critère j .
- pour chacune des actions, calcule de la moyenne des intensités des préférences sur toutes les alternatives, ce nombre est appelé flux de sortie de A et noté :

$$\Phi_{A^+} = \frac{\sum_{i \in I} \Pi(A, B_i)}{\text{Card}(I)} \quad (3.2)$$

Avec :

I : l'ensemble des actions pour lesquelles A surclasse B .

Le **flux entrant** de A est aussi calculé défini par :

$$\Phi_{A^-} = \frac{\sum_{i \in I} \Pi(B_i, A)}{\text{Card}(I)} \quad (3.3)$$

Avec :

I : l'ensemble des actions qui surclasse A .

Démarche d'utilisation

L'utilisation d'une méthode PROMETHEE entraîne l'exécution des étapes suivantes :

- Choix des critères généralisés : en utilisant une fonction de préférence, nous associons chaque critère, un critère généralisé ainsi, les effets d'échelle sont éliminés.
- Détermination d'une relation de surclassement : pour le but de quantifier les préférences du décideur, nous déterminons une relation de surclassement par le biais d'un indice de préférence.
- Evaluation des préférences : la prise en compte des flux entrant et sortant permet d'évaluer la préférence du décideur.

Pour plus de détail concernant la notion de critère généralisé et le choix des fonctions de préférence, le lecteur peut se référer (Hamdadou et al., 2016)

3.6 Synthèse des méthodes d'analyse multicritères

L'existence de plusieurs critères devant être pris en considération lors de différents processus décisionnels, peut garantir la fiabilité de la décision finale. Néanmoins, certains parmi eux, peuvent conduire à des situations conflictuelles. Pour cela, il est très important de trouver la meilleure méthode pour argumenter les solutions finales. La prise de décision peut être considérée donc comme une tâche délicate du fait qu'il n'existe pas réellement une manière pour répondre à la totalité des exigences.

A travers l'analyse multicritères, nous pouvons assurer la définition des objectifs, la définition de tous les critères de décision à prendre en compte ainsi que leurs importances.



En plus, grâce à l'analyse multicritères, nous concevons les alternatives du problème à résoudre.

Le choix de la méthode à appliquer est une tâche délicate. En effet, plusieurs méthodes existent pour résoudre, dans certains cas, le même type de problème. Néanmoins, les modes de fonctionnement diffèrent, et donc probablement, les résultats finaux sont différents.

Nous distinguons deux grandes écoles qui se différencient dans la manière de réflexion et qui sont présentées dans ce chapitre. La première regroupe les méthodes d'agrégation complète. Dans ce cas, nous cherchons à optimiser une fonction représentative du problème à résoudre. La deuxième catégorie regroupe les méthodes d'agrégation partielle. Ici, on se base sur la relation de surclassement en comparant les alternatives par paires.

Il est à noter donc, que chaque méthode d'analyse multicritères a ses propriétés, sa démarche d'utilisation et l'objectif à atteindre.

En appliquant une méthode donnée, la robustesse des résultats obtenus peut justifier l'utilité de la méthode pour le problème en cours et pour les différents paramètres utilisés.

L'application de plusieurs méthodes pour le même problème peut ne pas aboutir à des résultats satisfaisants. Néanmoins, ça peut conduire à mieux appréhender le problème et les résultats seront au moins bien argumentés.

3.7 Travaux connexes

Dans la littérature, il existe plusieurs exemples d'application de méthodologies multicritères, y compris la recherche agroforestière. Nous pouvons citer l'étude effectuée par les auteurs dans (Fontana et al., 2013) pour comparer les alternatives d'utilisation des terres. Dans cette étude, les auteurs ont obtenu une classification des alternatives par utilisation de la méthode PROMETHEE II.

Un autre exemple est celui présenté dans (Nordström et al., 2010). Dans cette étude, le processus de hiérarchie analytique (AHP) a été utilisé pour planifier la forêt urbaine en Suède, en déterminant la force de préférence pour un critère ou une alternative plutôt qu'un autre. Une étude critique des méthodes décisionnelles multicritères avec une référence particulière à la gestion et à la planification forestières est présentée dans (Ananda et Herath, 2009). Les auteurs dans (De Steiguer et al., 2003) ont annoté la bibliographie des modèles de décision multicritères pour la gestion de la foresterie et des ressources naturelles. L'évaluation participative multicritères des politiques de planification forestière dans des situations conflictuelles est présentée dans (Acosta et Corral, 2015).

Les auteurs dans (Fontana et al., 2013) ont utilisé le concept de services des écosystèmes pour définir une analyse de décision multicritères.

Dans (Greene et al., 2010), les auteurs développent une approche intégrée SIG-AMC qui intègre les phases d'exploration et d'évaluation dans l'étude d'un paysage dominé par la forêt. L'intégration d'un processus participatif avec le SIG et l'AMC pour la délimitation des aires protégées en Chine, est présentée par les auteurs dans (Zhang et al., 2013).

Dans (Montserrat et Serafín, 2017), les auteurs proposent une approche d'analyse décisionnelle multicritère et un système d'aide à la décision participative en gestion forestière.

Au sein du Laboratoire d'Informatique d'Oran (LIO), les travaux sur les systèmes

d'aide à la décision spatiale ont été largement abordés ; en particulier ceux qui sont basés sur les méthodes d'analyse multicritères.

En effet, dans (Hamdadou et al., 2013), les auteurs proposent un système d'aide à la décision spatiale de groupe en intégrant l'analyse multicritères et la théorie des jeux satisfaisants. L'intégration des méthodes d'analyse multicritères dans un système d'information décisionnel spatio-temporel a été, également, abordée dans (Zemri et Hamdadou, 2014). Une optimisation de la méthode multicritères Electre III a été proposée par les auteurs dans (Alnafie et al., 2013) où, la technique d'optimisation par la méthode des Colonies de Fourmis a été utilisée.

Dans (Younsi et al., 2009), les auteurs proposent un système interactif d'aide à la décision spatiale par l'intégration de l'analyse multicritères avec la télédétection et les SIG.

3.8 Conclusion

Au terme de ce chapitre, nous pouvons conclure que les deux manières de réflexion relatives aux méthodes d'agrégation multicritères, suivent la même logique issue de l'incertitude de la solution finale d'un problème donné. Ceci peut être justifié par l'ambiguïté dans l'expression des exigences et parfois par l'imprécision ou manque d'informations.

Du fait qu'il existe une différence remarquable entre le fonctionnement des deux catégories, il est donc très naturel de s'interroger sur la meilleure méthode qui sera plus adaptée pour les problématiques d'aide à la décision sur les espaces naturels. D'une manière générale, nous constatons que les méthodes d'agrégation partielle sont plus convoitées aux problématiques sociales et celles liées à la localisation spatiale des objets et des phénomènes. En revanche, pour les problématiques d'optimisation purement techniques, les méthodes d'agrégation complète sont généralement plus efficaces.

L'analyse multicritères est donc, très utile pour résoudre les problématiques avec des exigences conflictuelles dans les systèmes d'aide à la décision. Néanmoins, il y a d'autres problèmes liés à la consistance des indicateurs décisionnels, relatifs en particulier à la consistance de l'analyse.

Dans le chapitre suivant, nous allons présenter les concepts de vérification de la qualité des données utilisées, la qualité d'agrégation et la qualité de l'exploitation des outils SOLAP.

Chapitre 4

Qualité de l'analyse SOLAP

Dans ce chapitre, nous allons discuter la qualité des données et les contraintes d'intégrité (CI) dans les bases de données spatio-temporelles avant de présenter la qualité d'analyse spatio-multidimensionnelle. Les contraintes d'intégrité dans les cubes de données spatiales sont ensuite présentées avant de décrire la classification des langages de spécification des CI. Nous avons aussi mis l'accent sur le langage OCL et Spatial OCL avant de discuter quelques travaux effectués sur la qualité d'analyse dans les structures multidimensionnelles.

Sommaire

4.1	Introduction	65
4.2	Qualité des données géographiques	65
4.2.1	Critères quantitatifs	66
4.2.2	Critères qualitatifs	67
4.3	Contraintes d'intégrité dans les bases de données spatio-temporelles	67
4.4	Qualité d'analyse spatio-multidimensionnelle (SOLAP)	68
4.4.1	Selon la qualité de données	68
4.4.2	Selon la façon d'agrégation des mesures	68
4.4.3	Selon l'exploration des données	69
4.5	Contraintes d'intégrité dans les cubes de données spatiales	69
4.6	Classification des langages de spécification des CI	71
4.6.1	Langages naturels	71
4.6.2	Langages visuels	71
4.6.3	Langages logiques	72
4.6.4	Langages hybrides	72
4.7	OCL et Spatial OCL	72
4.8	Travaux effectués sur la qualité d'analyse dans les structures MD	74
4.8.1	Travaux sur la qualité d'agrégation	75
4.8.2	Travaux sur la qualité d'exploration	78
4.9	Conclusion	79

4.1 Introduction

Les outils SOLAP et les entrepôts de données spatiales (EDS) sont des technologies d'aide à la décision qui permettent l'analyse spatiale et multidimensionnelle des données.

La qualité des données dans les hypercubes spatiaux est importante car ces données sont utilisées pour prendre des décisions. En effet, une mauvaise qualité des données pourrait entraîner une mauvaise prise de décision. Sans tenir compte de la qualité des données dans les hypercubes spatiaux, les résultats obtenus peuvent être non fiables. En conséquence, les contraintes d'intégrité (CI) deviennent très importantes pour améliorer la cohérence logique de toute base de données, ce qui augmente la qualité d'analyse.

Dans les bases de données spatiales transactionnelles, des contraintes d'intégrité spatiale sont définies le long des modèles conceptuels afin de préserver la qualité des données spatiales (Mostafavi et al., 2004; Vallières et al., 2006). Cependant, dans le processus analytique, d'autres types de contraintes d'intégrité doivent être prises en compte pour maintenir une analyse cohérente à l'aide de l'hypercube. Elle est liée à la manière de naviguer et d'interroger les données en tenant compte des règles d'agrégation, des hiérarchies et d'autres concepts des modèles multidimensionnels.

Certains travaux de recherche ont étudié les contraintes d'intégrité pour les cubes de données (hypercubes) non spatiaux (Hurtado et al., 2005; Ghazzi et al., 2004). Cependant, pour étudier les contraintes d'intégrité d'un cube de données spatiales, il convient d'examiner son caractère spécifique lié en particulier aux caractéristiques des données spatiales.

4.2 Qualité des données géographiques

Les SIG peuvent être considérés comme des systèmes d'aide à la décision sur le territoire. En effet, ils permettent la manipulation des données à référence spatiale de sources multiples et hétérogènes, d'où l'importance de la qualité. Il est donc nécessaire de contrôler leurs paramètres géométriques, topologiques et de localisation. La définition de critères pour la vérification de ces paramètres permet l'intégration dans des bases de données géographiques, et l'utilisation correcte dans des systèmes d'information géographique.

Pour la description des données géographiques, l'échelle d'acquisition constitue un paramètre d'importance irréprochable. Néanmoins, l'utilisation des applications SIG permet de modifier l'échelle de présentation ce qui affecte les informations correspondant à l'échelle de détails des données.

Par exemple, un utilisateur de SIG fait changer l'échelle d'une carte d'une échelle 1/25000 à une échelle de 1/1000000, en conséquence, des petits détails tels que les routes secondaires vont apparaître dans la deuxième carte ce qui n'est pas valide. Il n'y aura donc pas de correspondance entre la description de l'implantation des objets et leur description sémantique. Ceci va générer un déséquilibre et des incohérences parmi les données qui sont générées par des traitements ultérieurs.

Dans la littérature, différents travaux sont proposés pour résoudre les problèmes de cohérence des objets.

Dans les applications SIG, l'implémentation des contraintes d'intégrité relatives en par-

ticulier à localisation et la sémantique des objets spatiaux, permet d'éviter la production des résultats erronés.

La définition des objets spatiaux correspond à leur implantation géographique (ponctuelle, linéaire ou zonale), et la description sémantique à travers des attributs descriptifs. En plus, une description de la topologie définissant les relations spatiales doit être définie.

En conséquence, l'intégrité spatiale cherche à assurer une cohérence géométrique. Par exemple nous devons assurer que les traçons d'une route soient bien connectés. En plus, une cohérence sémantique doit être respectée. Par exemple, deux parcelles différentes ne peuvent pas se chevaucher.

Vu que la définition de l'entité spatiale se fait au niveau structurel (structure de données) et au niveau conceptuel (forme, sémantique et topologie), la cohérence donc, doit tenir compte de tous ces aspects, puisque chacun d'eux peut être une source d'erreur.

La qualité des données géographiques peut être contrôlée à travers deux types de critères : critères quantitatifs (internes), relatifs à la qualité de production des données ; et critères qualitatifs (externes), relatifs à la qualité de l'exploitation, et la fiabilité de l'utilisation.

4.2.1 Critères quantitatifs

Les critères quantitatifs sont des critères internes, et concernent les éléments suivants :

Exhaustivité (Completeness)

L'exhaustivité des données vérifie s'il y a un manque ou des d'objets en superflu, leurs attributs et relations dans un jeu de données selon la norme (ISO/TC 211)¹. L'exhaustivité contrôle donc l'existence des deux cas :

- a) *Omission* : correspondant à une absence de données utiles dans le jeu de données ;
- b) *Commission* : c'est-à-dire la présence de données inutiles dans le jeu de données.

Précision de position

La précision de données correspond à la proximité d'une valeur mesurée par rapport à la valeur réelle représentée (Batini et Scannapieca, 2006). C'est-à-dire l'exactitude de la position spatiale des objets. Elle peut être donc absolue, relative.

Elle dépend donc de la fiabilité des moyens d'acquisition et de traitement de données. La précision peut être mesurée à travers la comparaison avec un autre jeu de données de référence.

Précision temporelle

La précision temporelle correspond à l'exactitude des attributs des entités spatiales. Elle concerne donc :

- *Validité temporelle* par rapport à un système de référence temporelle.

1. <https://www.iso.org/fr/committee/54904.html> (dernière visite le 24/05/2018)

- *Précision des mesures temporelles* par rapport aux valeurs de référence.
- *Cohérence temporelle* par rapport aux événements dans le temps.

Précision thématique (sémantique)

La précision thématique ou sémantique concerne les attributs alphanumériques. Ceci correspond donc à l'exactitude des attributs quantitatifs, la justesse des attributs qualitatifs et la classification de leurs relations.

Cohérence logique

Dans la cohérence logique, on s'intéresse à la description de la conformité des données aux règles de structures conceptuelle, logique et physique de données selon la norme (ISO/TC 211) sus-indiquée. C'est-à-dire l'absence de contradictions par rapport à ces règles. Elle peut être donc :

- *conceptuelle* : conformité des données aux schéma conceptuel ;
- *de domaine* : conformité des valeurs à leurs domaines ;
- *de format* : conformité du stockage
- *topologique* : conformité des caractéristiques topologiques.

4.2.2 Critères qualitatifs

Le deuxième type de critères sur la base desquels la qualité des données géographiques peut être mesurée, concerne les critères qualitatifs. Ils sont définis par la norme (ISO 19113 :2002)². Ils correspondent au :

- *But* : décrit la raison de création de ces données
- *Usage* : décrit les utilisations de ces données par le producteur ou autres utilisateurs.
- *Généalogie* : décrit les informations sur l'historique des données depuis leur création.
Ici, on doit avoir des informations sur les sources et les systèmes d'acquisition ; en plus, les traitements qui ont abouti à la création des données.

4.3 Contraintes d'intégrité dans les bases de données spatio-temporelles

Les systèmes SOLAP sont basés sur les données stockées dans les ED spatiales. Par conséquent, la qualité de l'analyse SOLAP inclut la qualité des données spatiales. Cette caractéristique distingue les cubes de données spatiales des cubes de données traditionnels. Ici, il faut prendre en compte la qualité interne définie par la précision géométrique et sémantique ; et la qualité externe qui est liée à la qualité des besoins de l'utilisateur dans le contexte d'utilisation.

Dans les bases de données spatiales, Cockcroft (Cockcroft, 1997) classe les CI spatiaux en trois catégories :

2. <https://www.iso.org/fr/standard/26018.html> (dernière visite : le 24/05/2018)

1. *CI topologiques* qui incluent toutes les propriétés géométriques et les relations sur les données ;
2. *CI sémantiques* qui concernent le sens des caractéristiques géographiques ;
3. *CI définis par l'utilisateur* qui sont des règles métier dans les bases de données spatiales.

Cependant, dans (Salehi, 2009), l'auteur note certaines limitations lorsqu'on essaie d'appliquer ces CI à des applications réelles.

Dans les bases de données temporelles, une classification formelle pour les CI temporels est introduite dans (Bohlen, 1994). En effet, les CI sont catégorisées en fonction du :

1. *temps de transaction* : quand un fait est stocké dans la base de données ;
2. *temps valide* : quand un fait était vrai dans la réalité, et
3. *temps de transaction et temps valide*, les deux à la fois.

Dans les bases de données spatiotemporelles, des CI statiques et de transition sont intégrées avec les CI spatiaux (Cockcroft, 1997). D'autres CI temporels peuvent être impliqués telle que la limitation du cycle de vie des objets comme la CI qui spécifie la période d'applicabilité d'une règle.

4.4 Qualité d'analyse spatio-multidimensionnelle (SOLAP)

L'analyse spatio-multidimensionnelle à travers les systèmes, peut se faire en se basant sur l'entrepôt de données spatiales. Pour ce fait, en exploitant la structure multidimensionnelle, l'utilisateur cherche à explorer les mesures stockées ou calculées selon les différentes dimensions, en appliquant des règles d'agrégation. Par conséquent, la qualité d'analyse ici, dépend de trois paramètres notamment : la qualité de données, la façon d'agrégation des mesures et la façon de l'exploration des données. (Boulil, 2012).

4.4.1 Selon la qualité de données

Pour déterminer la qualité d'analyse SOLAP vis-à-vis les données spatiales entreposées dans l'EDS, plusieurs études ont traité la façon et les critères devant être examinés. Ces études sont basées sur deux points de vue : (Boulil, 2012).

1. Le premier, concerne la qualité de production des données telles que la précision, l'exhaustivité, la cohérence logique, . . . etc.
2. Le deuxième point de vue, concerne l'adéquation des données pour l'utilisation, c'est-à-dire, est-ce qu'elles répondent bien aux besoins des utilisateurs.

4.4.2 Selon la façon d'agrégation des mesures

La façon dont les données sont agrégées représente un facteur d'une importance irréfutable pour assurer une bonne analyse dans les structures multidimensionnelles. En effet, en plus de la qualité des données dans l'EDS, les indicateurs d'aide à la décision, dépendent aussi de la manière dont les hiérarchies, les fonctions d'agrégation et le type

de mesures sont utilisés dans la manipulation du système SOLAP. Ça correspond à la possibilité d'effectuer des agrégations correctes lors du passage d'un niveau de détail à un autre. (Boulil, 2012) Pour ce fait, dans (Lenz et Shoshani, 1997), les auteurs définissent trois conditions notamment : la disjonction, la complétude et la compatibilité de types.

- *La condition de disjonction* : permet de vérifier le comptage en double des mesures ; et que les membres de dimension se trouvent aux mêmes niveaux.
- *La condition de complétude* : permet de vérifier que tous les membres de dimension et instances de table de faits soient présents pour éviter les agrégats incomplets.
- *La condition de comptabilité de types* : elle permet de vérifier que l'application de la fonction d'agrégation à une mesure à travers toute la dimension est valide. Pour cela, nous devons considérer le type de la mesure, le type de dimension et le type de fonction d'agrégation.

4.4.3 Selon l'exploration des données

L'exploration interactive des données représente la propriété caractéristique d'une application SOLAP. Cette exploration est multidimensionnelle et possède différents niveaux de détail.

Le SOLAP est donc utilisé pour des fins d'aide à la décision, par différents utilisateurs parfois des décideurs, qui peuvent ignorer les caractéristiques de l'analyse multidimensionnelle, et de la consistance des résultats obtenus suite à leurs requêtes.

En effet, ces requêtes multidimensionnelles sont formulées par des combinaisons d'indicateurs et des membres de dimension de façon invalide (Boulil, 2012).

Par exemple, une requête invalide peut être formulée par un utilisateur "Quelles sont les surfaces reboisées à l'union soviétique entre les années 2000 et 2010 ?".

Cette requête est correcte syntaxiquement, et les agrégats sont correctement appliqués. Néanmoins, l'exécution de cette requête retourne des résultats nuls. Ça peut être compris par l'utilisateur qu'il n'avait pas de reboisement à l'union soviétique entre ces deux dates ou bien il y a une absence de données dans l'EDS. Or, l'union soviétique n'existe plus après 1991.

La formulation de cette requête va conduire à des interprétations erronées qui peuvent diriger à prendre de mauvaises décisions. Ce problème se pose dans la plupart des applications SOLAP. Il correspond à une incohérence sémantique des requêtes des utilisateurs.

Il est pratiquement difficile de trouver des mécanismes automatiques qui examinent le sens des requêtes et leurs résultats. Néanmoins, des approches ont été proposées pour éviter ces problèmes. Elles sont basées généralement sur l'interdiction de l'exécution, ou bien la proposition de la reformulation de ces requêtes (Boulil, 2012).

4.5 Contraintes d'intégrité dans les cubes de données spatiales

Dans les applications de type SOLAP, la qualité de décision dépend de la qualité des données entreposées et de la façon dont elles sont explorées. Dans ce contexte, les

contraintes d'intégrité (CI) sont définies dans (Boulil, 2012) comme des assertions, généralement définies dans le modèle conceptuel, qui visent à empêcher l'apparition de données incorrectes dans l'entrepôt de données. Ici, une CI peut être utilisée pour :

1. identifier les données erronées stockées,
2. identifier les requêtes analytiques incorrectes et
3. définir des règles d'agrégation correctes.

Dans un SOLAP, une spécification de contrainte d'intégrité doit contenir une sémantique spécifique prenant en charge les concepts liés au modèle multidimensionnel.

Au niveau conceptuel, les CI des EDS définissent les conditions que les données, les métadonnées ou les requêtes analytiques doivent satisfaire.

Dans (Boulil, 2012), les auteurs proposent que la définition des CI au niveau conceptuel permette de considérer les problèmes de qualité dans les premiers stades du développement. En effet, ils proposent deux nouvelles classifications des CI pour les EDS :

1. La première classification catégorise les CI en fonction des éléments multidimensionnels qu'elles impliquent (*dite classification orientée concepts d'EDS*).

En effet, l'expression conceptuelle de toute CI dépend des concepts EDS concernés (métadonnées, membres, faits, agrégation, ...). La Figure 4.1 résume les différents cas possibles (Boulil, 2012).

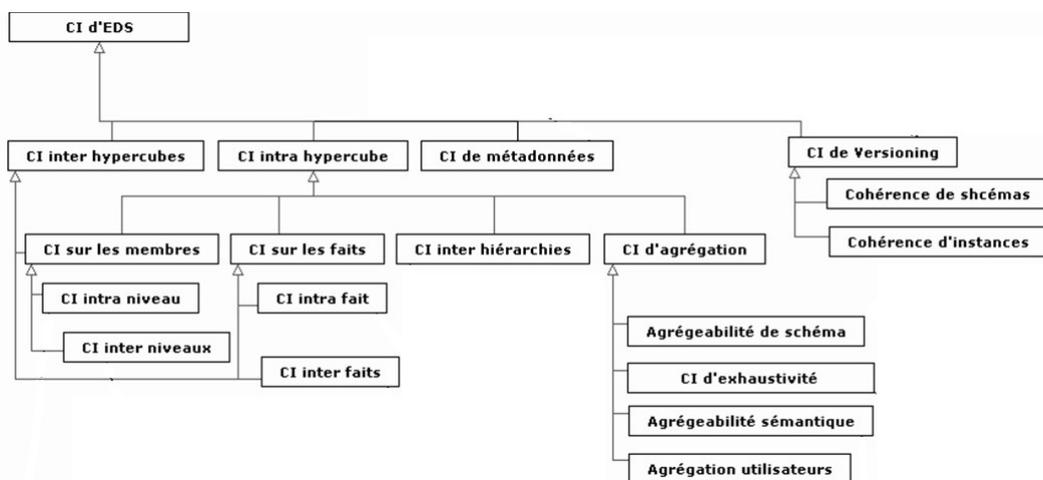


FIGURE 4.1 – Classification des CI en fonction des éléments MD

2. La deuxième classification classe les CI en fonction des niveaux de mise en œuvre dans l'architecture SOLAP (*dite Classification orientée implémentation*).

Dans cette classification, les auteurs distinguent trois catégories de contraintes :

- *Contraintes ETL* : définies au niveau de la solution ETL adoptée, pour éviter de charger des données erronées au cours du processus Extract-Transform-Load
- *Contraintes EDS* : définies sous forme de requêtes SQL au niveau de l'architecture ROLAP pour éviter les faits et les membres de dimension incorrects.

- *Contraintes SOLAP* : définies au niveau du serveur ou du client pour assurer l'utilisation correcte des opérateurs d'agrégation des mesures au sein des hypercubes.

4.6 Classification des langages de spécification des CI

Afin d'exprimer les contraintes d'intégrité des données dans les bases de données géographiques et les EDS, différents langages peuvent être utilisés. Ils peuvent être classés en 4 catégories (Salehi et al., 2007) :

4.6.1 Langages naturels

L'utilisation des langages naturels pour exprimer les CI, est plus facile et est plus expressif, du fait que les utilisateurs peuvent facilement comprendre les expressions en langage naturel. A leur tour, ces langages peuvent être classés en groupes : (Salehi, 2009 ; Bouليل, 2012)

1. Langages naturels libres

Ils correspondent à l'utilisation des langues naturelles sans aucune modification du vocabulaire. La syntaxe ainsi que la sémantique sont aussi respectées.

Par exemple : « les géométries des parcelles ne doivent pas se chevaucher ». Bien que ces langages sont facile à utilisés, ils présentent néanmoins, quelques inconvénients relatifs à l'ambiguïté du sens de certains mots. En plus l'implémentation automatique des expressions des CI en langages naturels est très difficile.

2. Langages naturels contrôlés

Dans cette catégorie, la sémantique et la syntaxe de la langue naturelle sont peu modifiées pour réduire l'ambiguïté de la catégorie précédente.

Les auteurs dans (Ubeda et Egenhofer, 1997), ont proposé un langage proche du langage naturel permettant la vérification de certaines CI.

Par exemple, l'expression qui permet d'interdire l'intersection entre un bâtiment et une route s'écrit comme suit :

(Bâtiment, Cross, Route, Forbidden)

4.6.2 Langages visuels

Dans cette catégorie, pour spécifier une CI, nous utilisons des notations visuelles ce qui permet une lisibilité facile par les utilisateurs.

La Figure 4.2 illustre un exemple de contrainte exprimée par un langage visuel. Cet exemple exprime l'interdiction pour les véhicules et les personnes de se trouver sur le passage piéton au même temps. (Salehi, 2009)

L'expression des CI avec ces langages présente quelques inconvénients, relatif en particulier à l'ambiguïté et la taille des expressions.

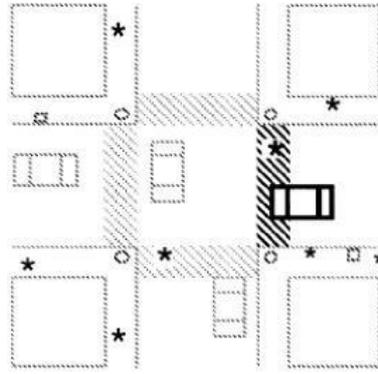


FIGURE 4.2 – CI exprimée en langage visuel

4.6.3 Langages logiques

L'utilisation des langages basés sur la logique du premier ordre pour exprimer les CI est possible. Les expressions précises permettent toute ambiguïté ou mauvaise interprétation.

L'inconvénient de ces langages réside dans la difficulté d'écriture basée sur des connaissances mathématiques chez le concepteur et l'utilisateur.

4.6.4 Langages hybrides

Les langages hybrides sont basés sur la combinaison entre les langages précédents. Ils sont aussi à leur tour, classés en deux groupes (Bimonte et al., 2016) :

1. *Langages hybrides visuels*

Les langages hybrides visuels combinent essentiellement l'utilisation des symboles visuels et la richesse expressive du langage naturel. Ils permettent ainsi d'enrichir la lisibilité chez les utilisateurs, néanmoins, ils ne peuvent exprimer qu'un nombre limité de CI.

2. *Langages hybrides naturels*

Ces langages utilisent essentiellement les langages naturels avec l'utilisation secondaire des pictogrammes ou des symboles.

Ici, nous distinguons donc, des langages qui, en plus de la langue naturelle, introduisent les pictogrammes, qui sont des symboles ressemblant aux objets réels. L'inconvénient de ce type est que les expressions ne sont pas compatibles avec les standards de modélisation (UML par exemple).

Un deuxième type de cette catégorie, intègre des symboles visuels spécifiques avec le langage naturel contrôlé. OCL (Object Constraint Language) et son extension spatiale (Spatial OCL) représentent le meilleur exemple de cette catégorie.

4.7 OCL et Spatial OCL

OCL (Object Constraint Language) est un langage hybride naturel, qui intègre des symboles visuels spécifiques avec un langage naturel contrôlé. Il représente un langage

formel conçu par IBM et adopté par l'OMG³, pour exprimer les CI dans les diagrammes UML. Il permet d'assurer la lisibilité et la facilité de l'expression sans aucune ambiguïté.

Il permet la spécification des conditions devant être satisfaites par les données sans préciser les processus qui permettent de vérifier ces conditions. En plus, l'évaluation d'une CI n'affecte pas l'état des différentes composantes du système.

A travers OCL, le concepteur peut exprimer les CI au stade de conception indépendamment de la plate-forme d'implémentation.

Il peut être, également, utilisé pour l'expression des requêtes (bouilil, 2012).

L'implémentation d'OCL est possible à travers plusieurs outils de modélisation tels que : Rational Rose et MagicDraw. De même que les générateurs de code (Eclipse, Delphi, ...) et à travers les différents langages tels que java, SQL, ...etc. A travers OCL, il est possible de spécifier des invariants sur les classes UML. L'expression des pré-conditions et post-conditions est aussi possible. Dans cette thèse, nous l'avons utilisé beaucoup plus pour la définition des invariants dans le profil UML (Abdallah Bensalloua et Hamdadou, 2018b).

Dans les expressions OCL, un invariant est utilisé pour préciser un ensemble de conditions devant être satisfaites.

Par exemple :

Context Forest inv :

self.type = 'domaniale' implies superficie >= 100

L'utilisation de l'extension spatiale de l'OCL, permet d'exprimer des CI spatiales en exprimant les types géométriques (Point, Ligne et Region), les relations spatiales topologiques (disjoint, meet, overlap, inside, contains, covers, coveredBy et equal) définies par le standard 9IM⁴.

Dans l'exemple suivant, une contrainte Spatial OCL est spécifiée. Elle considère que les géométries des cantons d'une forêt doivent être disjointes.

Exemple :

Context Canton inv :

Canton.allInstances -> forAll(c1, c2 | c1 cl<> c2 implies c1.geo.areDisjoint (c2.geo))

Afin d'implémenter les CI spatiales, nous devons fournir au générateur de code OCL2SQL⁵, les fichiers illustrés dans la Figure 4.3 : (Pinet, 2010)

1. le fichier XMI du modèle conceptuel de données (UML),
2. un fichier de contraintes Spatial OCL,

3. *Object Management Group*, association américaine créée en 1989 pour but de standardiser modèle objet sous toutes ses formes

4. Le modèle à neuf intersections dimensionnellement étendu (DE-9IM) est un modèle topologique et un standard utilisé pour décrire les relations spatiales de deux régions dans la géométrie, la topologie géospatiale et l'analyse spatiale.

5. OCL2SQL est la classe principale de l'outil OCL2SQL (java api). Il gère les différents composants et le flux d'informations entre eux, générant le code SQL. <http://www.docjar.org/docs/api/tudresden/ocl/sql/OCL2SQL.html> (dernière visite le 24/05/2018)

3. un fichier de métadonnées pour les attributs géométriques.

En sortie, nous aurons :

1. des scripts pour la création du schéma physique de la base de données spatiales
2. un ensemble de requêtes, vues et de triggers SQL pour implémenter les CI.

La vue permet de sélectionner les lignes qui ne satisfont pas la contrainte. Si le nombre de lignes n'est pas nul donc le déclencheur (triggers) renvoie un message d'erreur pour le non respect de la CI. (Boulil, 2012)

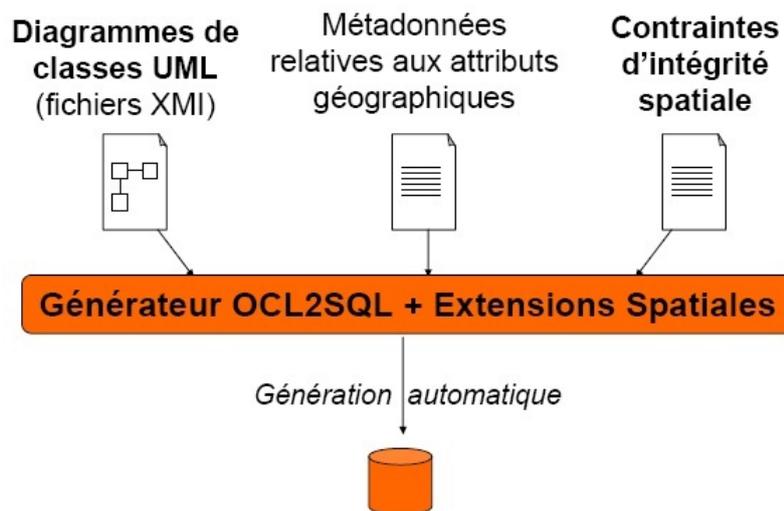


FIGURE 4.3 – CI spatiale avec le générateur de code Spatial OCL2SQL.

4.8 Travaux effectués sur la qualité d'analyse dans les structures MD

La gestion de la qualité dans les systèmes des entrepôts de données spatiales et SOLAP est un axe de recherche entamé par plusieurs études. En effet, la précision dans les EDS a été abordée dans (Siqueira et al., 2011) qui fournissent des modèles logiques et une technique d'indexation pour stocker et interroger des données spatiales.

Dans (Dyreson et al., 2003), les auteurs ont étudié le problème de l'exhaustivité dans l'ED classique. Ils ont proposé une solution prenant en compte les valeurs manquantes dans les hiérarchies. Les incohérences qui se réfèrent à l'existence de contradictions logiques dans les EDS peuvent également être contrôlées par des contraintes d'intégrité (CI). Comme mentionné dans (Boulil, 2012), l'expression des CI sur les modèles conceptuels est essentielle pour prendre en compte toutes les règles de qualité. Pour ce faire, les auteurs dans (Ghozzi et al., 2003) proposent des modèles conceptuels multidimensionnels ad-hoc permettant l'expression de certains CI de données à l'aide de prédicats logiques.

Dans (Malinowski et Zimanyi, 2008), les auteurs proposent une extension du modèle Entité-Relation (ER) pour la conception d'entrepôts de données spatio-temporels. Ils définissent un ensemble de pictogrammes ad-hoc pour exprimer des CI sur des données spatiales (relations topologiques entre membres spatiaux).

Les auteurs de (Glorio et Trujillo, 2008) proposent un profil UML pour les ED, mais ils ne considèrent qu'un très petit nombre de CI de données. Les problèmes d'agrégation sont présentés dans (Mazon et al., 2009) en définissant des contraintes de schémas simples avec des multiplicités UML.

Dans (Salehi, 2009), l'auteur conclut que plusieurs langages de spécification de CI dans des bases de données spatiales ont été proposés mais qu'ils ne sont pas efficaces pour définir des CI dans des hypercubes spatiaux. En effet, l'auteur présente un langage formel pour la spécification des contraintes d'intégrité dans le modèle conceptuel des hypercubes spatiaux. Ce langage est basé sur des langages naturels contrôlés ainsi que des langages hybrides naturels avec des pictogrammes.

Dans (Pinet et Schneider, 2009), les auteurs étudient les contraintes complexes d'agrégation structurelle. En se basant sur un modèle UML, (Boulil, 2012) montre que l'OCL Spatial qui est une extension de l'OCL pour les données spatiales, permet la définition d'un grand nombre de CI sur les données spatiales.

En outre, les langages de modélisation de données ne peuvent exprimer qu'un nombre très limité de contraintes d'intégrité comme par exemple les cardinalités des relations entre classes d'objets. Pour l'expression de CI, de nombreux travaux utilisent des langages non standards, à savoir : logiques, naturels, visuels ou hybrides. D'autres préfèrent OCL car il s'intègre facilement avec UML. Par conséquent, il est nécessaire d'utiliser des langages dédiés à la spécification des CI interopérables avec le langage de définition des structures de données.

Dans (Boulil, 2012), l'auteur présente une méthode pour la modélisation conceptuelle des CI SOLAP après avoir présenté les principaux concepts sur lesquels cette méthode est basée à savoir les profils UML et Spatial OCL.

Dans ce qui suit, nous allons explorer quelques travaux dans la littérature, qui traitent les problèmes de qualité d'agrégation et d'exploration des données spatiales à travers les systèmes (S)OLAP.

4.8.1 Travaux sur la qualité d'agrégation

Nous pouvons classer les conditions d'agrégabilité en deux catégories :

A. Conditions structurelles

Selon les auteurs dans (Mazón et al., 2009) les conditions structurelles d'agrégabilité relatifent à la disjonction et la complétude, sont liées aux types de relations d'agrégation et de hiérarchies de dimension d'une part, et aux relations fait-dimension d'autre part.

Selon Boulil (Boulil, 2012), les approches de modélisation conceptuelle qui traitent ces conditions peuvent être classées en quatre catégories :

1. La première catégorie, regroupe des approches qui interdisent la présence des structures multidimensionnelles complexes. Dans cette catégorie, nous pouvons citer par exemple le modèle ad hoc DFM (Golfarelli et al., 1998). D'autres travaux utilisent les dépendances fonctionnelles pour le même objectif (Niemi et al., 2001 ; Lechtenböcker et Vossen, 2003).

En effet, ils étendent ces concepts des bases de données relationnelles pour les modèles multidimensionnelles.

2. La deuxième catégorie, regroupe des approches qui proposent des modèles conceptuels avec des structures multidimensionnelles complexes sans fournir aucun mécanisme ou solutions pour résoudre les éventuels problèmes d'agrégabilité.

Dans cette catégorie, nous pouvons citer par exemple les modèles proposés dans les travaux de (Abelló et al., 2006 ; Lujan-Mora et al., 2006 ; Glorio et Trujillo, 2008 ; Pinet et al., 2010 ; Salehi et al., 2010)

3. La troisième catégorie, regroupe des approches qui proposent des modèles conceptuels qui comportent des structures multidimensionnelles complexes avec la proposition de mécanismes pour détecter la non agrégabilité mais ne propose pas de solution pour corriger ou rendre ces structures agréables.

Dans cette catégorie, nous pouvons citer par exemple les modèles proposés dans (Hurtado et Mendelzon, 2001) et (Hurtado et al., 2005). Ici les auteurs proposent des contraintes de dimension, pour déterminer l'agrégabilité des hiérarchies non strictes de différentes dimensions. Ces contraintes définissent, en utilisant un langage hybride naturel avec symboles, les chemins d'agrégation possibles des membres à travers des hiérarchies de dimension. Les auteurs ont aussi développé un prototype qui permet de tester l'agrégabilité de ces hiérarchies avec les contraintes de dimension.

Aussi, dans (Pinet et Schneider, 2009 ; Pinet et Schneider, 2010), les auteurs ont défini plusieurs types de contraintes d'agrégabilité structurelles, à travers OCL et UML. Les contraintes pouvant être exprimées, concernent plusieurs relations et chemins d'agrégation.

4. La quatrième catégorie, regroupe des approches qui proposent des modèles conceptuels contenant des structures multidimensionnelles complexes avec la proposition de mécanismes pour détecter la non agrégabilité, et propose des algorithmes pour transformer ces structures en structures agréables.

Dans cette catégorie, nous pouvons citer par exemple le modèle proposé dans (Pedersen et al., 1999). Ici, les auteurs ont proposé des algorithmes qui permettent de transformer les hiérarchies de dimension irrégulières en instances de hiérarchies régulières qui respectent les conditions structurelles de l'agrégabilité.

Les auteurs dans (Mazón et al., 2006) proposent un ensemble de relations QVT⁶ qui permettent de transformer le schéma conceptuel pour satisfaire certaines des formes normales multidimensionnelles, ce qui garanti l'agrégabilité structurelle.

6. Query-View-Transformation language. <http://www.omg.org/spec/QVT/About-QVT/> (dernière visite le 24/05/2018)

Les auteurs dans (Malinowski et Zimányi, 2008) proposent une définition d'un ensemble de techniques permettant d'implémenter les hiérarchies conceptuelles dans les SGBD disponibles, ce qui garantit les conditions structurelles de l'agrégabilité. Mais sans définir l'ordre de traitement s'il existe plusieurs irrégularités structurelles dans la hiérarchie de dimension.

D'autres travaux, dans le même objectif de cette catégorie, sont effectués dans le domaine des EDS. Nous pouvons citer à titre d'exemple, (Pedersen et Tryfona, 2001) dans lequel, les auteurs font une analyse de l'influence des relations spatiales topologiques sur l'additivité des mesures géométriques. Ils ont conclu par ailleurs que cette influence dépend de la disjonction des objets spatiaux. Ils ont proposé ainsi, une approche permettant la transformation du modèle pour garantir l'agrégabilité si les objets spatiaux ne sont pas disjoints.

dans (Jensen et al., 2004), les auteurs ont défini une technique permettant l'évaluation des chemins non précis dans d'agrégation. A travers cette technique, ils peuvent définir quel est le meilleur chemin parmi plusieurs alternatives.

Il est à noter que les travaux cités ici concernent le cas des hiérarchies de dimension irrégulières.

Dans le deuxième cas relatif aux relations fait-dimension non strictes ou incomplètes, peu de travaux qui ont traité le problème d'agrégabilité. Nous pouvons citer principalement (Mazón et al., 2008). Dans ce travail, les auteurs définissent des mécanismes permettant la conversion des modèles multidimensionnels conceptuels contenant des relations fait-dimension irrégulières en un modèle agrégable correctement.

Au final, nous constatons que si le modèle multidimensionnel conceptuel n'était pas agrégable dès sa conception, l'application de l'une de ces techniques va complexifier l'analyse SOLAP. En effet, l'utilisateur va trouver des résultats qui n'ont pas de sens du fait qu'une grande partie de données résultent des insertions arbitraires par ces méthodes.

B. Conditions sémantiques

Le deuxième type de condition d'agrégabilité est relatif aux conditions sémantiques. Ceci concerne en particulier la ressemblance entre la mesure de la fonction d'agrégation et de la dimension, impliquées dans la même requête.

Dans (Lenz et Shoshani, 1997), les auteurs proposent des solutions pour identifier la présence d'incompatibilité entre les mesures, les fonctions d'agrégation et les dimensions sans proposer des solutions pour le matérialiser au niveau du modèle conceptuel.

Dans (Malinowski et Zimányi, 2008) : les auteurs proposent des solutions pour représenter le type additif de la mesure seulement.

Les auteurs de (Malinowski et Zimányi, 2008) représentent le type de mesure par des symboles visuels.

Dans (Nguyễn et al., 2000b; et Ravat et al., 2005), les auteurs proposent des modèles pour exprimer les incompatibilités entre les mesures et les fonctions d'agrégation uniquement. Ils ne permettent pas la prise en compte des dimensions.

Les auteurs dans (Bimonte et al., 2009) ont défini un modèle spatio-multidimensionnel prenant en considération dans le cas d'agrégation des mesures spatiales, les dépendances entre les fonctions d'agrégation spatiales et alphanumériques. Pour contrôler la qualité de l'agrégation à travers les mesures, les auteurs proposent deux contraintes : une contrainte d'agrégation verticale qui examine le type d'additivité de la mesure et de la dimension ; et une contrainte d'agrégation horizontale qui examine le type de mesure et la fonction d'agrégation spatiale utilisée pour l'agrégation des valeurs géométriques.

Les auteurs dans (Prat et al., 2010) définissent différentes contraintes d'agrégation sémantiques à travers un modèle multidimensionnel UML en utilisant le langage PPR⁷ (Production Rule Representation language). Néanmoins, les auteurs ne proposent pas de mécanisme pour implémenter ces contraintes.

Dans (Salehi, 2009), six types de contraintes d'intégrité sont définies. Ces contraintes concernent l'agrégabilité illustrée par la Figure 4.4. Chacun de ces types définit une règle pour interdire une combinaison spécifique dans une agrégation.

En utilisant un langage naturel contrôlé, ils expriment par exemple la contrainte M-D-AF qui permet l'interdiction d'agréger une mesure par une fonction à travers une dimension donnée.

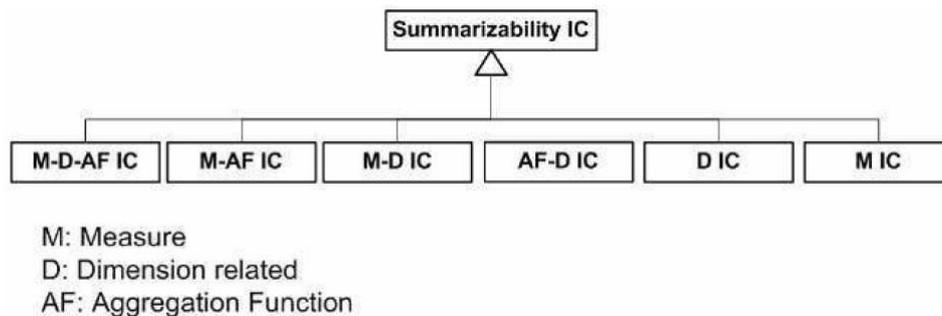


FIGURE 4.4 – Types des CI d'agrégabilité

4.8.2 Travaux sur la qualité d'exploration

Plusieurs travaux existent dans la littérature traitant les problèmes de qualité d'exploration dans le domaine de l'OLAP. Nous citons à titre d'exemple celui effectué par les auteurs dans (Sapia, 1999), dans lequel, ils modélisent une séquence de requêtes pour d'analyse OLAP, en se basant sur la notation graphique montrée dans la Figure 4.5.

Les auteurs définissent donc, les requêtes OLAP par deux composants : le template exprimé par 'SELECT' définissant les mesures et les niveaux d'agrégation devant s'afficher ; et le filtre exprimé par 'WHERE' pour les conditions de sélection des membres des dimensions figurant dans la requête. En plus, certaines contraintes structurelles sont définies pour l'enrichissement de cette proposition.

Les auteurs dans (Böhnlein et al., 2002) proposent des contraintes pour vérifier que la mesure et le niveau d'agrégation appartiennent au même cube de données. En plus, leur

7. <http://www.omg.org/spec/PRR/About-PRR/> (dernière visite le 24/05/2018)

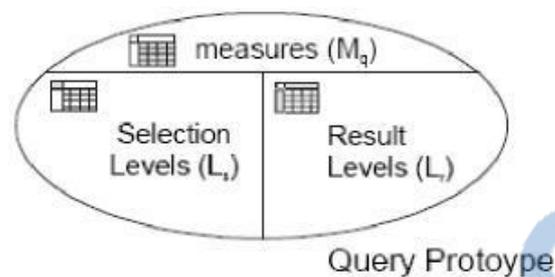


FIGURE 4.5 – Représentation d'un patron de requêtes OLAP

modèle interdit de combiner des niveaux d'agrégation qui appartiennent aux hiérarchies différentes pour une dimension quelconque.

Dans le domaine des systèmes SOLAP, et pour éviter les conséquences d'une mauvaise interprétation des résultats, les auteurs de (Levesque et al., 2007) proposent une approche pour l'aide à gérer les risques de mauvaise utilisation des cubes de données spatiales.

4.9 Conclusion

L'assurance de la qualité des données et leur cohérence, et par conséquent, la fiabilité d'un système d'information décisionnel, peuvent être atteintes grâce au contrôle des CI de différents types. Celles-ci peuvent être spécifiées au stade de la conception de modèles ; ce qui permet de rajouter la sémantique aux modèles.

Pour ce fait, des solutions pour la description des contraintes d'intégrité sont proposées. Elles sont destinées aux bases de données spatiales de même que pour les structures multidimensionnelles notamment les EDS.

Différents outils et méthodes existent permettant la spécification, au même temps, des contraintes sur les données traditionnelles et sur les données spatiales. Nous parlons ici de l'OCL et son extension Spatial-OCL.

Dans ce chapitre, nous avons traité les problèmes de contrôle de la qualité d'analyse dans les systèmes SOLAP. Nous avons montré que cette qualité dépend de trois aspects : la qualité des données entreposées, la qualité d'agrégation et la qualité d'exploration.

Nous avons souligné la différence entre la qualité interne des données spatiales entreposées dans les structures multidimensionnelles notamment les EDS ; et la qualité externe relative à leur efficacité par rapport aux exigences des utilisateurs.

Nous avons ensuite présenté les critères quantitatifs de qualité des données, relatifs à l'exhaustivité, la précision spatiale, la cohérence logique, etc. en plus les critères qualitatifs sont aussi présentés.

Nous avons, également, montré que la qualité d'agrégation dans un système Spatial OLAP, dépend de deux types de conditions. En effet, le premier type concerne les conditions structurelles relatives aux types de relations fait-dimension ainsi que les hiérarchies. Le deuxième type concerne les conditions sémantiques relatives à la cohérence sémantique entre les mesures et les fonctions d'agrégation appliquées d'une part, et de la hiérarchie

de dimension impliquée d'une autre part.

Nous avons souligné aussi que la qualité d'exploration dépend de la cohérence structurelle et sémantique des requêtes SOLAP.

Nous avons indiqué ensuite, que les Contraintes d'Intégrité sont la meilleur façon jusqu'au temps présent, pour vérifier la cohérence logique des données. Dans la même optique, nous avons discuté les différents langages permettant la spécification des contraintes d'intégrité dans les bases de données spatiales ainsi que les structures multidimensionnelles tout en soulignant l'avantage du langage OCL et son extension spatiale.

Nous avons exploré ensuite une nouvelle CI pour les faits dans le schéma en constellation. Aussi nous avons effectué une classification des contraintes d'intégrité pour les utilisateurs d'un SOLAP.

Nous avons terminé ce chapitre par un état de l'art qui explore différents travaux effectués pour exprimer les contraintes d'intégrité dans les entrepôts de données spatiales.

Dans la deuxième partie de cette thèse, nous allons exposer nos contributions relatives à l'implémentation, dans le chapitre 5, d'un système d'aide à la décision nommé Silivicultura fondé sur une approche intégrée SOLAP-AMC. Ensuite, dans le chapitre 6, nous allons explorer notre deuxième système pour la vérification des contraintes d'intégrité relatives aux utilisateurs de SOLAP.

Deuxième partie

Contributions

Chapitre 5

Approche intégrée (SOLAP-AMC) pour l'aide à la décision. Application dans la gestion agroforestière

Dans ce chapitre, nous allons présenter notre première contribution dans cette thèse. Ceci concerne le système 'Silvicultura' qui est basé sur une approche intégrée SOLAP-AMC pour l'aide à la décision dans la gestion agroforestière. Une étude de cas est présentée pour l'expérimentation de différents modèles et prototype du système 'Silvicultura' avec un jeu de données réelles, relatives aux espaces agroforestiers de la wilaya de Mostaganem en Algérie.

Sommaire

5.1	Introduction	83
5.2	Description du système d'aide à la décision spatiale proposé "Silvicultura"	84
5.2.1	Sous-système de « base de données »	85
5.2.2	Sous-système de « surveillance »	85
5.2.3	Sous-système de « traitement des problèmes »	85
5.2.4	Sous-système « Interface utilisateur »	86
5.3	Le modèle décisionnel adopté	86
5.3.1	Structuration du Modèle	87
5.4	Exploitation du modèle	88
5.4.1	Proposition de critères de gestion	88
5.4.2	Les méthodes d'analyse multicritères utilisées	88
5.4.3	Déduction des indicateurs et des vérificateurs	89
5.4.4	Analyse spatio-multidimensionnelle (SOLAP)	89
5.4.5	Proposition d'actions candidates	89
5.5	La démarche décisionnelle adoptée par "silvicultura"	90
5.6	Modelisation UML pour "Silvicultura"	91

5.6.1	Diagramme d'activité de " <i>Silvicultura</i> "	91
5.6.2	Diagramme des cas d'utilisation de " <i>Silvicultura</i> "	91
5.6.3	Diagramme de séquence de " <i>Silvicultura</i> "	92
5.7	Modélisation MD de l'Entrepôt de Données Spatiales pour l'Agroforesterie	93
5.7.1	Profil UML pour les cubes de données spatiales	93
5.7.2	Schéma en constellation pour l'EDS de l'agroforesterie	96
5.8	Etude de Cas : Résultats et Discussion	99
5.8.1	Données utilisées	99
5.8.2	« ELECTRE I » pour le choix des critères de gestion	100
5.8.3	Traitement analytique en ligne spatial	101
5.8.4	Analyse Multicritères (ELECTRE III) pour classer les actions à exécuter	104
5.8.5	Agrégation	105
5.8.6	Exploitation	105
5.9	Conclusion	105

5.1 Introduction

Les espaces forestiers sont en dégradation continue depuis des années. Cette dégradation concerne toutes les composantes écologiques relatives aux espèces végétales et animales ; et même dans certains cas, concerne la topologie des sols. Les causes de ces phénomènes sont multiples. Les facteurs liés aux phénomènes atmosphériques affectent, en grande partie, la stabilisation de ces composantes. Néanmoins, le facteur humain représente le premier ennemi de la forêt. En effet, la majorité des feux de forêts sont d'origine humaine intentionnellement ou pas. En plus, l'exploitation abusive des ressources forestière, peut conduire à l'épuisement du patrimoine forestier.

Il est à noter deux caractéristiques principales de la forêt algérienne. La première relative au caractère fragile des écosystèmes, par conséquent, elle est considérée comme forêt de protection et non pas de production. La deuxième est relative à l'existence de l'agglomération rurale intensive, et qui considère la forêt comme une source de survie.

Actuellement, nous remarquons des efforts des responsables de la gestion forestière, à plusieurs axes de réflexion, pour le but de préserver, régénérer et développer le patrimoine existant, à travers des plans de gestion et d'aménagement des forêts.

Dans cette optique, l'intégration (SOLAP-AMC) peut aider les décideurs à gérer efficacement les zones agroforestières. En effet, cette intégration permet de bénéficier de trois solutions à savoir :

1. l'analyse spatiale et la représentation cartographique assurées par le SIG ;
2. l'analyse multidimensionnelle offerte par les systèmes OLAP ; et
3. l'AMC qui permet la résolution des problèmes complexes avec des exigences souvent conflictuelles, notamment les critères écologiques, économiques techniques et sociales.

5.2 Description du système d'aide à la décision spatiale proposé "Silvicultura"

Un système d'aide à la décision spatiale (SADS) est un système interactif conçu pour aider à la prise de décision. Il est développé pour résoudre un problème avec une base de données spatiale et / ou lorsque la solution aura une dimension spatiale.

Le SADS que nous proposons, à savoir "Silvicultura", est basé sur la détermination du rôle de SOLAP et de l'AMC dans un processus d'aide à la décision pour la gestion des écosystèmes agroforestiers comme montré dans la Figure 5.1 (Abdallah Bensalloua et Hamdadou, 2018a).

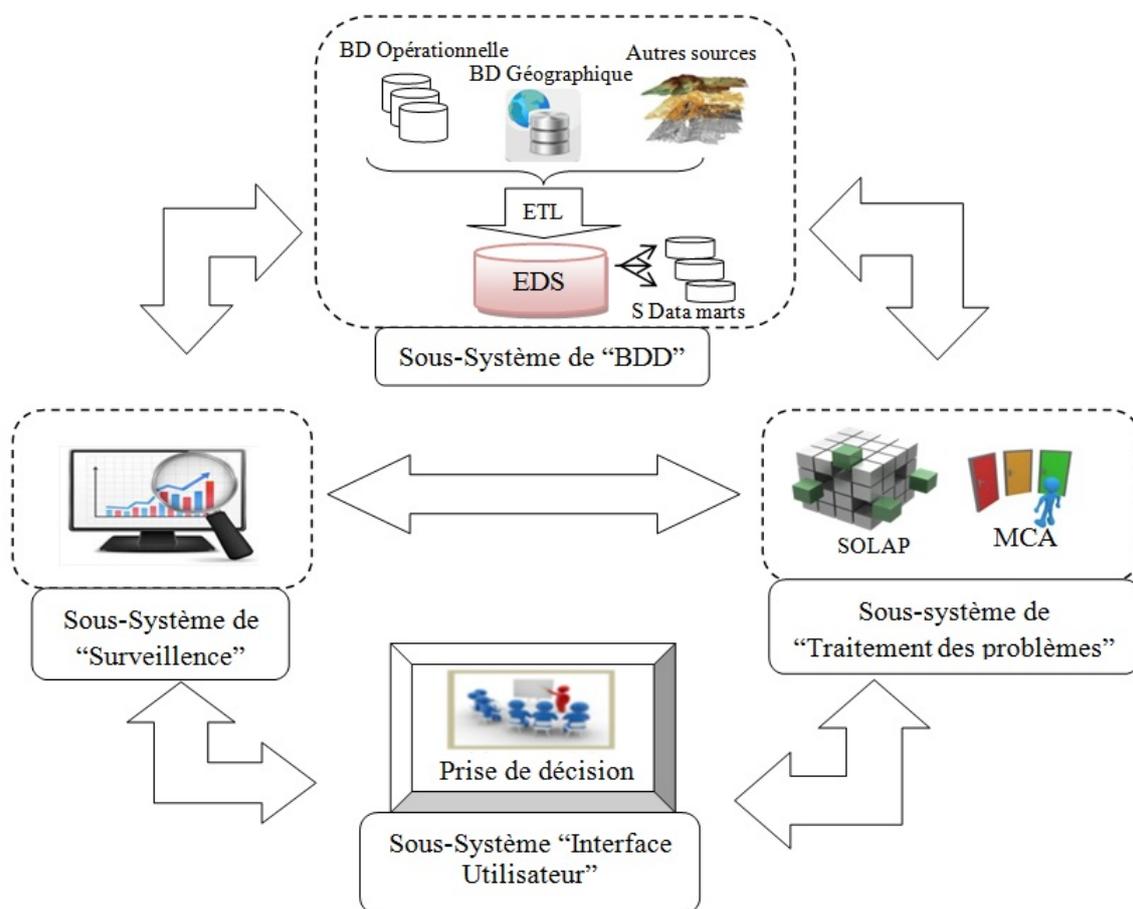


FIGURE 5.1 – Le système d'aide à la décision spatiale "Silvicultura". Une vue d'ensemble

Le développement du système d'aide à la décision spatiale "Silvicultura" résulte de la combinaison de quatre sous-systèmes, à savoir :

1. Sous-système de base de données,
2. Sous-système de surveillance,
3. Sous-système de traitement des problèmes et
4. Sous-système d'interface utilisateur.

Nous détaillons dans les sections suivantes les différents sous systèmes de "Silvicultura".

5.2.1 Sous-système de « base de données »

Le sous-système de base de données est développé afin de construire l'entrepôt de données spatiales (SDW) incluant toutes les données conventionnelles et spatiales liées à la gestion de l'agroforesterie. Les sources de données sont issues du processus opérationnel, de la base de données géographique liée aux applications SIG et d'autres sources telles que la télédétection. Le processus ETL est appliqué pour l'extraction, la transformation et le chargement des données dans le SDW.

5.2.2 Sous-système de « surveillance »

Le sous-système de surveillance permet de bénéficier des connaissances des experts. La tâche des experts forestiers consiste à proposer des critères de gestion. Les auteurs dans (Guillermo et Phil, 2000) estiment que les critères de gestion sont le sens de l'application du principe général dans la politique de gestion forestière. Par exemple nous pouvons considérer le critère de gestion suivant :

« Les principaux processus et fonctions de l'écosystème forestier sont maintenus »

afin assurer le principe général disant :

« pour la gestion durable des forêts, l'intégrité de l'écosystème est maintenue ».

Des paramètres subjectifs sont également intégrés pour l'application de l'AMC afin de réduire la liste des critères de gestion ou *management_criteria* (MC) dans le sous-système de traitement des problèmes. Les experts utiliseront la liste réduite des critères de gestion pour générer des indicateurs qui représentent des variables caractérisant l'état des critères ; et aussi pour générer des vérificateurs, qui sont des données qui renforcent la facilité d'évaluation des indicateurs (Guillermo et Phil, 2000). Ces derniers seront transformés en expressions multidimensionnelles (MDX). Après avoir obtenu le résultat de l'analyse spatiale multidimensionnelle, le processus de surveillance est démarré. En effet, selon leur expérience, les experts peuvent voir s'il existe une menace pour l'écosystème agroforestier, puis ils proposeront une liste d'actions afin d'éliminer le risque. Sinon, ils affineront les critères de gestion afin d'améliorer l'équilibre de l'écosystème.

5.2.3 Sous-système de « traitement des problèmes »

Le sous-système de traitement des problèmes contient des méthodes et des algorithmes à exploiter dans "Silvicultura".

À ce titre, l'analyse spatiale multidimensionnelle est appliquée grâce à SOLAP, le cœur du système. Cela permet de répondre aux requêtes MDX envoyées par le sous-système de surveillance.

En outre, l'AMC est appliquée dans deux situations : Premièrement, elle est utilisée pour choisir une liste réduite de critères de gestion (MC) proposée par le sous-système

de surveillance à prendre en compte lors des étapes futures du processus "Silvicultura". Deuxièmement, l'AMC est utilisée pour classer la liste des actions proposées par les experts pour l'élimination de l'écosystème agroforestier à risque.

5.2.4 Sous-système « Interface utilisateur »

Le sous-système Interface utilisateur est une interface homme-machine utilisée principalement par les décideurs et les experts pour l'introduction de paramètres et de critères pour l'analyse multicritères. Il est, également, utilisé pour l'introduction de requêtes MDX pour l'analyse spatiale MD. Les résultats sont, ensuite, affichés sous forme tabulaire, graphique et cartographique. Cette manipulation est automatisée grâce à l'interface graphique qui facilite l'application de différents opérateurs d'agrégation SOLAP.

5.3 Le modèle décisionnel adopté

Le modèle que nous proposons est adapté de celui proposé dans (Hamdadou et Thérèse, 2011). Il met en évidence tous les éléments de la gestion forestière, à savoir les unités de gestion (cantons), les experts et les décideurs d'une part ; et les opérations affectant ses espaces d'une autre part, en leur permettant d'exprimer leurs caractéristiques et leurs influences les unes sur les autres.

Le modèle de décision proposé est itératif et structuré en trois phases principales (Pictet, 1996) :

1. *Structuration du modèle* : Dans cette phase, nous définissons tous les éléments du modèle, relatifs à la gestion de l'agroforesterie.
2. *Exploitation* : Dans cette phase, nous rassemblons tous les éléments utilisés dans les différents processus du modèle.
3. *Réalisation* : Cette phase implique la mise en œuvre des actions et du contrôle.

Nous notons que nous ne prendrons pas en compte la troisième phase qui concerne le lancement des actions et le contrôle des décisions. La Figure 5.2 illustre les différentes composantes de notre modèle décisionnel. (Abdallah Bensalloua et Hamdadou, 2018a)

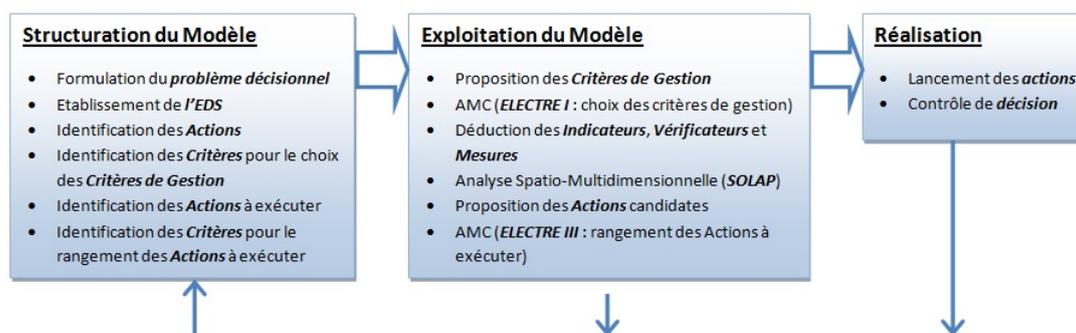


FIGURE 5.2 – Modèle décisionnel de "Silvicultura"

5.3.1 Structuration du Modèle

Dans cette section, nous définissons différents éléments du modèle d'aide à la décision adopté par "Silvicultura".

A. Identification des acteurs

L'identification des acteurs est l'une des tâches les plus importantes lors de la modélisation d'un système d'aide à la décision. En effet, l'absence d'un acteur va nécessairement altérer le résultat du processus qui va générer des recommandations incohérentes. Les acteurs impliqués dans "Silvicultura" sont, principalement :

1. *Le décideur* : c'est le responsable principal de la gestion des zones forestières.
2. *Les experts* : Ils sont directement impliqués dans la stratégie globale de gestion des forêts.
3. *L'homme d'étude (l'analyste)* : Il met en œuvre les méthodes et les modèles d'aide à la décision.

B. Formulation du problème de décision

Le problème de décision doit être formulé dans cette étape. En effet, la question à poser est la première composante à définir. Dans notre étude, le décideur pose la question : «*quelles actions à exécuter et où ?*». Cela permet de déterminer les acteurs intervenant et les autres composants du processus de décision.

C. Établissement d'un entrepôt de données spatiales

Pour l'établissement de l'entrepôt de données spatiales, nous devons commencer par une étape préliminaire de l'analyse des sources de données et de l'identification des besoins des utilisateurs. Le processus d'extraction, de transformation et de chargement des données est ensuite démarré. De même, la structure logique et physique de l'entrepôt de données spatiales doit être prise en compte à ce niveau (Kimball et Ross, 2003).

D. Identification des critères pour choisir les critères de gestion

Cela inclut la famille des critères sur laquelle l'AMC sera appliquée pour choisir une liste réduite de critères de gestion proposés par les experts. Dans (Guillermo et Phil, 2000), les auteurs proposent que l'évaluation des critères de gestion soit assurée par l'AMC. En effet, cette opération doit couvrir tous les acteurs de la gestion forestière ; et elle doit prendre en compte les données quantitatives et qualitatives.

E. Identification des actions à exécuter

La liste des actions est proposée par des experts d'après les résultats de l'analyse spatiale et multidimensionnelle avec SOLAP. Les actions sont liées à différents axes à savoir :

- la protection des espèces de la faune et de la flore,
- la régénération après feu ou la chasse excessive,
- la protection des zones humides,
- l'exploitation des ressources forestières ... etc.

F. Identification des critères de classement des actions à exécuter

La liste des actions doit faire l'objet de l'AMC pour effectuer le classement, selon une liste de critères liés au temps d'exécution, le coût, l'intérêt écologique ... etc, et d'autres paramètres fixés par le décideur et les experts.

5.4 Exploitation du modèle

Notre approche vise à atteindre les points suivants :

5.4.1 Proposition de critères de gestion

Dans la présente étude, nous considérons que les experts sont chargés de fournir la liste des critères de gestion liés à leurs objectifs à défendre. Les critères de gestion peuvent être liés à :

1. la préservation des zones humides, des zones protégées et des espèces naturelles ;
2. la protection des bassins versants, le maintien de l'équilibre naturel, ... etc ;
3. l'animation de l'agriculture de montagne et d'autres fins socio-économiques.
4. le suivi des affaires juridiques de tous les événements qui se produisent dans les zones agro forestières ;

5.4.2 Les méthodes d'analyse multicritères utilisées

L'analyse multicritères est un outil d'aide à la décision dans le cas d'une situation impliquant une ou plusieurs actions possibles évaluées sur un ensemble de critères. Dans le cadre des problématiques liées à la gestion des espaces agroforestiers, elle permet de :

1. éviter les conflits en fournissant une base pour le dialogue acceptable par tous les acteurs présents et influents sur les espaces agroforestiers.
2. aider les décideurs à formuler et/ou à modifier leurs préférences, ou à prendre une décision en fonction de leurs objectifs écologiques, économiques et sociaux.
3. réduire la complexité de la décision dans les cas de présence de plusieurs décideurs et des actions sur le territoire agroforestier (Hamdadou, 2008).

Selon les données fournies et la sensibilité du modèle aux différents paramètres, nous avons adopté deux méthodes d'analyse multicritères procédant par agrégation partielle et appartenant à la famille ELECTRE :

1. *ELECTRE I : Choix des critères de gestion pertinents*

Cette méthode traite le problème décisionnel de choix, appelé Alpha (α). Il s'agit de regarder dans toutes les actions envisagées, un sous-ensemble contenant les meilleures actions «les plus satisfaisantes» (Roy, 1985).

Dans cette étude, nous utilisons cette méthode pour choisir un sous-ensemble de tous les critères de gestion proposés par les experts. Cela permet de réduire leur nombre en choisissant le plus satisfaisant d'entre eux.

2. *ELECTRE III : Actions de classement à exécuter*

Cette méthode aborde le problème du classement. Cependant, ELECTRE III ne cherche pas seulement deux types de surclassement, fort et faible, mais examine une famille qui va du totalement dur au totalement faible ou inexistant, à travers toutes les nuances qui permettent une échelle continue entre ces deux extrêmes. (Roy, 1985)

Dans cette étude, nous utilisons cette méthode pour classer la liste des actions proposées par les experts selon une liste de critères. L'utilisation de cette méthode présente les avantages suivants : (Roussat, 2007 ; Chakhar, 2006 ; Hamdadou et al., 2016)

- Utilisation Mixte de critères quantitatifs et / ou qualitatifs ;
- Intégration du concept d'incomparabilité des alternatives dans toute la procédure de classification ou de classement ;
- Traitement de la non-comparabilité, afin de se concentrer sur les alternatives qui présentent des caractéristiques particulières ;
- Simplicité des comparaisons et, par conséquent, compréhension des résultats ;
- Exploitation des paramètres subjectifs (poids, indifférence, préférence et seuils de veto) attribuant un haut niveau d'efficacité au processus d'aide à la décision.

5.4.3 Déduction des indicateurs et des vérificateurs

Après l'application de l'AMC (ELECTRE I) pour le choix de la liste réduite des critères de gestion, les indicateurs sont déduits. Et puis, les vérificateurs sont identifiés pour chaque indicateur. Pour ces vérificateurs, les mesures relatives sont donc, identifiées.

5.4.4 Analyse spatio-multidimensionnelle (SOLAP)

Les vérificateurs sont similaires aux mesures dans le contexte SOLAP. Ainsi, pour chaque mesure, les requêtes MDX sont lancées en utilisant l'interface graphique SOLAP pour l'analyse spatiale et multidimensionnelle.

5.4.5 Proposition d'actions candidates

Les actions candidates sont proposées par des experts pour résoudre des risques éventuels menaçant les écosystèmes agroforestiers. Ils sont liés à la nature du risque et aux ressources humaines, financières et naturelles disponibles. Par exemple : définition des parcelles pour l'utilisation du bois dans des cantons déterminés ; définition des parcelles réservées à la chasse dans des cantons déterminés, etc.

5.5 La démarche décisionnelle adoptée par "*silvicultura*"

Chaque expert, selon son axe d'intérêt, et sur la base d'un ensemble de paramètres liés à l'unité de gestion (canton), propose un ensemble de critères de gestion (MC).

Après avoir défini tous les critères de gestion, la méthode (ELECTRE I) est appliquée pour choisir une liste réduite parmi celles initialement proposées par les experts à prendre en considération. Cela a pour objectif de choisir les critères de gestion les plus importants afin de faciliter les prochaines étapes du processus.

L'application de la méthode ELECTRE I nécessite la définition d'un ensemble de critères de choix avec leurs poids. Ceci permet de dresser la matrice de performances, Voir Figure 5.11.

Après cela, à chaque critère de gestion de la liste finale correspond un ensemble d'indicateurs. A partir de ces indicateurs, nous déduisons les vérificateurs qui sont similaires aux mesures du concept SOLAP. Ils font l'objet d'une analyse SOLAP. Les experts peuvent alors, déterminer la présence ou non de risques naturels, en s'appuyant sur leurs connaissances. S'il n'y avait pas de risque, une reformulation des critères de gestion est lancée pour améliorer la situation actuelle en respectant les principes fondamentaux et la stratégie globale de gestion forestière. Nous notons ici le mécanisme itératif du processus. Sinon, il y a présence de risque identifié par les experts. Dans ce cas, un ensemble d'actions sera proposé pour enlever ou réduire le risque.

La méthode (ELECTRE III) est appliquée pour classer ces actions, en tenant compte d'une liste de critères tels que la priorité, le coût . . . etc ; et d'autres paramètres notamment les poids de critères, les seuils de préférence, d'indifférence et de vito, Voir Tableau 5.2. La liste finale représente le résultat du processus et fera l'objet de la phase de réalisation.

La démarche décisionnelle adoptée par «*silvicultura* » est décrite dans la section suivante à travers les diagrammes d'activité et de séquences d'UML. Voir Figure 5.3 et Figure 5.5 respectivement. L'algorithme itératif du processus peut s'écrire comme suit :

```
Algorithm Silvicultura process  
Input : C = list of cantons  
Output : Ac = list of actions  
Begin  
Variable  
Risk : Boolean ; MC, Cr, SP1, I, V, MDX, Results : list ;  
{Management criteria, criteria for ELECTRE I, subjective parameters for ELECTRE  
I, indicators list, verifiers list, multidimensional expressions list}  
Get_cantons_characteristics () ;  
Risk ← false ;  
While (risk=false) do  
Begin  
MC ← Propose_management_criteria() ;  
Cr ← propose_criteria_for_choosing_MC() ;  
SP1 ← Propose_subjective_parameters() ;  
MC ← MCA1() ; {MC list is reduced}
```

```
I ← get_indicators(MC);           {indicators are deducted}
V ← get_verifiers(I);             {verifiers are deducted}
MDX ← deduct_MDX(V);             {MDX list are deducted}
Results ← SOLAP(MDX); {spatial_multidimensional analysis launching}
Risk ← Expertise(results); {check risk presence}
If (risk=true) then
  Begin
  Ac ← propose_list_actions();     {proposing actions}
  C ← propose_criteria_for_ranking_Ac(); {propose criteria}
  SP2 ← Propose_subjective_parameters(); {proposition subjective parameters for ac-
tions ranking }
  Ac ← MCA2();                    {ranking the list of actions by ELECTRE III}
  End If
  End While
  ToPronounce (Ac);
  Launching (Ac);
  Controlling (Ac);
END Algorithm.
```

5.6 Modélisation UML pour "*Silvicultura*"

La définition des composants du système permet de définir toutes les interactions au sein de "*Silvicultura*". En fait, cela peut être facilement modélisé grâce au langage UML. Ce dernier a été largement utilisé comme un langage de modélisation pour la conception de divers systèmes et logiciels. Il devient une norme pour la modélisation utilisant le concept orienté objet.

5.6.1 Diagramme d'activité de "*Silvicultura*"

Le diagramme d'activité est utilisé pour modéliser le flux d'activité de "*Silvicultura*". Il donne une vue de haut niveau du système qui est principalement destinée pour les utilisateurs professionnels. La Figure 5.3 illustre le diagramme d'activité de "*Silvicultura*".

5.6.2 Diagramme des cas d'utilisation de "*Silvicultura*"

Le diagramme des cas d'utilisation permet de définir les interactions des acteurs. C'est un ensemble d'actions qui sont exécutées par les acteurs du système. La Figure 5.4 représente le diagramme des cas d'utilisation du processus décisionnel de "*Silvicultura*", dans lequel nous avons considéré trois acteurs et dix cas d'utilisation. Les acteurs sont :

1. *le décideur* : qui effectue la définition du projet, introduit les paramètres subjectifs, contribue avec les experts à proposer des critères pour l'AMC et prend la décision finale;

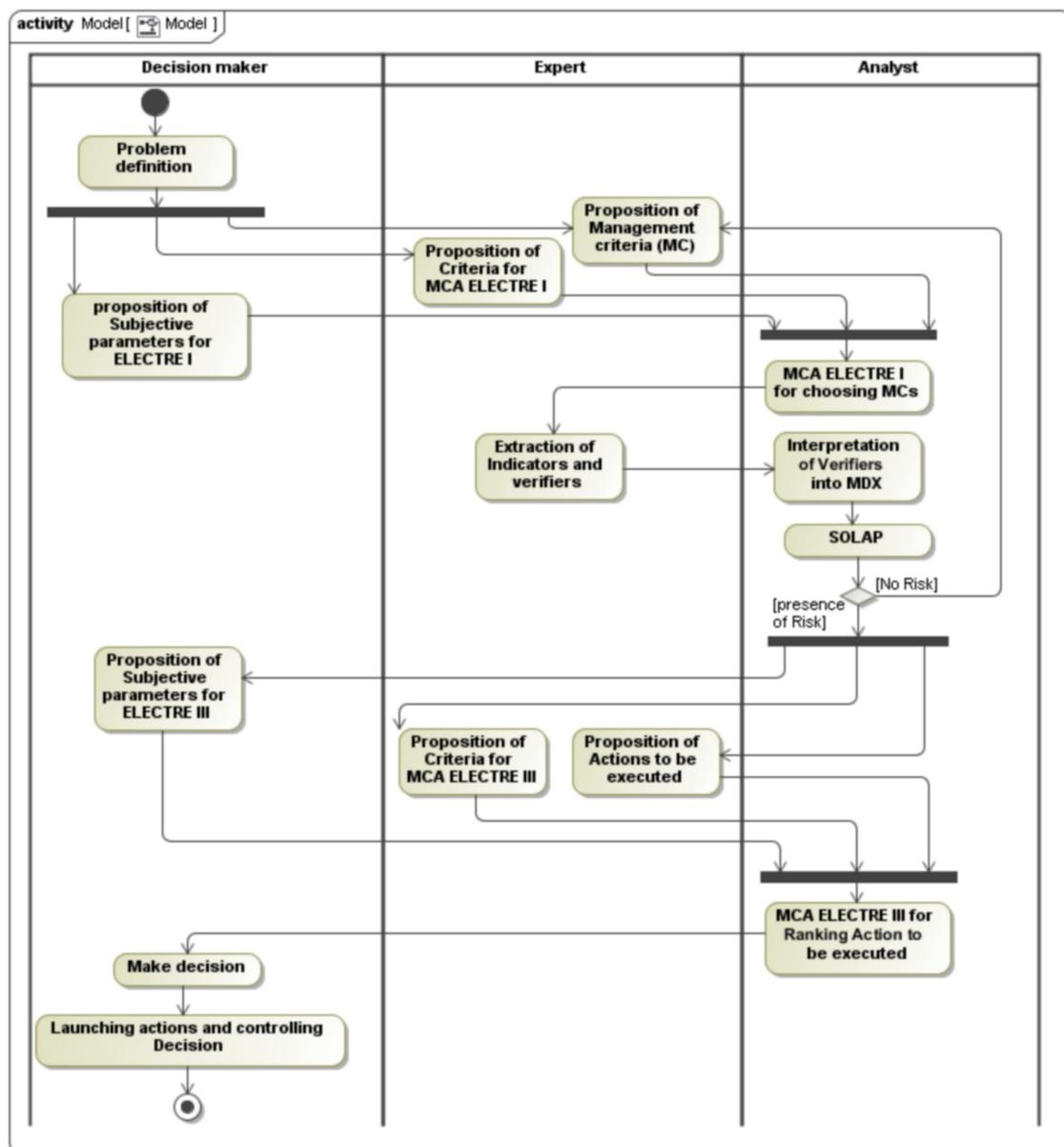


FIGURE 5.3 – Diagramme d'activité de "Silvicultura"

2. *l'expert* : qui effectue la proposition de la liste initiale des critères de gestion, la formulation des vérificateurs et des indicateurs, propose des critères pour l'AMC avec la contribution du décideur, propose des actions à exécuter et effectue l'analyse SOLAP avec la participation de l'analyste ;
3. *l'analyste* : qui effectue les analyses Electre I, Electre III et SOLAP.

5.6.3 Diagramme de séquence de "Silvicultura"

Le diagramme de séquence est une technique de modélisation dynamique. Il se concentre sur l'identification des interactions au sein du système. La Figure 5.5 représente le diagramme de séquence UML pour "Silvicultura" où les interactions entre les acteurs sont

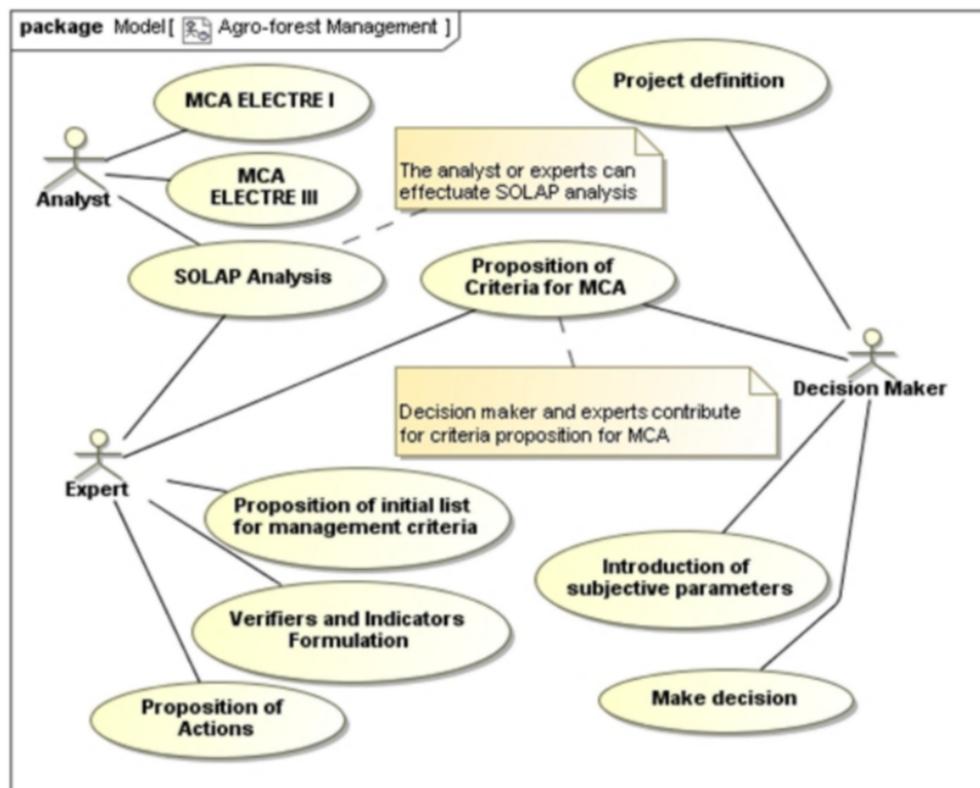


FIGURE 5.4 – Diagramme des cas d'utilisation de "Silvicultura"

montrées. Le diagramme montre trois types d'acteurs : le décideur, les experts et l'analyste.

5.7 Modélisation MD de l'Entrepôt de Données Spatiales pour l'Agroforesterie

La modélisation de l'EDS (SDW pour Spatial Data Warehouse) pour l'agroforesterie est basée sur l'utilisation d'une extension de l'UML pour la modélisation multidimensionnelle des faits, des dimensions et des niveaux hiérarchiques que notre approche comprend (Salehi et al., 2010, Octavio et al., 2012).

5.7.1 Profil UML pour les cubes de données spatiales

L'utilisation de l'UML pour la modélisation MD peut être effectuée en l'étendant avec le concept du profil pour l'adapter à des plates-formes ou domaines spécifiques. Cela permet de personnaliser les métaclasses UML avec trois mécanismes : les stéréotypes, les valeurs étiquetées et les contraintes (Juan et al., 2009).

Le langage OCL (Object Constraint Language) peut être utilisé pour formaliser les contraintes qui affinent les définitions des stéréotypes et des valeurs marquées (Boulil et al., 2014).

Dans notre proposition, le package « SDW core model » est adapté de celui présenté

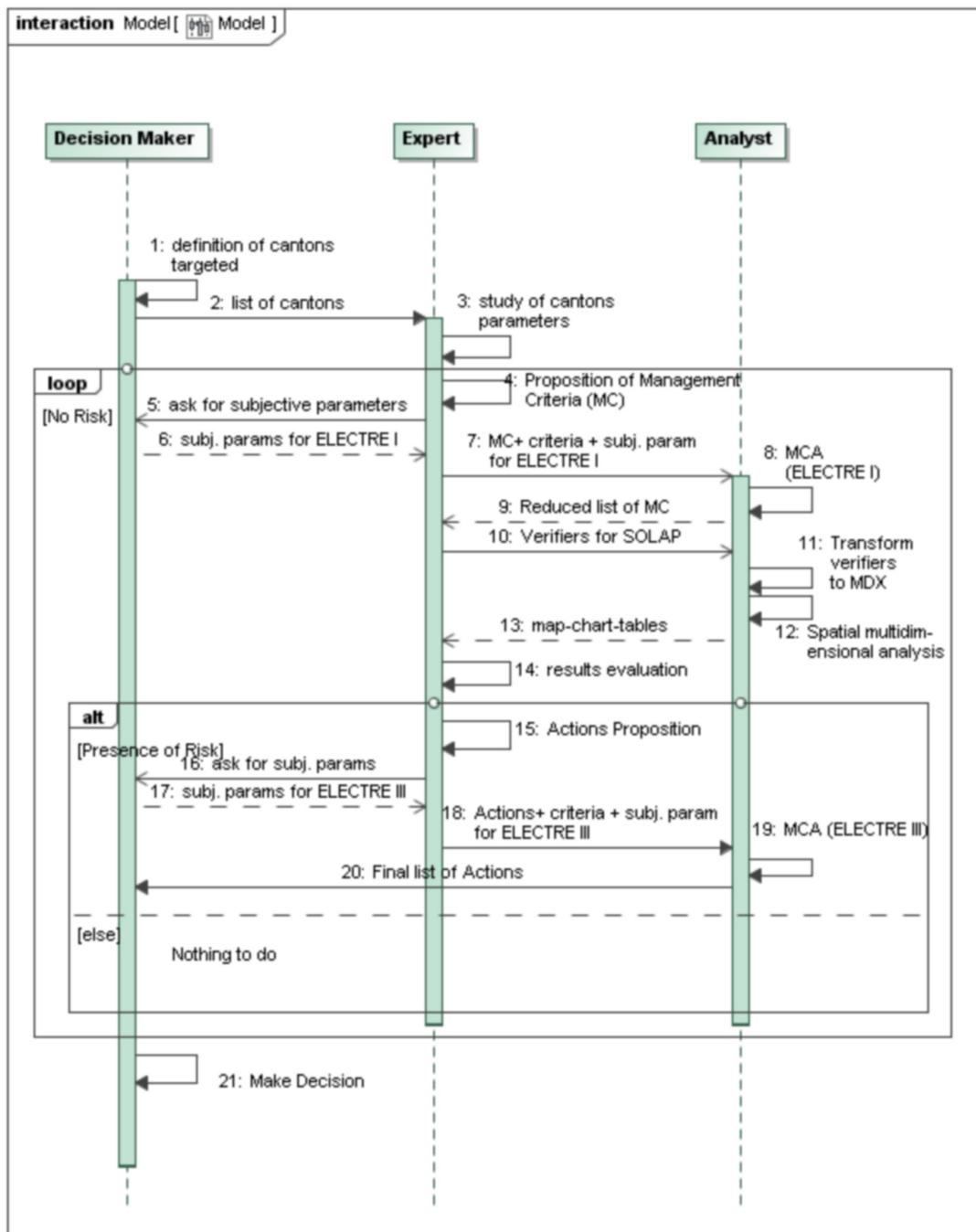


FIGURE 5.5 – Diagramme de séquence du processus décisionnel

dans (Boulil et al, 2015). Il permet de représenter les principaux concepts statiques du SDW. Nous l'avons utilisé grâce à son exhaustivité et sa clarté.

Cependant, nous proposons quelques modifications que nous avons jugé nécessaire, comme le montre la Figure 5.6 :

1. Contrairement au monde OLAP traditionnel, le modèle dimensionnel unifié (UDM) permet d'avoir plusieurs tables de faits dans un cube (Sivakumar et al., 2009). Donc, sur la base de ce principe, nous proposons que la cardinalité dans la relation de

composition entre le stéréotype de l'hypercube et celle de la classe de faits soit "one-to-many" voir Figure 5.6.

2. En plus de la relation de composition entre l'hypercube et les stéréotypes de dimension, nous proposons une association entre les stéréotypes de faits et de dimensions, car un hypercube peut contenir plusieurs tables de faits. Ceci est obligatoire pour définir quelle table de faits concerne quelles dimensions dans le même hypercube. (Abdallah Bensalloua et Hamdadou, 2018a)
3. Une relation d'association est créée entre les faits et les stéréotypes niveau_d'agrégation « AggLevel » pour faciliter le dimensionnement.
4. Une relation d'association est également créée entre les stéréotypes de mesure et de hiérarchie. Cela peut définir quelle fonction d'agrégation est applicable sur quelle hiérarchie.

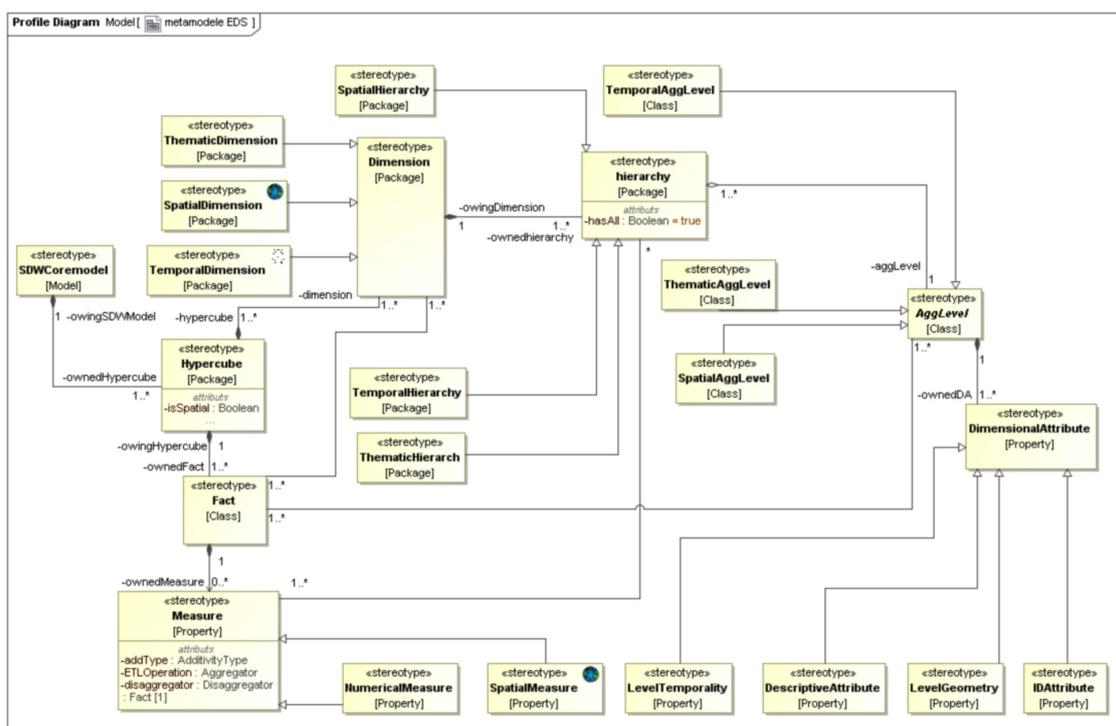


FIGURE 5.6 – Stéréotypes du profil SDW pour "Silvicultura"

Les auteurs dans (Bouilil et al., 2015) considèrent qu'un hypercube spatial doit contenir au moins une dimension spatiale ou au moins une mesure spatiale, cependant, dans notre proposition, nous définissons une contrainte d'intégrité dans le contexte du stéréotype hypercube qui considère qu'un hypercube spatial doit contenir au moins une dimension spatiale et au moins une mesure spatiale dans chaque table de faits (voir Figure 5.7). En effet, dans notre proposition, nous devons effectuer une analyse spatiale avec chaque mesure spatiale, telle que la localisation, la zone tampon ... etc.

Pour l'implémentation du profil UML, nous utilisons l'outil "*MagicDraw*"¹ qui permet

1. <https://www.nomagic.com/products/magicdraw> (dernière visite le 24/05/2018)

```
context Hypercube inv AtLeastOneSpatialDimensionAndOneSpatialMeasure:  
isSpatial=true implies (  
ownedMember->select (m|m.oclIsTypeOf(SpatialDimension))->size())>0  
and ownedMember->exists (m|m.oclIsTypeOf(Fact) and  
m.ownedAttribute->select (a|  
a.oclIsTypeOf(SpatialMeasure))->size())>0))
```

FIGURE 5.7 – Contrainte OCL dans le contexte du stéréotype hypercube

de vérifier les contraintes OCL du profil. En effet, les contraintes sont vérifiées par "MagicDraw" au niveau de l'abstraction conceptuelle qui empêche la modélisation incorrecte.

5.7.2 Schéma en constellation pour l'EDS de l'agroforesterie

SDW doit fournir des réponses à toutes les questions des experts. Nous devons assurer la modélisation de toutes les mesures et dimensions requises dans le processus d'analyse. De plus, nous devons prendre en compte la modélisation des hiérarchies utilisées pour les agrégations de données.

Afin de sélectionner les dimensions et les mesures basées sur les données collectées, nous devons prendre en compte les objectifs de l'analyse liée à la gestion agroforestière.

1. **Les Dimensions :** les faits sont analysés selon un ensemble de dimensions décrivant trois catégories :
 - a) *Dimension temporelle* : pour effectuer une analyse au niveau mensuel, trimestriel ou annuel, une dimension temporelle est requise.
 - b) *Dimensions spatiales* : nous utilisons deux dimensions spatiales administratives et écologiques.
 - c) *Les dimensions thématiques* : ces dimensions sont nombreuses et liées aux processus métier.

Dans la Figure 5.8, les dimensions et les hiérarchies sont affichées.

La dimension "*Activity_classification*" enregistre la classification des activités sylvicoles à deux niveaux d'agrégation. Ces niveaux d'agrégation sont formés en regroupant les attributs dimensionnels en sous-ensembles disjoints en fonction des besoins d'analyse (Boulil et al., 2015).

Par exemple, la dimension spatiale "*Ecological_divisioning*" se compose d'une hiérarchie composée de deux niveaux d'agrégation, tandis que l'autre dimension "*Administrative_divisioning*" est composée de trois niveaux.

2. **Mesures :** Ils sont similaires aux vérificateurs indiqués ci-dessus. Ils sont obtenus à partir de requêtes à formuler par les experts en fonction de leurs critères de gestion. Dans notre contexte d'analyse, les mesures peuvent être numériques ou spatiales.

Dans notre étude de cas, les experts s'intéressent à l'analyse de cinq aspects liés directement à leurs compétences à savoir : les activités sylvicoles, le contrôle des événements, la dendrométrie, la gestion de la faune et les aires protégées.

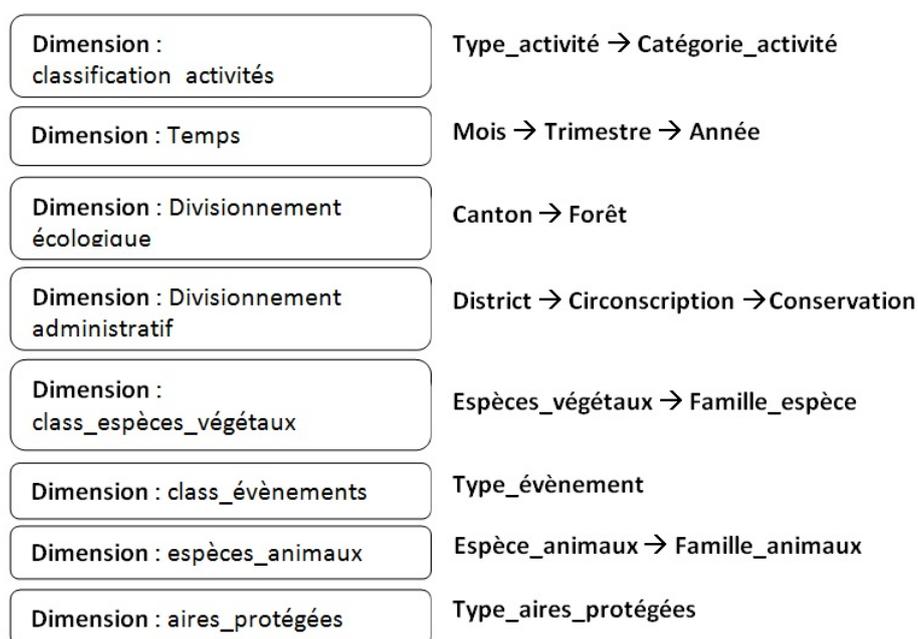


FIGURE 5.8 – Dimensions et hiérarchies

Ces aspects sont définis à l'aide d'un hypercube avec cinq tables de faits partageant certaines tables de dimension. Dans notre proposition, l'hypercube spatial n'est pas lié à une activité mais à un ensemble d'activités. Cela permet de dessiner le schéma en constellation.

Le diagramme de classes de l'EDS pour l'agroforesterie est représenté sur la Figure 5.9. Il est considéré comme spatial car il comporte au moins une mesure spatiale en plus des dimensions spatiales.

Comme le montre la Figure 5.9, la table de faits « Activité_silvicoles » (*silvicultural_act*) est décrite par trois mesures où l'une est spatiale (*localisation*) et deux numériques (*volume_fin* et *superficie*); ainsi que la table de faits « Contrôle_événements » (*control_event*). La table de faits « Contrôle_dendrométrie » (*control_dendrometry*) est décrite par trois mesures numériques (*cir*, *haut* et *age*), en plus d'une mesure spatiale (*localise*). La table de faits « Aires_protégées » (*protected_area*) est décrite en utilisant une mesure spatiale (*localisation*). En fin, la table de faits « Gestion_de_la_faune » (*wild_life*) est décrite en utilisant deux mesures numériques et une mesure spatiale.

Après avoir défini le modèle SDW, nous traitons sa mise en œuvre. Nous utilisons une architecture OLAP relationnelle composée d'un système de gestion de bases de données relationnelles (SGBD) *MS-SQL Server*² avec son « *Analysis Service* »³, qui permet de créer le cube de données et de fournir les opérateurs OLAP tels que : Roll up, Drill down, Slice, Dice, etc.

L'architecture OLAP relationnelle est composée de quatre niveaux :

2. <https://docs.microsoft.com/fr-fr/sql/sql-server/sql-server-technical-documentation?view=sql-server-2017> (dernière visite le 24/05/2018)

3. <https://docs.microsoft.com/en-us/sql/analysis-services/analysis-services?view=sql-analysis-services-2017> (dernière visite le 24/05/2018)

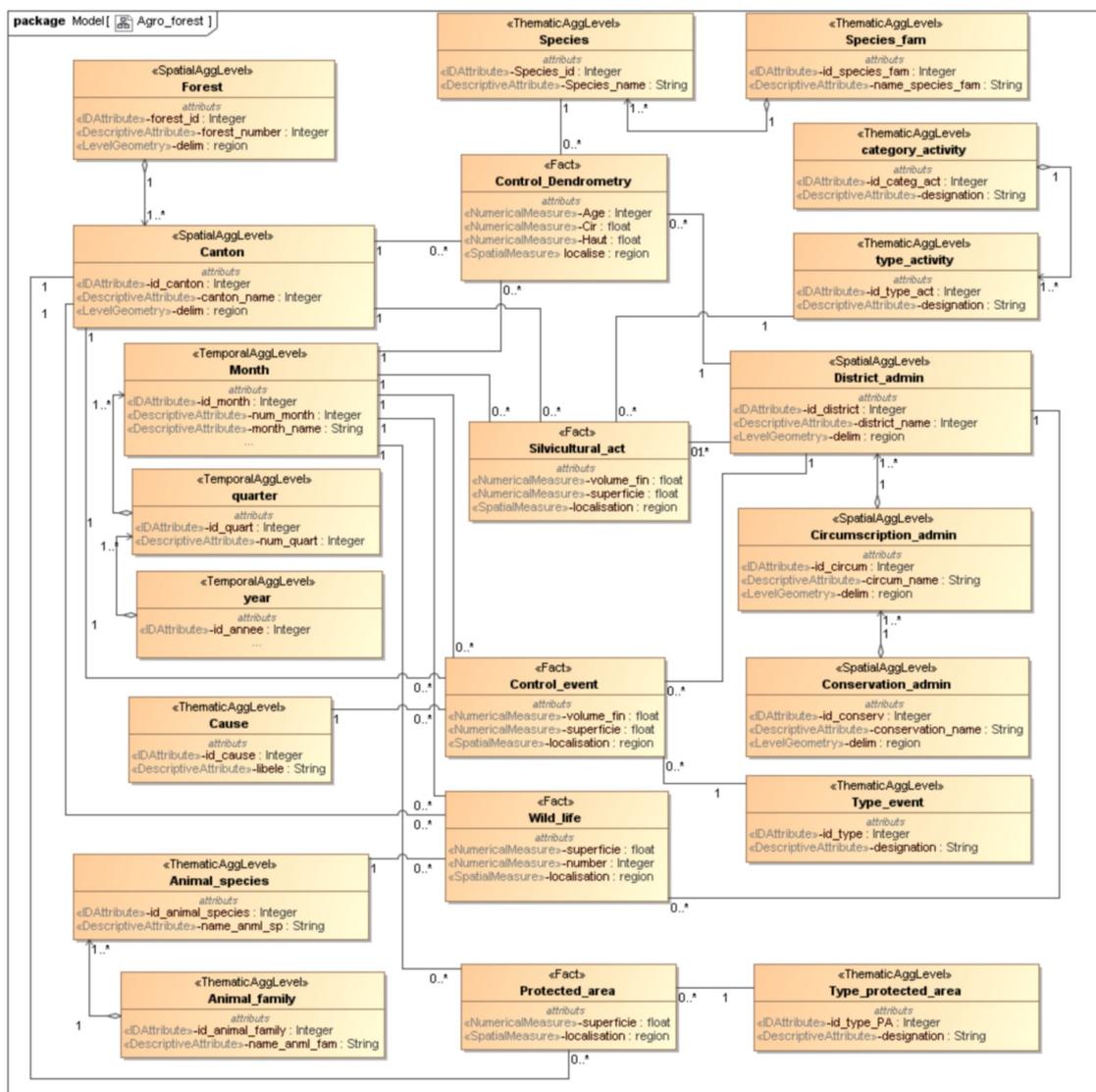


FIGURE 5.9 – Diagramme de classe de l'hypercube "Agro-forest"

1. le niveau ETL (Extract-Transform-Load),
2. le niveau de stockage des données,
3. le serveur OLAP et
4. le client OLAP.

Nous notons que les techniques et procédures ETL complexes dépassent le cadre de cette thèse.

Le client SOLAP à savoir "Sylvicole" est développé avec JAVA Eclipse (Luna Service Release 2 (4.4.2)), avec l'intégration de *ArcGIS Runtime SDK*⁴. En fait, la solution fournit une synchronisation cartographique avec des affichages tabulaires et diagrammes comme le montre la Figure 5.12.

4. <https://developers.arcgis.com/arcgis-runtime/> (dernière visite le 24/05/2018)

L'analyste et les experts peuvent utiliser le client SOLAP pour effectuer une analyse spatiale et multidimensionnelle. Les résultats permettent de dresser la liste des actions candidates, objet de la phase suivante du processus décisionnel.

5.8 Etude de Cas : Résultats et Discussion

Périodiquement, les responsables du secteur forestier en Algérie, réalisent des études globales sur différentes activités et un inventaire des ressources forestières. Le but de ces études est de fournir un outil d'aide à la décision pour la planification des activités pour les futurs programmes dans ce pays.

5.8.1 Données utilisées

Dans ce qui suit, nous utilisons notre système "Silvicultura", avec un jeu de données concernant les forêts méditerranéennes de la wilaya de "Mostaganem" situé au nord ouest de l'Algérie.

L'alimentation de la base de données géographique peut être effectuée après des opérations de prétraitement ; en particulier le géoréférencement et la numérisation des couches d'information afin d'être intégrées dans la base de données.

Pour la constitution des couches de fond, nous avons utilisé *ARCGIS 10.3*⁵. Pour cela différentes sources d'informations peuvent être exploitées tels que :

- Les cartes topographiques : permettent d'extraire d'autres couches utiles à savoir les limites administratives, le réseau routier, l'hydrographie, ... etc.
- Les limites des forêts, limites des cantons, plans cadastraux et fascicules des propriétés et de gestion.
- Les classes de formations forestières
- Le modèle numérique de terrain et produits dérivés (pentes, exposition, altimétrie, ... etc.)
- ... etc.

La Figure 5.10 illustre une composition cartographique formée de plusieurs couches d'information concernant la wilaya de Mostaganem : (limites administratives, forêts, réseau routier, points d'eau, nomenclature, ... etc.).

Remarque :

Une partie des données utilisées est la propriété de l'administration forestière de Mostaganem et certains détails sont confidentiels. Ils sont généralement liés à des noms de personnes ou d'autres paramètres que nous n'avons pas obtenus mais nous considérons qu'ils ne sont pas très intéressants pour le fonctionnement du système proposé dans la présente étude. Nous devons indiquer par ailleurs, que nous avons sollicité l'expérience des experts forestiers dans différentes phases de cette expérimentation.

5. <https://www.esri.com/fr-fr/store/arcgis-desktop> (dernière visite le 24/05/2018)

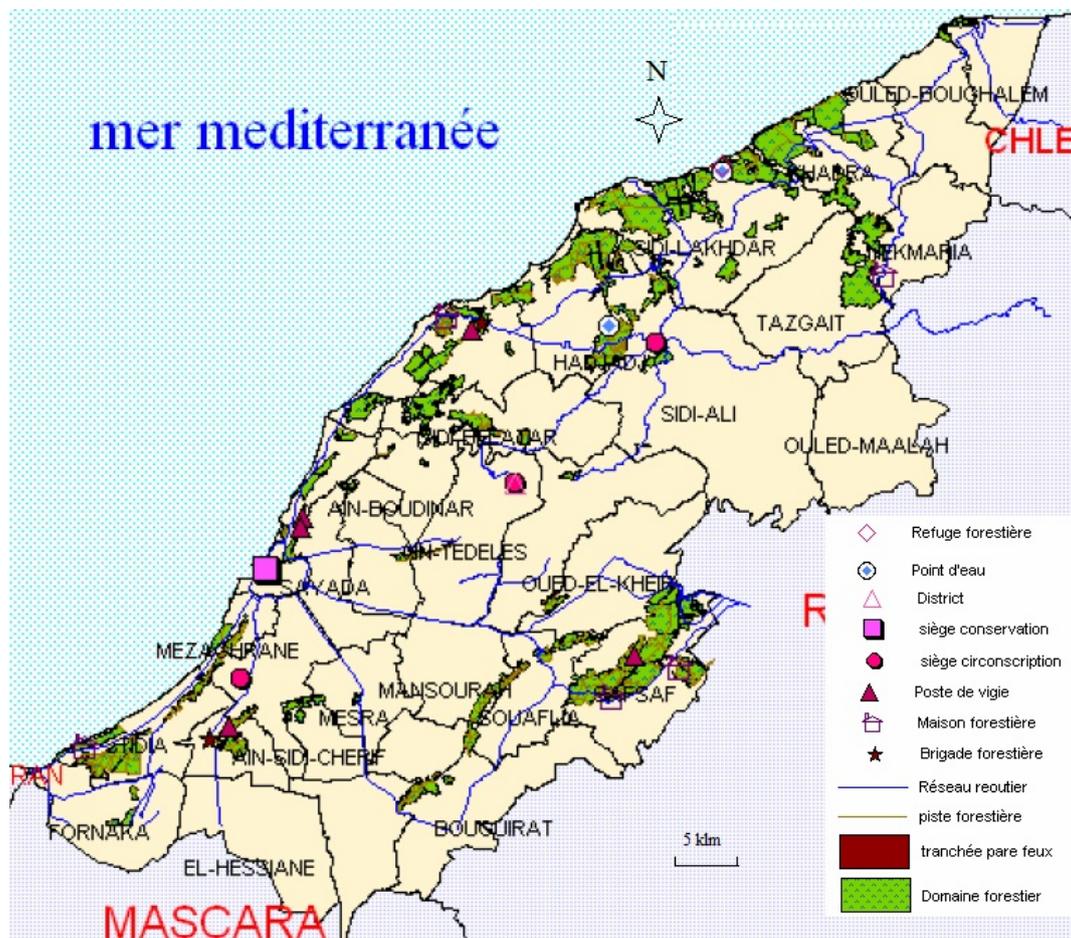


FIGURE 5.10 – Espaces forestiers de la wilaya de Mostaganem

Les données de différents sources feront ensuite, l'objet du processus ETL afin d'être intégrés dans l'entrepôt de données. Ensuite, le cube de données spatiales est créé et les différentes agrégations sont calculées.

5.8.2 « ELECTRE I » pour le choix des critères de gestion

Tout d'abord, le décideur pose le problème décisionnel suivant : "quelles activités devraient être menées à travers un ensemble d'unités de gestion (cantons) ?"

Les experts sont sollicités pour fournir toutes les caractéristiques techniques des cantons en matière de pédologie, de traditions agricoles, de faune et de flore, etc. Sur la base de ces données, ils proposent, l'ensemble des critères de gestion (MC) propres aux cantons ciblés et relatifs à leurs différents axes de réflexion.

Après cela, l'analyste effectue l'AMC (ELECTRE I) pour choisir une liste réduite de critères de gestion (MC) à prendre en considération. Cette analyse vise à prendre en compte toutes les exigences des experts. Les paramètres subjectifs à savoir les poids des critères identifiés, sont fournis par le décideur.

Afin de définir la liste des critères de gestion à prendre en compte, nous réalisons une AMC (ELECTRE I). La liste des critères de gestion « Management Criteria » (MC) est

initialement la suivante :

- MC1 : Préservation du patrimoine existant
- MC2 : Régénération du patrimoine existant
- MC3 : Développement du patrimoine existant
- MC4 : Réhabilitation de l'environnement naturel
- MC5 : Préservation des zones humides
- MC6 : Développement des zones humides
- MC7 : Conservation des aires protégées
- MC8 : Consolidation des aires protégées
- MC9 : Préservation de la frange côtière
- MC10 : Maintenir l'équilibre naturel des écosystèmes
- MC11 : Protection des bassins versants
- MC12 : Animation de l'agriculture de montagne

La participation des experts au processus d'aide à la décision permet de définir la liste de critères suivante (Cr_i, i = 1..5) pour l'analyse des critères de gestion (MC_i, i=1..12) :

- Cr1 : Protection de la flore
- Cr2 : Protection de la faune
- Cr3 : Protection des zones humides
- Cr4 : Extension dans la forêt
- Cr5 : Besoins socio-économiques

Afin d'associer un poids pour chaque critère, nous proposons une échelle qualitative (important = 3, moyen = 2 et faible = 1). Pour ELECTRE I nous proposons la liste des poids suivante : Cr1 = 3; Cr2 = 3; Cr3 = 2; Cr4 = 1; Cr5 = 1. Il est à noter que le poids des critères peut varier selon le processus itératif de "*Silvicultura*".

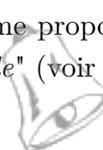
La génération de la matrice de performances est réalisée en collaboration avec tous les experts impliqués dans la définition de la liste initiale des critères. Nous avons utilisé une application développée dans notre équipe de recherche "*Ingénierie de l'Aide Multicritères à la Décision Collective, Distribuée et Temps Réel pour la Gestion Environnementale, le Diagnostic médical et la Surveillance Epidémiologique*", au Laboratoire d'Informatique d'Oran "*LIO*", pour l'obtention des résultats du procédé ELECTRE I. Dans la Figure 5.11, sont présentées les matrices de Performances, de Concordances et de Discordances, et enfin la solution (le sous ensemble choisi).

La solution est le sous ensemble sélectionné : MC4, MC7, MC10, MC12. Nous notons que le choix des paramètres de l'AMC affecte ce résultat. Cette solution signifie que nous serons intéressés par ces critères de gestion. Comme indiqué ci-dessus, ces critères seront transformés en indicateurs et plus tard en vérificateurs.

Dans la Table 5.1, sont décrits quelques exemples de transformation des critères de gestion (MC) en indicateurs puis en vérificateurs.

5.8.3 Traitement analytique en ligne spatial

L'analyse spatiale et multidimensionnelle est le noyau du système proposé. A cet effet, nous avons développé une application SOLAP nommée "*sylvicole*" (voir Figure 5.12).



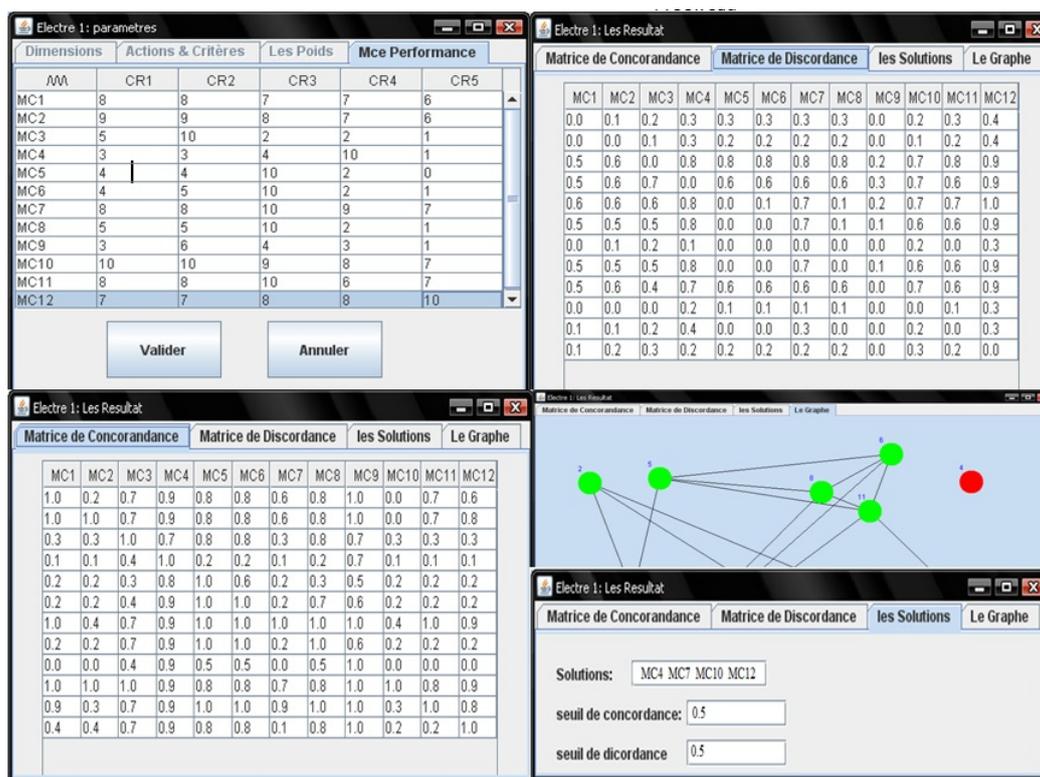


FIGURE 5.11 – Résultat d'ELECTRE I

TABLE 5.1 – Critères de gestion, Indicateurs et Vérificateurs

critères de gestion	Indicateurs	Vérificateurs
MC 4	<ul style="list-style-type: none"> • Enquêter sur les activités de réhabilitation en forêt 	<ul style="list-style-type: none"> • Mesurer le volume financier des activités de réhabilitation réalisées en forêt • Mesurer la zone des activités de réhabilitation réalisées dans la forêt
MC 7	<ul style="list-style-type: none"> • Enquêter sur la situation des aires protégées 	<ul style="list-style-type: none"> • Identifier les aires protégées • Comptabiliser les infractions dans les zones protégées
MC 10	<ul style="list-style-type: none"> • Étude de l'état de la faune et de la flore 	<ul style="list-style-type: none"> • Identification des zones de chasse excessive • Faire le recensement de la zone des espèces de la flore et mesurer la dendrométrie
MC 12	<ul style="list-style-type: none"> • Etudier la situation de l'agriculture de montagne 	<ul style="list-style-type: none"> • Recenser les autorisations d'exploitation des ressources forestières • Identifier les zones d'exploitation des autorisations de ressources forestières

Grâce à cette application, l'analyste peut lancer des requêtes analytiques multidimensionnelles (MDX). Ces requêtes sont automatisées dans l'interface graphique. Cela permet de visualiser les résultats sous forme de tableaux, de cartes ou de graphiques.

Dans la Figure 5.12, un exemple de visualisation avec l'application "sylvicole" est illustré. En exploitant la première table de faits « *Dendrométrie* » (*Control_dendometry*),

nous pouvons voir l'âge (mesure "age") des espèces forestières (*dimension thématique*) dominantes dans les forêts de "Mostaganem" par "Canton" (*dimension spatiale*). Nous pouvons voir la mesure spatiale "délimitation" qui affiche les parcelles dendrométriques correspondantes.

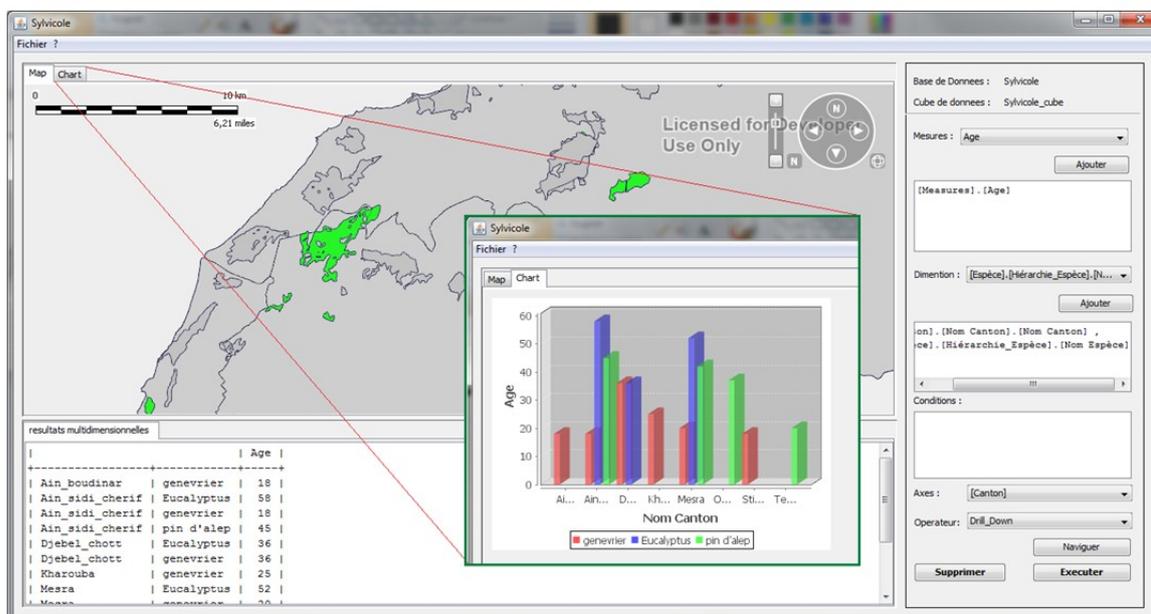


FIGURE 5.12 – Exemple de visualisation avec le client SOLAP "Sylvicole"

Afin de lancer une opération de "Régénération", une telle analyse permet d'examiner quelles sont les bosquets contenant les espèces les plus vieilles et ce pour chaque unité de gestion.

Dans l'exemple de la Figure 5.12, nous pouvons facilement voir que les arbres d' "Eucalyptus" dans les cantons "Ain-Sidi-Cherif" et "Mesra" sont les plus âgés, suivis par le "pin d'Alep" dans les deux cantons ce qui implique un besoin de "Régénération". D'autre part, l'absence de cette espèce est remarquable dans les cantons "Ain-Boudinar", "Kharouba" et "Stidia". Cela pourrait qualifier une «reforestation» de cette espèce comme une action candidate dans la troisième phase du processus d'aide à la décision.

De même, un deuxième sujet d'analyse peut être lancé avec la deuxième table de faits « Contrôle_des_événements » (*Control_event*). En effet, les analyses de mesures : « volume_financier », « zone » et « localisation » peuvent se faire dans différentes dimensions. Cela nous permet de savoir quel type d'événement est arrivé à un site, à quel moment et de quel type (feux de forêt, coupe de bois illégale, attaque d'insectes ou de parasites ... etc.). Les résultats de ces analyses proposent quelques opérations telles que : Reforestation, Ouverture de tranchée pare-feu, intensification des brigades de contrôle,... etc. selon le type de sol, la distance, l'élévation, le climat, etc. Ces opérations sont proposées par des experts qualifiés pour la troisième phase du processus d'aide à la décision.

La troisième table de faits « Travaux_silvicoles » (*silvicultural_work*) peut également être utilisée pour effectuer une analyse multidimensionnelle de ses mesures en fonction de

différentes dimensions thématiques, temporelles et spatiales. En effet, grâce à l'application SOLAP, nous pouvons par exemple, poser la question : *Quelles parties de la forêt et depuis quand se sont déroulées les opérations de «coupe de bois»?* , pour savoir s'il est nécessaire de lancer une telle opération.

Cette opération permet le renouvellement de la forêt d'une part, et permet aux riverains de la forêt de bénéficier des ressources naturelles d'autre part.

Les mesures de la quatrième table de faits « Gestion_de_la_faune » (*Wild_life*), peuvent être analysées dans différentes dimensions. Nous pouvons savoir si le nombre de certaines espèces faunistiques dans une région représente un danger pour les autres espèces végétales ou animales. Dans ce cas, pour conserver l'équilibre naturel, les experts peuvent proposer des opérations telles que «ouverture de chasse» ou «encourager des ennemis naturels».

Le cinquième sujet d'analyse peut être entamé par l'analyser des mesures de la dernière table de faits « Aires_protégées » (*protected_areas*). L'analyse multidimensionnelle de ces mesures dans différentes dimensions permet de contrôler les changements écologiques dans ces zones. Par exemple nous pouvons connaître l'évolution de la superficie (mesure numérique) s'il y a une expansion ou une dégradation annuelle. L'analyse spatiale de la mesure «localisation» permet d'examiner la présence de l'infrastructure routière ou de l'expansion urbaine à proximité des aires protégées. Cela peut être considéré comme une menace. Les experts peuvent offrir des «opérations de gabionage» ou «réalisation de clôtures» pour protéger ces zones.

5.8.4 Analyse Multicritères (ELECTRE III) pour classer les actions à exécuter

La liste des actions concernant l'AMC n'est pas connue à l'avance. En effet, elle ne sera fixée par les acteurs (experts) qu'après l'analyse SOLAP. En se basant sur les résultats des requêtes SOLAP, la liste initiale des actions est la suivante :

Ac1 : Construction de stations d'observation à des points élevés avec des coordonnées définies

Ac2 : Réalisation de points d'eau à des coordonnées définies

Ac3 : Contrôle des brigades de diffusion sur des sites bien situés

Ac4 : Lancement d'activités sylvicoles (entretien, assainissement ...).

Ac5 : Lancer un projet de reboisement sur des parcelles bien définies

Ac5 : Lancement du développement ou ouverture des pistes forestières

Ac7 : Lancement d'un travail de développement ou d'un pare-feu de tranchée

Ces actions sont évaluées en tenant compte de cette liste de critères (Ci, i = 1..5) :

C1 : Temps d'exécution

C2 : Coût

C3 : Effectif

C4 : Intérêt écologique

C5 : Génération de produits industriels

L'expert peut alors évaluer les actions selon différents critères. De plus, les Poids, les

seuils de préférence, les seuils d'indifférence et les seuils de Veto sont introduits comme indiqué dans la Table 5.2.

TABLE 5.2 – La matrice de Performances et paramètres de ELECTRE III

Actions	Critères				
	C1	C2	C3	C4	C5
Ac1	1	2	2	8	0
Ac2	2	3	3	8	0
Ac3	1	2	1	7	0
Ac4	7	5	5	7	6
Ac5	8	9	9	10	0
Ac6	8	8	9	7	7
Ac7	7	7	9	7	9
Poids	2	2	2	3	1
Sens de	↓	↓	↑	↑	↑
Indifférence	2	10	1	2	3
Préférence	4	20	2	2.8	6
Véto	8	40	4	5	9

En utilisant une application développée dans notre équipe de recherche, les résultats de (ELECTRE III) peuvent être illustrés en deux phases à savoir l'agrégation et l'exploitation :

5.8.5 Agrégation

La Figure 5.13 montre les résultats de l'agrégation. Elle présente les matrices de concordance globale, de discordances et de crédibilité.

5.8.6 Exploitation

La deuxième phase exploite les résultats pour le classement. La Figure 5.14 montre les résultats de l'exploitation. Elle présente le processus de distillation et le classement final.

Le classement final des actions à exécuter représente la dernière étape du processus "Silvicultura". Nous notons que la réalisation et le contrôle de cette solution représentent les phases suivantes. Cependant, cela dépasse les limites de la présente étude.

5.9 Conclusion

SOLAP combine l'analyse OLAP avec des visualisations cartographiques des systèmes SIG, permettant de comprendre les distributions géographiques des données à référence spatiale. Le système SOLAP étend OLAP avec des opérateurs spatiaux (par exemple, Spatial Roll-up) pour permettre des explorations spatiales multidimensionnelles interactives

The screenshot displays three matrices from the ELECTRE III aggregation results:

- Concordance Globale:** A matrix comparing 7 actions. Values range from 0.7000 to 1.0000.
- Matrices Discordances:** A matrix where all values are 0.0000.
- Matrice Crédibilité:** A matrix where all values are 0.0000.

FIGURE 5.13 – Résultats de l'agrégation de ELECTRE III

The screenshot shows the final ranking results from the ELECTRE III exploitation:

- Ascendante (Ranking from highest to lowest):**

RANG	ACTION
7	Action 3
5	Action 1
5	Action 2
4	Action 4
3	Action 6
2	Action 5
1	Action 7
- Descendante (Ranking from lowest to highest):**

RANG	ACTION
1	Action 5
3	Action 6
3	Action 7
4	Action 4
5	Action 2
7	Action 1
7	Action 3
- Classification Summary:**
 - Classement 1 ==> Action 5 ; Action 7 ;
 - Classement 2 ==> Action 6 ;
 - Classement 3 ==> Action 4 ;
 - Classement 4 ==> Action 2 ;
 - Classement 5 ==> Action 1 ;
 - Classement 6 ==> Action 3 ;

FIGURE 5.14 – Résultats de l'exploitation d'ELECTRE III

de données analytiques.

Cependant, la technologie SOLAP ne peut, à elle seule, résoudre une situation liée à des problèmes spatiaux. La solution, dans notre proposition, est de combiner cette technologie avec les outils d'analyse multicritères (AMC). En effet, l'AMC offre plusieurs avantages en termes d'aide à la prise de décision lorsque nous devons prendre en compte des intérêts

conflictuels.

L'objectif principal de cette contribution est de développer un Système d'aide à la décision spatiale multicritères à savoir "Silvicultura" en intégrant un moteur d'analyse multicritères et SOLAP. Cette intégration peut aider les décideurs à gérer efficacement les espaces agroforestiers et leurs activités.

"*Silvicultura*" est composé de quatre sous-systèmes, à savoir le sous-système de base de données, le sous-système de surveillance, le sous-système de traitement des problèmes et le sous-système d'interface utilisateur. Ces sous-systèmes sont interconnectés par des flux de données et de requêtes pour que la décision soit annoncée dans l'interface utilisateur. Notre modèle décisionnel est implémenté à l'aide d'une architecture OLAP relationnelle composée d'un système de gestion de base de données relationnelle (SGBD) qui peut implémenter des indicateurs complexes au moyen des expressions multidimensionnelles (MDX). En outre, l'outil SIG permet l'analyse spatiale et l'exploration de données spatiales.

Notre proposition est testée avec des données relatives aux espaces agroforestiers de la wilaya de *Mostaganem (Algérie)*, en tenant compte de la politique de gestion actuelle dans ce pays. Quant à l'analyse multicritères, nous avons utilisé des outils ELECTRE I et ELECTRE III, développés au sein du laboratoire *LIO*.

Comme tout outil, le système proposé présente quelques limites relatives notamment à la qualité de l'analyse SOLAP. En effet, la consistance de cette analyse dépend, en particulier, de la qualité des données entreposées, de l'agrégation et de la qualité des requêtes formulées par les utilisateurs de tel système.

Afin de résoudre ce problème, nous allons exposer, dans le chapitre suivant, notre deuxième contribution relative à l'implémentation du système *UIC-SOLAP*, permettant le contrôle des contraintes d'intégrité relatives aux utilisateurs du système SOLAP.

Chapitre 6

Un système "UIC-SOLAP" pour le contrôle d'utilisation dans SOLAP

Dans ce chapitre, notre deuxième contribution dans cette thèse est présentée. Nous visons la mise en œuvre de système "UIC-SOLAP", pour le contrôle de la qualité d'utilisation d'un système décisionnel, notamment « Silvicultura » présenté au chapitre précédent. Un jeu de données relatives aux espaces agroforestiers de la wilaya de Mostaganem en Algérie, est utilisé pour tester la démarche de ce système.

Sommaire

6.1	Introduction	108
6.2	Nouvelle CI pour les faits dans le schéma en constellation	109
6.3	Contraintes d'intégrité pour les utilisateurs d'un SOLAP	109
6.3.1	CI-Utilisateurs liés aux données spatiales	110
6.3.2	CI-Utilisateurs liées à la politique d'exploitation du SOLAP	111
6.3.3	CI-Utilisateurs liées aux requêtes SOLAP	111
6.3.4	CI-Utilisateurs liées aux agrégations	112
6.4	Le système "UIC-SOLAP" pour les contraintes d'intégrité utilisateur dans SOLAP	113
6.5	Le processus adopté par "UIC-SOLAP"	113
6.6	Etude de Cas	115
6.7	Discussion des Résultats	117
6.8	Conclusion	119

6.1 Introduction

Les entrepôts de données spatiales (EDS) et les systèmes SOLAP sont utilisés pour l'exploration et l'analyse des données dans les systèmes d'information décisionnels spatiaux (SDSS). Dans la gestion des forêts, le processus de prise de décision nécessite l'utilisation d'un modèle multidimensionnel incluant plusieurs tables de faits dans le schéma en constellation.

Comme indiqué au chapitre 4, la qualité d'analyse SOLAP dépend de trois aspects qui concernent les données, l'agrégation et les requêtes formulées. Afin d'assurer cette qualité, les contraintes d'intégrité (CI) ont été reconnues comme une excellente approche. En effet, ils définissent les conditions qui doivent être remplies pour améliorer l'exactitude, la cohérence et l'exhaustivité des bases de données. Les CI peuvent être définies dans des modèles conceptuels pour permettre de traiter d'éventuels problèmes aux premiers stades du développement.

6.2 Nouvelle CI pour les faits dans le schéma en constellation

Selon (Salehi, 2009), une hyper-cellule est constituée d'une paire (L, MS) où L est un ensemble fini de niveaux de dimension et MS est un ensemble fini de mesures d'un hypercube. L'auteur classe les CI de faits en trois catégories. En effet, un fait peut être défini pour un seul fait d'une hyper-cellule (f-Inter0), pour plusieurs faits de la même hyper-cellule (f-Inter1), et pour plusieurs faits de plusieurs hyper-cellules (f-Inter2) comme le montre la Figure 6.1.

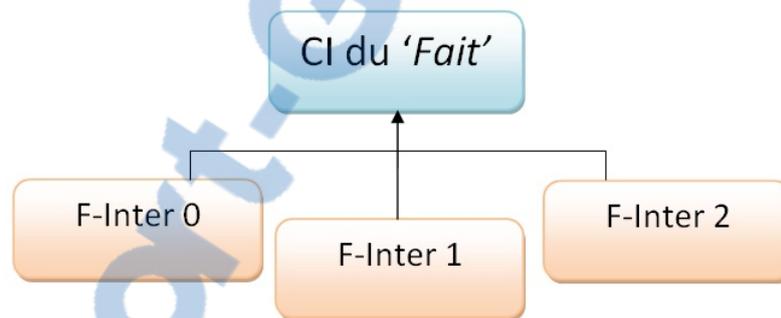


FIGURE 6.1 – Les catégories des CI de 'Fact'

Dans cette classification, l'auteur considère les CI des faits (*Fact IC*) d'une ou de plusieurs hyper-cellules correspondant au même table de faits.

Dans notre proposition, nous considérons de nouvelles CI pour des mesures dans plusieurs tables de faits au sein du même hypercube comme le montre la Figure 6.2. Cette situation peut se produire lorsqu'on considère un schéma en constellation d'un hypercube avec plusieurs tables de faits. Dans notre étude de cas, ces tables de faits sont liées à des sujets d'analyse en gestion forestière.

6.3 Contraintes d'intégrité pour les utilisateurs d'un SOLAP

Les CI liées aux utilisateurs du système SOLAP sont nombreuses. Dans notre proposition, nous utilisons un hypercube avec un schéma de constellation avec plusieurs utilisateurs, nous classifions ces CI en quatre catégories comme le montre la Figure 6.3.

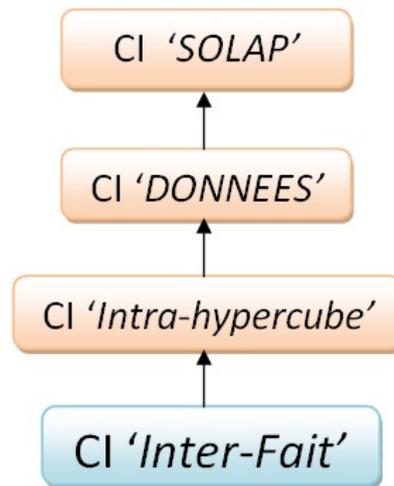


FIGURE 6.2 – CI inter-fact dans un hypercube.

Cette classification peut ne pas être exhaustive mais nous nous sommes concentrés sur ces parties que nous considérons directement liées à l'utilisateur d'un tel système.

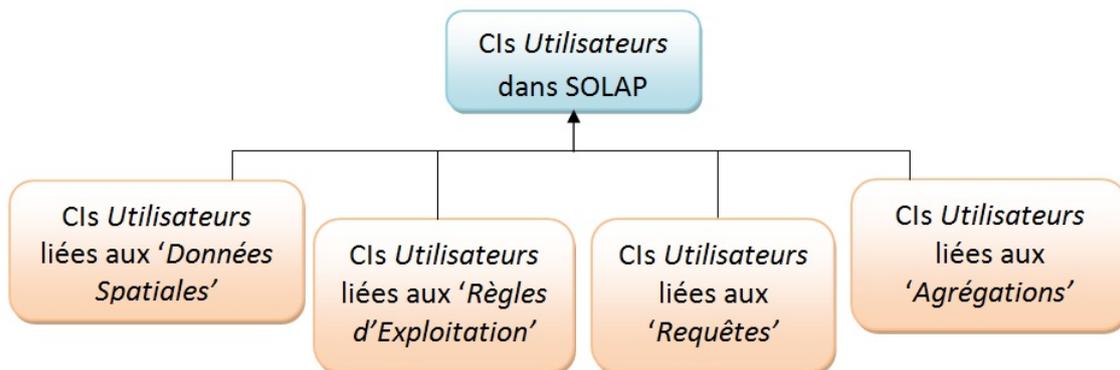


FIGURE 6.3 – CI-utilisateurs dans un SOLAP.

6.3.1 CI-Utilisateurs liés aux données spatiales

Les contraintes d'intégrité s'appliquent à la base de données, un ensemble d'états qui sont valides en vertu des propriétés des objets qui doivent être stockés en fonction de chaque type d'utilisateur final du système SOLAP.

Les propriétés géométriques des données spatiales sont liées aux relations topologiques, métriques et d'ordre. Ici, les CI sémantiques concernent la signification des entités géographiques. L'utilisateur peut définir des contraintes d'intégrité qui permettent de maintenir la cohérence de la base de données en fonction de ses besoins (Cockcroft, 1997).

Après la spécification des contraintes d'intégrité, nous devons vérifier si elles sont cohérentes. En effet, avant de vérifier si une base de données satisfait un ensemble de contraintes sémantiques spatiales, il faut s'assurer que les contraintes elles-mêmes ne doivent pas être en conflit.

Par exemple, une règle d'utilisateur doit être activée pour localiser la station d'essence à une distance donnée par rapport à la forêt. Cela ne doit pas être en contradiction avec une autre CI défini par l'utilisateur.

6.3.2 CI-Utilisateurs liées à la politique d'exploitation du SOLAP

Lors de la modélisation, le profil UML peut être enrichi avec un stéréotype appelé «*UserProfile*», qui permet de spécifier des contraintes en fonction d'informations particulières d'un utilisateur ou d'un groupe d'utilisateurs (Villarroel et al., 2006). Par conséquent, nous personnalisons les informations en fonction des caractéristiques de l'utilisateur qui demande ces informations.

La Figure 6.4 illustre des CI-utilisateur pour un exemple de modélisation multidimensionnelle sécurisée adaptée de (Villarroel et al, 2006). Il montre une partie du modèle sécurisé, appelé «Hôpital», qui est basé sur un système de soins médicaux. Dans cet exemple, les auteurs notent les considérations suivantes relatives à l'utilisateur du système :

- Les patients peuvent accéder à leurs propres informations en tant que "patients".
- refuser l'accès aux données des patients ayant été traités avant la date du contrat initial.
- refuser l'accès aux informations d'admission aux utilisateurs dont la zone de travail est différente de la zone d'une instance d'admission particulière.

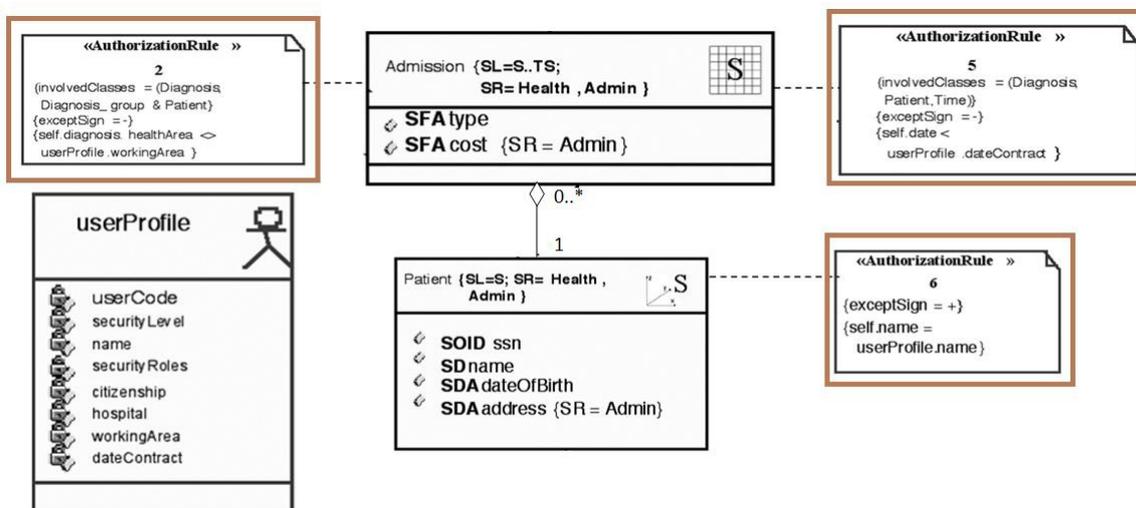


FIGURE 6.4 – Exemple de CI-Utilisateur.

6.3.3 CI-Utilisateurs liées aux requêtes SOLAP

Une requête SOLAP est une combinaison de mesures et de membre de différentes dimensions. Une CI vérifie la validité de cette combinaison qui peut être effectuée par n'importe quel utilisateur.

Par exemple, il est absurde de demander la répartition géographique des sylvicultures effectuée par la «conservation» de l'administration forestière «*Mostaganem*» en 1985 dans

la forêt de «*Zemoura*». Ceci par ce que cette forêt n'est plus rattachée à "*Mostaganem*" après 1984.

Ici, selon le modèle multidimensionnel représenté avec un schéma de constellation proposé dans la Figure 5.9, nous avons une requête analytique qui combine :

- Table de faits : ["sylviculture"].
- Measure Spatial : ["localisation"].
- Niveau d'agrégation spatiale : [forest = "Zemoura"].
- Niveau d'agrégation spatiale : [conservation = "Mostaganem"].
- dimension Time : [année = "1985"].

Le résultat démontrera qu'il n'y a pas d'activités sylvicoles; tandis que, selon la division administrative après février 1984, la forêt "*Zemoura*" est devenue sous la direction d'une autre conservation à savoir "*Relizane*" (ouest de *l'Algérie*). Cela peut produire une décision peu fiable. Ainsi, une contrainte d'intégrité de requête est capable d'interdire cette combinaison de paramètres.

6.3.4 CI-Utilisateurs liées aux agrégations

Bouilil (Bouilil, 2012) a implémenté des CI d'agrégation dans le profil UML pour éviter l'implémentation de modèles de l'EDS incorrects sémantiquement et structurellement.

Par exemple, pour forcer l'utilisateur à ne pas agréger des mesures non additives en utilisant la fonction d'agrégation «sum», l'auteur définit l'instruction OCL qui interdit cela comme indiqué sur la Figure 6.5.

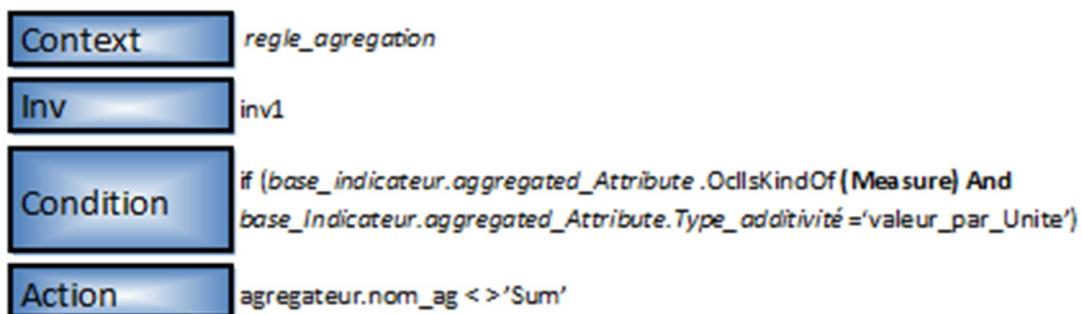


FIGURE 6.5 – Exemple de CI relative aux agrégations.

Généralement, ce type de règles d'agrégation est incontournable pour chaque utilisateur. Cependant, tout utilisateur doit utiliser un nombre limite de mesures liées à une table de faits spécifique avec un groupe de niveaux de dimension. Par conséquent, nous pouvons attacher les règles d'agrégation au stéréotype "*UserProfile*" défini dans la section Étude de cas.

En fait, il révèle que les CI peuvent être spécifiées en utilisant OCL et Spatial OCL au niveau conceptuel. Cependant, cette solution peut être dans certains cas insuffisante. Cela concerne en particulier les CI des métadonnées, les CI d'agrégation et les CI de visualisation.

Dans ce travail de thèse, nous proposons que la solution puisse être au stade de modélisation, en ajoutant dans le profil UML : (Abdallah Bensalloua, C. et Hamdadou, D. 2018b)

1. plus de valeurs étiquetées (tagged values) représentant les métadonnées principales,
2. plus de stéréotypes permettant aux utilisateurs de définir leurs propres CI d'agrégation,
3. une formalisation des profils des utilisateurs pour spécifier les CI de visualisation en fonction de leurs préférences Dans les sections suivantes, nous allons expliquer notre solution en détail.

6.4 Le système "UIC-SOLAP" pour les contraintes d'intégrité utilisateur dans SOLAP

Dans la présente étude, nous proposons un système appelé "UIC-SOLAP" (Abdallah Bensalloua et Hamdadou 2018b) pour vérifier les contraintes d'intégrité qui contrôlent les interactions entre l'utilisateur et le système SOLAP, comme le montre la Figure 6.6 (Abdallah Bensalloua et Hamdadou 2018b) .

Ce système intègre toutes les contraintes d'intégrité définies précédemment et qui sont liées aux utilisateurs du système SOLAP dans les différentes étapes de ce dernier.

En effet, dans les sources de données qui peuvent être à référence spatiale, les propriétés des objets peuvent être stockées en fonction de chaque type d'utilisateur final.

Dans l'interaction avec SOLAP Server, l'utilisateur utilise une combinaison de mesures et de membres dans des requêtes analytiques ; et utilise aussi des règles d'agrégation. Ceux-ci doivent respecter les contraintes d'intégrité.

En outre, lors de l'utilisation de l'interface utilisateur d'une application SOLAP, les contraintes d'intégrité doivent être définies pour respecter les informations particulières d'un utilisateur, de sorte que seules les données concernées soient accessibles.

6.5 Le processus adopté par "UIC-SOLAP"

Tout au long de construction et l'utilisation d'un SOLAP, le processus adopté par le système "UIC-SOLAP", peut être décrit comme suit : (Abdallah Bensalloua et Hamdadou 2018b)

- Premièrement, les utilisateurs peuvent faire des assertions différentes dans la base de données spatiale en fonction de leurs besoins. Dans cette tâche, nous devons veiller à ce que ces assertions ne soient pas en conflit entre elles ;
- Deuxièmement, pour identifier les requêtes analytiques incorrectes, nous devons vérifier la validité de la combinaison des mesures et des membres ;
- Ensuite, pour définir les règles d'agrégation correctes, nous devons définir dans la phase de modélisation, des stéréotypes permettant aux utilisateurs de définir leurs propres CI d'agrégation ;

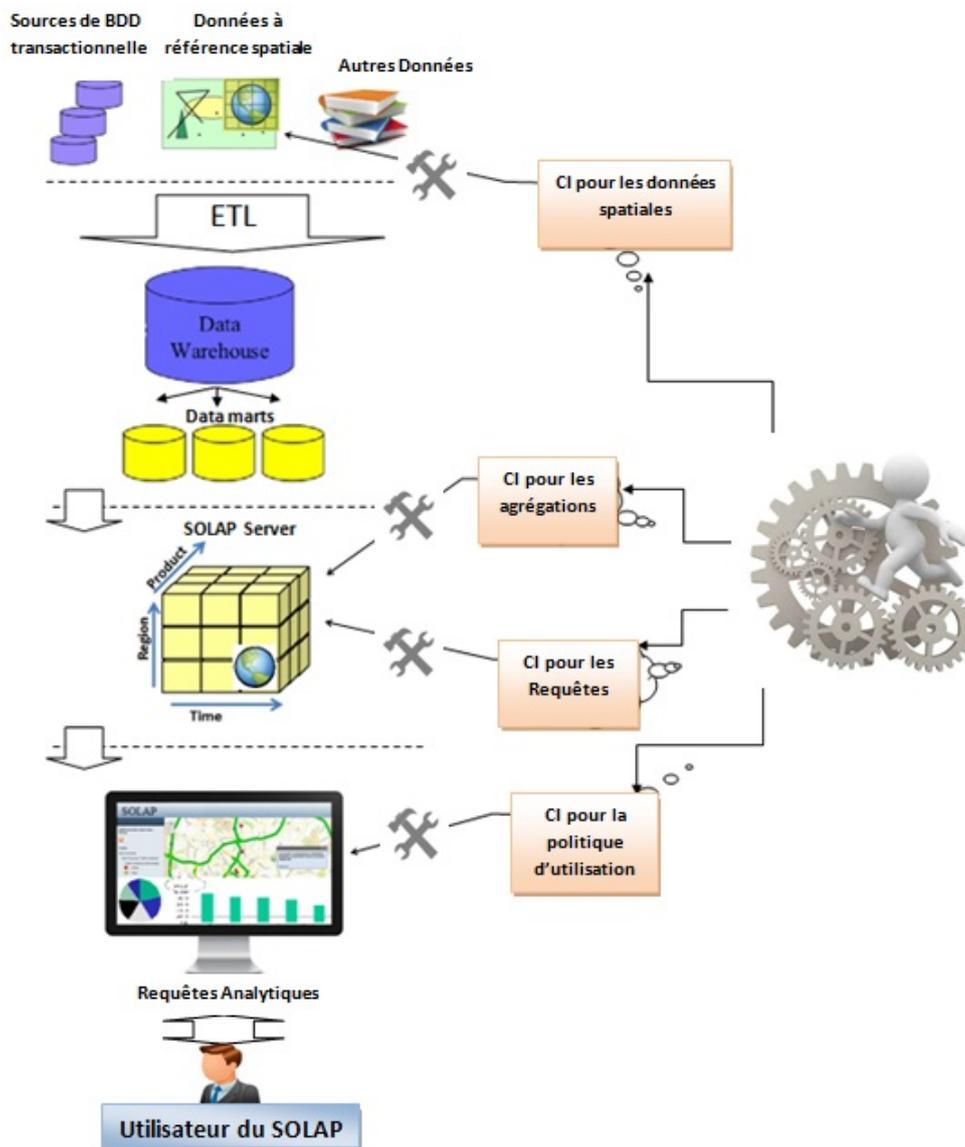


FIGURE 6.6 – “UIC-SOLAP” Contraintes d’intégrité Utilisateurs de SOLAP.

- Et enfin, la formalisation du concept *user_profile* est assurée permettant de spécifier des contraintes en fonction des informations particulières de chaque type d'utilisateurs.

L’algorithme de fonctionnement pour appliquer "UIC_SOLAP" peut être comme suit :

Algorithm

{ check data user_integrity_constraints when analyzing data, before applying “ETL” process }

S ← ∅ { set of spatial data properties }

P ← ∅ { final list of properties initially empty }

For(every kind of final_user of SOLAP)

Begin

S ← Read (database_properties); { user will enter spatial properties according to him }

```

Verify_contradictory (P,S); {to verify if entered properties are contradictory with the others}
While (not ok) {While S is in contradictory with at least one of database_properties}
  Begin
  redefine_properties(S) {S will be redefined}
  Verify_contradictory (P,S);
  End_while
  P ← P+S; {add S to the final list of properties}
end_for;
{checking exploitation policy user integrity constraints When connecting to SOLAP}
Read (user_id)
Apply_profile (user_id) {applying a set of parameters allowing user to see only information related to his specialty}
{ ... When querying the SOLAP}
While (user is connected)
  Begin
  {checking querying user integrity constraints}
  Read(query)
  Verify_combin({dimension_members}, measure)
  If (not OK) then show_message("inconsistent query will generate inconsistent results");
  {checking querying user integrity constraints}
  Read (aggregation)
  Verify_aggregation(query,aggregation_model)
  {to Verify if requested aggregation function is allowed to be applied}
  If (not OK) then show_message("not allowed aggregation will generate inconsistent results")
  End_while
End_Algorithm.

```

Le diagramme d'activité du système "UIC-SOLAP" proposé, est représenté sur la Figure 6.7

6.6 Etude de Cas

L'étude de cas utilisé pour illustrer notre proposition concerne l'analyse multidimensionnelle (MD) dans un Système d'Aide à la Décision Spatiale (SDSS) pour la gestion forestière.

Nous utilisons une extension de UML pour la modélisation MD avec trois mécanismes : les stéréotypes, les valeurs étiquetées (*tagged values*) et les contraintes dans le profil UML (Juan et al., 2009). Nous utilisons également le langage OCL (Object Constraint Language) pour formaliser les contraintes d'intégrité qui affinent les définitions des stéréotypes et des valeurs étiquetées (Boulil et al., 2014).

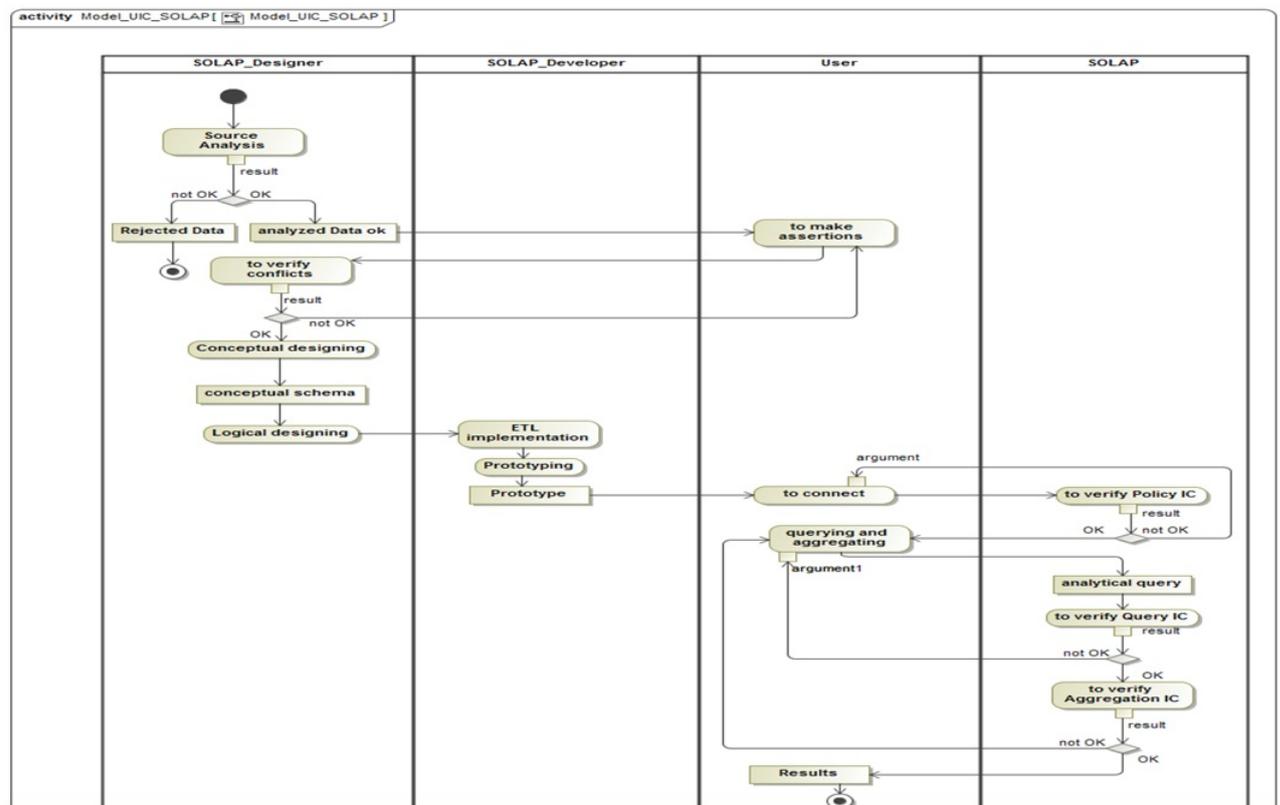


FIGURE 6.7 – Diagramme d’activité pour "UIC-SOLAP" .

Nous avons utilisé le profil-UML de l’EDS des forêts décrit au chapitre précédent, dans la Figure 5.6. Nous utilisons également l’hypercube de données décrit au chapitre précédent, dans la Figure 5.9.

Néanmoins, nous allons nous intéresser à l’analyse de trois aspects uniquement, correspondant aux trois tables de faits partageant certaines tables de dimension qui sont expliquées au chapitre précédent et qui sont :

1. Table de faits (travaux_sylvicoles) « *Silvicultural_act* » ;
2. Table de faits (événement) « *Control_event* » ;
3. Table de faits (Dendrométrie) « *Control_dendrometry* ».

Les dimensions considérées sont :

1. Une Temporelle à trois niveaux : (mois, trimestre ou année) ;
2. Deux dimensions Spatiales : administrative (district, circonscription et conservation) ; et écologique (canton et forêt) ;
3. Quatre dimensions thématiques : *Espèce*, *Type_activité*, *Cause* et *Type_event*.

Par ailleurs, nous définissons le stéréotype UserProfile pour spécifier la CI en fonction des informations particulières de l’utilisateur. Dans le diagramme de classes, nous avons une classe de ce stéréotype qui n’a aucune association avec d’autres classes. Il est utilisé uniquement pour la formulation de la CI comme indiqué sur la Figure 6.8. (Abdallah Bensalloua et Hamdadou 2018b)



FIGURE 6.8 – Stereotype et Class UML pour UserProfile.

Pour le nouveau type de CI inter-faits dans l’hypercube cité dans la section 4.1, dans le schéma multidimensionnel de l’EDS des forêts, nous introduisons l’assertion qui vérifie les conditions prospectives entre deux tables de faits ou plus.

La Figure 6.9 montre une CI inter-faits entre deux tables de faits "sylviculturals" et "events" du même hypercube. Elle précise que les superficies de reboisement doivent être supérieures aux superficies détruites par les incendies.

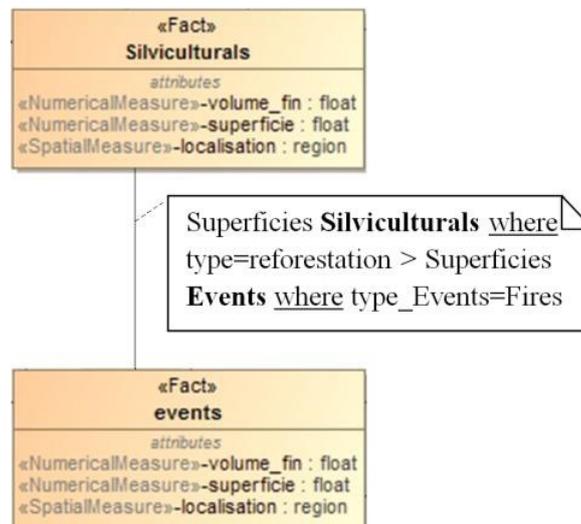


FIGURE 6.9 – Contraintes d’intégrité Inter-facts.

6.7 Discussion des Résultats

Nous avons développé une application SOLAP *Forest* qui intègre le système de vérification des contraintes d’intégrité-Utilisateur.

En utilisant un jeu de données concernant les forêts de la wilaya de Mostaganem, différents utilisateurs peuvent transmettre leurs requêtes analytiques à l’aide de l’interface interactive. Avec leurs différents aspects d’intérêt, ils peuvent manipuler le système de différentes façons. Cela peut générer des résultats analytiques incohérents. Par conséquent, la formulation de CI à différentes étapes peut éviter les problèmes d’intégrité.

Dans les exemples suivants, nous allons nous concentrer sur certains CI liées aux utilisateurs de l’application SOLAP.

Premièrement, pour les CI utilisateurs liées aux données spatiales, nous prenons en compte les CI «définis par l'utilisateur». Ce sont des règles métier qui contrôlent l'intégrité des données stockées et des résultats. Ici nous avons trois exemples liés aux utilisateurs de chacune des tables de faits de l'hypercube :

1. "La distance entre une *forêt* et une *station-service* doit être supérieure à 500 mètres".
2. "Le boisement peut être effectué dans les espaces forestiers nus".
3. "il est autorisé de couper un arbre quand sa circonférence = 4m ; hauteur > = 20m ; âge > = 50 et il est proche d'une route ".

Ces CI peuvent être contradictoires car les utilisateurs sont hétérogènes et les données seront utilisées dans des tables de faits différentes dans le même hypercube.

Dans notre proposition, nous essayons d'éliminer ces problèmes contradictoires. Puisque notre prototype peut être utilisé par n'importe quel type d'utilisateur, il peut interroger n'importe quelle table parmi les tables de faits. Les intérêts des utilisateurs peuvent être différents ; par conséquent, un utilisateur peut ne pas accepter le boisement à proximité d'une station-service, même s'il s'agit d'un espace forestier nu. Cela peut être dangereux pour d'autres considérations.

Pour les CI-utilisateurs liées à la politique d'utilisation, en utilisant le concept "*User-Profile*", une "règle utilisateur" ne sera activée que si l'utilisateur concerné se connecte. Nous créons trois profils : *Event_role*, *Silvicultural_role* et *dendrometry_role* liés aux utilisateurs des tables de faits *Events*, *Silviculturals* et *dendrometry* consécutives.

Dans ce cas, si le rôle est "*Event_role*" alors seul le message de CI correspondant apparaît "La station de service est à moins de 500 mètres de la forêt" ; malgré "l'existence de très vieux arbres près de la route à couper" et "l'existence d'espaces forestier nu à reboiser" voir Figure 6.10.

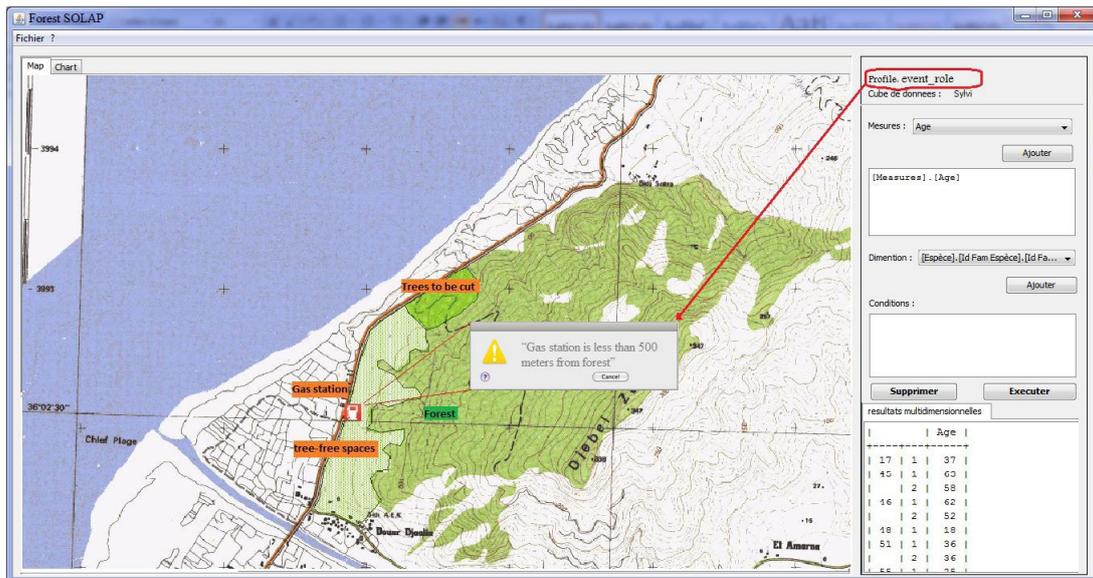


FIGURE 6.10 – Notification de violation d'une CI de Requête.

Nous notons que le problème décisionnel peut être résolu par l'application d'un système

d'aide à la décision collective. Cependant, cela va au-delà de l'objectif de la présente étude.

Pour les CI-utilisateurs liées aux requêtes, compte tenu de la nouvelle division administrative, le système que nous proposons vérifie la justesse des combinaisons de mesures et de membres de dimension dans les requêtes utilisateur.

Concernant l'exemple ci-dessus concernant la combinaison de mesures du table de faits *Silviculturals* et de la dimension spatiale écologique "forêt = *Zemoura*" avec dimension spatiale administrative "conservation = Mostaganem" est faux lorsque le membre dimension temporelle est > février 1984. Ainsi, le message de la Figure 6.11 apparaît.

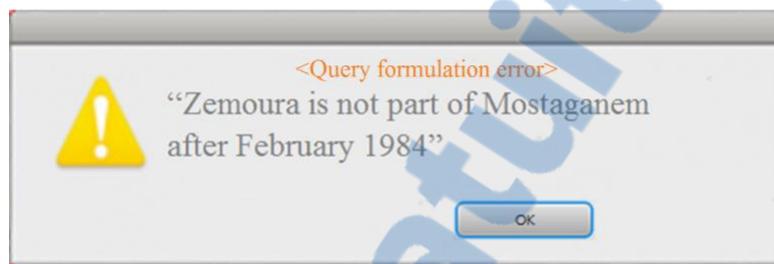


FIGURE 6.11 – Notification de violation d'une CI de Requête.

Et enfin, les CI utilisateurs liées à l'agrégation définissent comment l'utilisateur veut agréger des mesures pour calculer différents indicateurs. Ici, nous indiquons que chaque utilisateur dispose d'un ensemble d'indicateurs pour l'analyse multidimensionnelle liés à sa spécialité. Ainsi, dans le modèle d'agrégation, pour chaque indicateur nous devons spécifier quelle fonction doit être utilisée afin d'agréger la mesure.

Dans ce cas, puisque la CI est liée à l'indicateur et que ce dernier est associé à l'utilisateur, nous pouvons donc définir une CI associant la spécialité de '*UserProfile*' à la table de faits concernée. En effet, cette association peut enrichir la cohérence de l'analyse, puisque chaque utilisateur est en mesure d'effectuer des requêtes correctes relatives à sa spécialité.

Nous notons que l'interrogation d'un entrepôt de données est la tâche de l'analyste (l'homme d'étude) ou du décideur. Cependant, dans la présente étude, nous nous concentrons sur le cas où nous utilisons un hypercube avec plusieurs tables de faits. Par exemple, dans les systèmes de décision de groupe, ils peuvent exister différents décideurs. Chacun d'eux interroge les faits de son intérêt et ignore comment utiliser d'autres mesures. Cela semble être facultatif, mais il peut être considéré comme un contrôle plus à l'avance pour l'agrégation des mesures. Par exemple, en ce qui concerne la table de faits dendrométrie, seuls les utilisateurs pour lesquels *Userprofile.specialty* = "*dendrometry*" pourront agréger ses mesures.

6.8 Conclusion

L'objectif principal de la présente contribution détaillée dans ce chapitre est de mettre en évidence les CI essentiellement liées aux utilisateurs de la solution SOLAP pour la gestion efficace des zones forestières. En effet, nous proposons une modélisation de l'EDS de la forêt grâce au profil UML. Ensuite, nous l'utilisons pour la modélisation du schéma en constellation avec trois tables de faits partageant des dimensions spatiales, temporelles

et thématiques. Ensuite, nous proposons une classification des CI liées aux utilisateurs des systèmes SOLAP. Nous proposons, également, un nouveau type de CI inter-faits liées à l'utilisation de l'hypercube avec plusieurs tables de faits. Enfin, les résultats de notre proposition sont montrés en utilisant des données concernant les zones forestières de *Mos-taganem* en *Algérie*.

Conclusion Générale et Perspectives

« *Qui voit de haut voit bien, qui voit de loin voit juste.* »

Victor Hugo

Les forêts représentent des ressources naturelles très chères. Leur importance est remarquable pour l'équilibre des écosystèmes écologiques et même pour la stabilisation des caractéristiques climatiques, économiques et sociales.

Les espaces agroforestiers, à travers le monde, et en particulier en Algérie, se détériorent de jour en jour. Cette dégradation est due aux différents facteurs naturels interprétés par le changement climatique et la désertification d'une part ; et d'une autre part, l'influence humaine est devenue l'ennemi fondamental qui menace l'existence future de ces ressources vitales.

En outre, les riverains de la forêt doivent collaborer à la protection et la préservation des ressources forestières qui garantissent leur survie.

La forêt algérienne est en général, caractérisée par un milieu naturel fragile et perturbé qui nécessite une gestion intégrée tenant en compte toutes les exigences écologiques, économiques et sociales.

Dans cette thèse, nous avons abordé la problématique d'aide à la décision dans le domaine de la gestion agroforestière. En effet, le système décisionnel est un moyen qui permet de faciliter la mise en œuvre de la stratégie de préservation des ressources naturelles. Ceci est garanti en :

- bordant l'analyse des activités déjà réalisées selon différents axes de réflexion pour en tirer des préventions concernant les activités futures,
- Utilisant des données de différentes époques pour les entités spatiales sujettes de l'analyse.

Il s'agit donc de permettre aux différents intervenants du domaine de bien comprendre l'environnement, d'ajuster les interactions entre eux, et surtout d'exprimer les meilleurs choix des exigences et d'actions nécessaires.

Un système décisionnel permet la bonne gestion des prévisions, de planification et du suivi. En plus les interactions entre les intervenants permettent le partage des différentes techniques d'apprentissage relatives à l'acquisition des connaissances, de compétences et du savoir faire.



Un système d'information décisionnel (*SID*) est basé généralement sur la modélisation multidimensionnelle inventé par Ralph Kimball (Kimball, 1996). Elle est dédiée à l'analyse, en vue d'aider à la décision, d'un processus métier notamment la gestion agroforestière. Dans un modèle multidimensionnel, les données sont représentées en fonction des besoins d'analyse dans un entrepôt de données (ED). En effet, les activités réalisées sont décrites par des faits contenant des *mesures* qui représentent les indicateurs de gestion. Ces mesures sont analysées selon différentes *dimensions* thématiques et temporelles, grâce aux outils d'analyse en ligne '*On-Line Analytical Processing*' (OLAP).

L'introduction de la dimension spatiale dans le processus d'analyse permet d'aider les décideurs à résoudre des problèmes spatiaux complexes. Ceci en utilisant les entrepôts de données spatiales (EDS) et spatial-OLAP (SOLAP). Ces derniers permettent l'intégration des capacités de modélisation spatiale, la gestion des données à référence spatiales et la représentation cartographique des données et des résultats d'analyses spatiales.

Dans le cadre de la gestion agroforestière, l'analyse doit concerner les différentes activités liées aux espaces forestiers à travers les différents axes de réflexion à savoir :

1. la dimension spatiale représentant la répartition géographique écologique et administrative ;
2. la dimension temporelles représentant les différentes époques ; et
3. un ensemble de dimensions thématiques, représentant les différents points de vue relatives à la gestion agroforestière.

En présence de certains paramètres complexes lorsque la décision concerne des espaces agroforestiers, il y a l'intervention de différents acteurs qui affectent le processus d'aide à la décision. Par conséquent, l'utilisation de l'outil SOLAP seul ne permet pas la prise en compte au même temps, des considérations techniques, écologiques, économiques et sociales.

Dans ce cas, il est très bénéfique le recours à l'analyse multicritères (AMC), qui désigne l'ensemble d'outils développés pour le but de résoudre des problèmes décisionnels, en présence de plusieurs critères conflictuels.

Dans la présente thèse, la solution proposée est l'intégration des outils SOLAP-AMC pour gérer efficacement les zones agroforestières. Ceci permet de tirer profit de l'analyse spatiale des SIG et de l'analyse multidimensionnelle offerte par le SOLAP d'un côté ; et de l'AMC qui permet la résolution des problèmes complexes avec différentes exigences.

A cet effet, nous avons tout d'abord étudié les systèmes d'information décisionnels, les entrepôts de données et les outils OLAP. Notre intérêt s'est porté sur les différents niveaux de l'architecture d'un système d'information décisionnel basé sur les entrepôts de données (ED). Le modèle multidimensionnel ainsi que les différentes modélisations conceptuelle, logique et physique ont été présenté.

Nous avons abordé ensuite les systèmes d'aide à la décision spatial (*Spatial Decision Support Systems* : (SDSS)) qui englobent des composants de DSS et des systèmes d'information géographique (SIG). Pour ce fait, les concepts de base de ces derniers sont aussi présentés. Par la suite, nous avons mis l'accent sur les avantages des systèmes SOLAP qui permettent l'intégration SIG-OLAP bénéficiant des atouts des deux solutions. La mo-

délisation spatio-multidimensionnelle est entamée pour présenter les différentes approches notamment (UML, Entité-Association et les modèles Ad-hoc).

La méthodologie multicritères d'aide à la décision est ensuite étudiée. Dans ce contexte, les concepts de base ainsi que les types de problèmes décisionnels sont présentés. Pour la méthodologie d'AMC, nous nous sommes concentrés sur les méthodes procédant par agrégation partielle, notamment les méthodes *ELECTRE* et *PROMETHEE* en expliquant les différentes démarches d'utilisation.

La deuxième partie de ce document correspond à nos contributions. En effet, nous avons présenté *Silvicultura*, le système d'aide à la décision proposé pour la gestion des espaces agroforestiers, développé dans cette thèse, en se basant sur une approche intégrée SOLAP-AMC.

Silvicultura est composé de quatre sous-systèmes interconnectés par des flux de données et de requêtes, notamment :

1. le sous-système de base de données
2. le sous-système de surveillance
3. le sous-système de traitement des problèmes et
4. le sous-système d'interface utilisateur.

Nous avons basé notre proposition de la modélisation multidimensionnelle sur UML qui peut être étendu, pour modéliser l'entrepôt de données spatiales relatives à l'agroforesterie, grâce au profil-UML

Pour tester le système «*Silvicultura* », l'implémentation de la solution SOLAP «*Silvicole* » a été réalisée par le langage de programmation (*Java Eclipse*) à l'aide d'une architecture ROLAP grâce au système de gestion de base de données (*MS-SQL Server*). En outre, l'analyse spatiale et l'exploration de données spatiales sont effectuées dans (*ARCGIS 10.3*).

Nous avons utilisé par ailleurs, un ensemble d'outils pour l'AMC notamment *ELECTRE I* et *ELECTRE III*, qui sont développés au sein du laboratoire *LIO*.

Nous avons testé notre proposition à l'aide d'un jeu de données concernant les espaces forestiers de la wilaya de *Mostaganem (Algérie)*. Nous avons sollicité l'expérience des experts et des responsables forestiers dans différentes phases de cette expérimentation pour le but de tenir compte de la politique de gestion forestière actuelle du pays.

Notre deuxième contribution dans cette thèse, consiste au développement d'un système de contrôle des contraintes d'intégrité spécifiques aux utilisateurs du SOLAP notamment «*UIC-SOLAP* ». Pour tester l'efficacité de ce système nous avons développé l'application «*Forests* » permettant la vérification de l'ensemble de ces CI.

Dans cette thèse, nous avons eu l'occasion d'étudier et d'intégrer les techniques d'analyse en ligne (OLAP), d'analyse spatiale (SIG) et celles de l'analyse multicritères (AMC). L'objectif de cette intégration est le développement d'un système d'aide à la décision dans la gestion agroforestière. Nous avons adopté pour cela, une démarche qui tire profit des travaux réalisés dans ce domaine connu pas la multitude des efforts des auteurs en matière de techniques d'analyse et de développement de solutions. En outre, la multitude des intervenants et la complexité des problématiques liées à la décision sur le territoire, rendent

le rôle de l'homme d'étude et de l'expert, incontournable dans le processus décisionnel.

Nos travaux futurs concernent l'extension de *Silvicultura* en incluant un système de base de connaissances spatiales. Ce sera un mécanisme intelligent avec la capacité de lier l'entrepôt de données spatiales avec le sous-système de surveillance, d'une part. D'autre part, dans le sous-système de base de données, l'entrepôt de données spatiales sera amélioré en enrichissant et optimisant le profil UML par l'introduction de nouvelles contraintes d'intégrité spatiale et des mécanismes de contrôle de qualité des données dans les hypercubes hétérogènes. En outre, nous pensons à l'exploitation d'autres méthodes de l'AMC qui peuvent améliorer la résolution d'autres situations de conflit spatial dans le processus décisionnel.

Nous pensons, également, à la mise en œuvre de systèmes décisionnels plus avancés gérant des données fraîches. Ces données sont mises à jour en quasi temps réel pour le but de servir un centre de gestion des urgences (catastrophes naturels), ou bien pour les serveurs web.

Dans ce cas, les entrepôts de données sont alimentés fréquemment. Par conséquent, se poseront des problèmes d'optimisation de la gestion et de la recherche de l'information au sein de ces entrepôts de données.

Nous visons aussi, à l'extension du système *sylvicultura* pour le cas de plusieurs décideurs en optant pour une modélisation basée agents.

Quant au système de contrôle de qualité d'analyse SOLAP, nous visons à traiter d'autres cas où les contraintes d'intégrité concernant les utilisateurs SOLAP peuvent être posées. Cela peut être atteint en enrichissant notre classification présentée dans cette thèse. Nous aborderons également d'autres aspects liés à la qualité d'analyse dans les entrepôts de données spatiales et à la vérification de la cohérence logique dans les systèmes SOLAP.

Publications issues de la présente thèse

- C. Abdallah Bensalloua and D. Hamdadou. Spatial OLAP and Multicriteria Integrated Approach for Decision Support System : Application in Agroforestry Management (030117-074730), International Journal of Decision Support System Technology (IJDSST) 10(3), 2018
- C. Abdallah Bensalloua and D. Hamdadou. Users Integrity Constraints in SOLAP Systems- Application in Agroforestry, International Journal of Interactive Multimedia and Artificial Intelligence. 5(1) :47-56, 2018
- C. Abdallah Bensalloua, D. Hamdadou and B. Beldjillali. SOLAP for Spatial Multidimensional Analysis in Decision-Support Purpose - Case study : Natural Hazards effect on forest ecosystems managing. Journal of Network and Innovative Computing. 3 :128-137, 2015
- C. Abdallah Bensalloua, D. Hamdadou and B. Beldjillali. Sharing and Security of Spatially Referenced Data in SOLAP : An Application to Agro-Forestry Management. Journal of Information Assurance and Security. 9 :306-315, 2014
- C. Abdallah Bensalloua, D. Hamdadou and B. Beldjillali. SOLAP : A type of user interface and Security Oriented System for Agro-forestry management”. Information Assurance and Security Letters. 3 :008-013, 2012
- C. Abdallah Bensalloua, D. Hamdadou and B. Beldjillali. Une solution SOLAP pour la planification des activités de pêche- Cas de la wilaya de Mostaganem. 2nd International Conference on Software Engineering and New Technologies (ICSENT 2013) Hammamet, Tunisia. December 21—23, 2013
- C. Abdallah Bensalloua, D. Hamdadou and B. Beldjillali. L’information spatiale dans les systèmes d’aide à la décision, exemple : SOLAP et la gestion forestière. Journées Nationales sur l’informatique et ses applications, JNIAK’2012, Khenchela, 29-30 Avril 2012
- C. Abdallah Bensalloua, D. Hamdadou and B. Beldjillali. SIG et partage de données dans la gestion des espaces agro forestiers. Journée nationale sur l’environnement du littoral Oranais ELO-01, Oran, 16 Avril 2012
- C. Abdallah Bensalloua, D. Hamdadou and B. Beldjillali. Vers une Solution SOLAP (Spatial On-Line Analytical Processing) pour un Outil d’Aide à la Décision : Ap

plication à l'Agroforesterie », 1st International Conference on Information Systems and Technologies "ICIST 2011" 24, 25 et 26 avril, Tébessa, Algérie, 2011

- C. Abdallah Bensalloua, D. Hamdadou and K. Benhanifia. SIG et partage de données dans la gestion des espaces agro forestiers. 1er séminaire international euro-méditerranéen sur l'aménagement du territoire, la gestion des risques et la sécurité civile, (Géomatique des risques spatialisés, de la recherche à l'action territoriale) du 27-28- 29 avril , Batna, Algérie, 2010

Bibliographie

- E. Abdi, B. Majnounian, A. Darvishsefat, Z. Mashayekhi, and J. Sessions. A gis-mce based model for forest road planning. *Journal of Forest Science*, 55(4) :171–176, 2009.
- A. Abello, J. Samos, and F. Saltor. Yam2 : a multidimensional conceptual model extending uml. *Information Systems*, 31(6) :541–567, 2006.
- M. Acosta and S. Corral. Participatory multi-criteria assessment of forest planning policies in conflicting situations : The case of tenerife. *Forests*, 6 :3946–3969, 2015.
- E. Akay, A. and A. Erdogan. Gis-based multi-criteria decision analysis for forest fire risk mapping. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-4/W4 :25–30, 2017. doi : 10.5194/isprs-annals-IV-4-W4-25-2017.
- E. Alnafie, D. Hamdadou, A. Benyelloul, and M. Messaoudi. Optimisation de la méthode multicritère electre iii par les colonies de fourmis. In *CIP2013 9ème Conférence Internationale Conception Production Intégrées*. Université Aboubekr Belkaid Tlemcen Algérie 21-23 Octobre 2013.
- J. Ananda and G. Herath. A critical review of multi-criteria decision making methods with special reference to forest management and planning. *Ecological Economics*, 68 : 2535–2548, 2009.
- H.I Ansoff. *Stratégie du développement de l'entreprise : une approche méthodologique du management stratégique dans le dernier quart du XXe siècle*. organisation edition, 1989.
- A.D. Ayadi. Optimisation multicritere de la fiabilite : Application du modele de goal programming avec les fonctions de satisfactions dans l'industrie de traitement de gaz. these de doctorat en science de l ingénieur. université d'angers-université de sfax. 2010.
- M. Bakillah, M. A. Mostafavi, and Y. Bédard. Développement d'une approche géosématique intégrée pour ajuster les résultats des requêtes spatiotemporelles dans le domaine forestier. In *Proceedings of the Géomatique Au cœur des processus colloquy*. Montreal Canada 2006.
- M. Barr and L. Bellatreche. Approche dirigée par les fourmis pour la fragmentation horizontale dans les entrepôts de données relationnels. *Revue Nature et Technologie*, 06 :16–24, Janvier 2017.

- C. Batini and M. Scannapieca. Data quality : Concepts, methodologies and techniques. *Springer-Verlag New York Inc*, 2006.
- Y. Bédard and J. Han. Geographic data mining and knowledge discovery, chapter fundamentals of spatial data warehousing for geographic knowledge discovery. taylor and francis. 2009.
- O. Bedel. Un système d'information logique pour l'organisation et la recherche de données géolocalisées. thèse, université de rennes 1, rennes, france, 2009.
- M. Belhadj-Aissa, A. Belhadj-Aissa, and Y. Smara. Application du sig et de la télédétection dans la gestion des feux de forets en algerie. In *2nd FIG Regional Conference Marrakech, December 2-5*. Morocco 2003.
- L. Bellatreche. Une méthodologie pour la fragmentation dans les entrepôts de données. *INFORSID*, pages 2–5, 2003.
- L. Bellatreche, M. Schneider, H. Lorinquer, and M. Mohania. Bringing together partitioning, materialized views and indexes to optimize performance of relational data warehouses. *Y. Kambayashi et al. (Eds.) : DaWak 2004*, 3181 :15–25, 2004.
- V. Belton and T.J. Stewart. *Multiple Criteria Decision Analysis - An Integrated Approach*. Kluwer Academic Publishers, Boston, 2002. ISBN 079237505X.
- G. Bentrup and T. Leininger. Agroforestry mapping : the way gis. *Journal of Soil and Water Conservation*, 57(6) :148–153, 2002.
- S. Bimonte. Intégration de l'information géographique dans les entrepôts de données et l'analyse en ligne : de la modélisation à la visualisation. thèse de doctorat en informatique. institut national des sciences appliquées de lyon, france, 2007.
- S. Bimonte, A. Tchounikine, and M. Miquel. Towards a spatial multidimensional model. In *Proceedings of the 8th ACM international workshop on Data warehousing and OLAP*, pages 39–46. Bremen. Germany, 2005.
- S. Bimonte, A. Tchounikine, and M. Miquel. Spatial olap : Open issues and a web based prototype. In *International Conference on Geographic Information Science*, pages 1–11, 2007.
- S. Bimonte, M. Villanova-Oliver, and J. Gensel. A multidimensional model for correct aggregation of geographic measures. evolving application domains of data warehousing and mining. trends and solutions. *IGI Global*, pages 162–183, 2009.
- S. Bimonte, K. Boulil, and F. Pinet. Cohérence logique dans les systèmes olap spatiaux : un état de l'art. *Revue Internationale de Géomatique, Lavoisier*, 26(1) :97–131, 2016.
- M. Bohnlein, M. Plaha, and A. Ulbrich-Vom-Ende. *Visual Specification of Multidimensional Queries based on a Semantic Data Model*. In : von Maur E., Winter R. (eds) *Vom Data Warehouse zum Corporate Knowledge Center*. Physica, Heidelberg, 2002.

- K. Boulil. Une approche automatisée basée sur des contraintes d'intégrité définies en uml et ocl pour la vérification de la cohérence logique dans les systèmes solap : Applications dans le domaine agri-environnemental. thèse de doctorat en informatique. université blaise pascal - clermont-ferrand ii, france, 2012.
- K. Boulil, L.B. Florence, B. Sandro, G. Corinne, and C. Flavie. Multidimensional modeling and analysis of large and complex watercourse data : an olap-based solution. *Ecological Informatics*, 24 :90–106, 2014.
- K. Boulil, B. Sandro, and P. Francois. Conceptual model for spatial data cubes : A uml profile and its automatic implementation. *Computer Standards and Interfaces*, 38 : 113–132, 2015.
- G Bressy and C. Konkuyt. *Economie d'entreprise*. Dollaz, Paris, first edition, 2004.
- S. Chakhar. Cartographie décisionnelle multicritère : Formalisation et implémentation informatique. thèse de doctorat en informatique. université paris dauphine, france, 2006.
- S. Cockcroft. A taxonomy of spatial data integrity constraints. *Geoinformatica*, 1(4) : 327–343, 1997.
- E.F. Codd. Providing olap (on-line analytical processing) to user-analysts : An it mandate. technical reports. ibm, 1993.
- M. L. Damiani and S. Spaccapietra. *Spatial Data Warehouse Modelling. in Processing and Managing Complex Data for Decision Support*, pages 1–27. Hershey, IGI Global, 2006.
- F.S. Danks and D.R. Klein. Using gis to predict potential wildlife habitat : a case study of nuskoxen in northern alaska. *International Journal of Remote Sensing*, 23(21) :4611–4632, 2001. doi : 10.1080/01431160110113890.
- J.E. De Steiguer, L. Liberti, A. Schuler, and B. Hansen. Multi-criteria decision models for forestry and natural resources management : An annotated bibliography. united states department of agriculture, forest service, northeastern forest experiment station. upper darby, pa, usa, 2003.
- C.E. Dyreson, T.B. Pedersen, and C.S. Jensen. Incomplete information in multidimensional databases. *Multidimensional databases. IGI Global publishing*, page 282–309, 2003.
- A. Fediaevsky. Contribution à la mise en place d'un système d'information géographique pour l'analyse du risque « peste bovine » en centrafrrique. thèse de doctorat, école nationale vétérinaire d'alfort, paris, france, 2002.
- M. E. Feeney and I. Williamson. The role of institutional mechanisms in spatial data infrastructure development that supports decision making. *Cartography Journal*, 312 : 21–37, 2002.
- E. Ferragu. *Modélisation des Systèmes d'Information Décisionnels : Techniques de Modélisation Conceptuelle et Relationnelle des Entrepôts de Données*. Vuibert, 2013.

- V. Fontana, A. Radtke, V.B. Fedrigotti, U. Tappeiner, E. Tasser, S. Zerbe, and T. Buchholz. Comparing land-use alternatives : Using the ecosystem services concept to define a multi-criteria decision analysis. *Ecological Economics*, 93 :128–136, 2013.
- F. Ghozzi, F. Ravat, O. Teste, and G. Zurfluh. Modèle multidimensionnel à contraintes. *Revue d'Intelligence Artificielle*, 17(1-3) :43–55, 2003.
- O. Glorio and J. Trujillo. An mda approach for the development of spatial data warehouses. *Data Warehousing and Knowledge Discovery. I.-Y. Song, J. Eder and T. Nguyen, Springer Berlin - Heidelberg*, 5182 :23–32, 2008.
- M. Golfarelli and S. Rizzi. *Data Warehouse Design : Modern Principles and Methodologies*. McGraw-Hill. New York, 2009.
- J. Gray, S. Chaudhuri, A. Bosworth, A. Layman, D. Reichart, M. Venkatrao, F. Pellow, and H. Pirahesh. Data cube : A relational aggregation operator generalizing group-by, cross-tab, and sub-totals. *Data Mining and Knowledge Discovery*, 1(1) :29–53, 1997.
- R. Greene, R. Luther, J.E. Devillers, and B. Eddy. An approach to gis-based multiple criteria decision analysis that integrates exploration and evaluation phases : Case study in a forest-dominated landscape. *Forest Ecology and Management*, 260 :2102–2114, 2010.
- A.M. Guillermo and M. Phil. Application de l'analyse multicritère à l'évaluation des critères et indicateurs- manuels de critères et indicateurs pour la gestion durable des forêts. technical report. french version, cirad : International cooperation center in agronomic research for development, department of forests, cifor, montpellier – france, 2000.
- L.E. Guimond. Conception d'un environnement de découverte des besoins pour le développement de solutions solap. master's thesis presented at high studies faculty, as part of the masters program in geomatics for obtaining the degree of master of science, laval university, quebec, 2005.
- D. Hamdadou. Un modèle de prise de décision pour l'aménagement de territoire, une approche multicritère et une approche de négociation. thèse de doctorat en sciences, université d'oran1 ahmed ben bella, algérie, 2008.
- D. Hamdadou and L. Thérèse. A multicriteria group decision support system for industrial diagnosis. *INFOCOMP : Journal of Computer Science*, 10(3) :12–24, 2011.
- D. Hamdadou, S. Oufella, and K. Bouamrane. Un système d'aide à la décision spatiale de groupe : Couplage analyse multicritère et théorie des jeux satisfaisants. In *Conférence sur les Avancées des Systèmes Décisionnels (ASD'2013)*. Marrakech, Maroc, 2013.
- D. Hamdadou, K. Bouamrane, and Naoui O.k. *Méthodologie Multicritères d'Aide à la Décision, Cours et Exercices*. Département Informatique, Faculté des sciences Exactes et Appliquées, Université d'Oran1 Ahmed Ben Bella, Algérie., 2016.

- O. Huber. *Information-processing operators in decision making. In Process and structure in human decision-making.* ed. H. Montgomery and O. Svenson, 3–21. New York : John Wiley and Sons, Inc, 1989.
- C.A. Hurtado and A.O. Mendelzon. A multicriteria group decision support system for industrial diagnosis. *Proceedings of the 8th International Conference on Database Theory, Springer-Verlag*, pages 375–389, 2001.
- C.A. Hurtado, C. Gutierrez, and A.O. Mendelzon. Capturing summarizability with integrity constraints in olap. *ACM Transactions on Database Systems*, 30(3) :854–886, 2005.
- W.H. Inmon. *Building the Data Warehouse.* 4th Edition, Wiley press, 2005.
- C.S. Jensen, A. Kligys, T.B. Pedersen, and I. Timko. Multidimensional data modeling for location-based services. *The VLDB Journal*, 13(1) :1–21, 2004.
- T. Juan, S. Emilio, F.M. Eduardo, and P. Mario. A uml 2.0 profile to define security requirements for data warehouses. *Computer Standards and Interfaces*, 31 :969–983, 2009.
- P. Julien. Conception d’une légende interactive et forable pour le solap. master’s thesis presented at high studies faculty, as part of the masters program in geomatics for obtaining the degree of master of science, laval university, quebec, canada, 2004.
- C.P. Keller. Decision making using multiple criteria”, in ncgia core curriculum, application issues in gis eds goodchild, m f, kemp, k, national center for geographical information analysis. lecture 57. santa barbara, canada, 1989.
- K.K. Kemp. *Encyclopedia of geographic information science.* Thousand Oaks, CA : SAGE Publications, Inc, 2008.
- J. Kilgo, D. Gartner, B. Chapman, J. Dunning, K. Franzels, and S. Gauthreaux. A test of an expert based bird-habitat relationship model in south carolina. *Wildlife Society Bulletin*, 30(3), 2002.
- D.M. Kilgour and C. Eden. *Handbook of Group Decision and Negotiation.* Springer Science and Business Media : New York, NY, USA, Volume 4, 2010.
- R. Kimball. *The data warehouse toolkit.* John Wiley and Sons, 1996.
- R. Kimball and M. Ross. *Entrepôts de données, Guide pratique de modélisation dimensionnelle.* Vuibert Informatique Press, Paris, France, 2003.
- J. Lechtenborger and G. Vossen. Multidimensional normal forms for data warehouse design. *Information Systems*, 2003.
- H.-J. Lenz and A. Shoshani. *Summarizability in OLAP and statistical data bases.* IEEE, 1997.

- M. Levesque, Y. Bédard, M. Gervais, and R. Devillers. Towards managing the risks of data misuse for spatial datacubes. In *5th international symposium of spatial data quality, june 13-15, Enschede, Netherland, 2007*.
- S. Lujan-Mora, J. Trujillo, and I.Y. Song. A uml profile for multidimensional modeling in data warehouses. *Data and Knowledge Engineering*, 59(3) :725–769, 2006.
- E. Malinowski and E. Zimanyi. Hierarchies in a multidimensional model : From conceptual modeling to logical representation. *Data and Knowledge Engineering*, 59(2) :348–377, 2006.
- E. Malinowski and E. Zimanyi. *Advanced Data Warehouse Design : From Conventional to Spatial and Temporal Applications*. Springer, 2008. ISBN 978-3-540-74404-7.
- M. Maryvonne, Y. Bédard, and B. Alexandre. Conception d’entrepôts de données géospatiales à partir de sources hétérogènes, exemple d’application en foresterie. *Ingénierie des Systèmes d’information*, 7(3) :89–111, 2002.
- J.N. Mazon, J. Trujillo, and J. Lechtenbörger. A set of qvt relations to assure the correctness of data warehouses by using multidimensional normal forms. In *25th international conference on Conceptual Modeling. Tucson, AZ, Springer-Verlag : 385-398, 2006*.
- J.N. Mazon, J. Lechtenbörger, and J. Trujillo. A survey on summarizability issues in multidimensional modeling. *Data and Knowledge Engineering*, 68(12) :1452–1469, 2009.
- A. Montserrat and C. Serafin. Multicriteria decision analysis and participatory decision support systems in forest management. *Forests review*, 8 :116, 2017.
- L. Naoum. Un modèle multidimensionnel pour un processus d’analyse en ligne de résumés flous. thèse de doctorat, université de nantes, france., 2006.
- T. Niemi, J. Nummenmaa, and P. Thanisch. Logical multidimensional database design for ragged and unbalanced aggregation hierarchies. In *3rd International Workshop on Design and Management of Data Warehouses : 7, 2001*.
- E. Nordstrom, L.O. Eriksson, and K. Ohman. Integrating multiple criteria decision analysis in participatory forest planning : Experience from a case study in northern sweden. *Forest Policy and Economics*, 12 :562–574, 2006.
- G. Octavio, J.N. Mazon, G. Irene, and J. Trujillo. A personalization process for spatial data warehouse development. *Decision Support Systems*, 52 :884–898, 2012.
- S. Oufella, D. Hamdadou, and K. Bouamrane. Conception d’un système interactif d’aide à la décision collective en localisation spatiale. In *Conférence Internationale sur l’Informatique et ses Applications (CIIA’09), Saida, Algérie, 2009*.
- M. Paegelow. Géomatique et géographie de l’environnement. de l’analyse spatiale à la modélisation prospective. habilitation à diriger des recherches. université toulouse le mirail - toulouse ii, france, 2004.

- T.B. Pedersen and C.E. Jensen, C.S. and Dyreson. Extending practical pre-aggregation in on-line analytical processing. In *25th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc. : 663-674*, 1999.
- T.B. Pedersen and N. Tryfona. Pre-aggregation in spatial data warehouses. In *7th International Symposium on Advances in Spatial and Temporal Databases, Springer-Verlag : 460-480*, 2001.
- T.B. Pedersen, C.S. Jensen, and C.E. Dyreson. A foundation for capturing and querying complex multidimensional data. *Information Systems*, 26(5) :383–423, 2001.
- J. Pictet. *Dépasser l'évaluation environnementale. Procédure d'étude et insertion dans la décision globale*. Romandes Polytechnics and Universities Press. Lausanne, Switzerland, 1996.
- F. Pinet. Modélisation des contraintes d'intégrité dans les systèmes d'information environnementaux. habilitation à diriger les recherches. université blaise pascal, clermont-ferrand : 112, france, 2010.
- F. Pinet and M. Schneider. A unified object constraint model for designing and implementing multidimensional systems. *Journal on Data Semantics*, 13 :37–71, 2009.
- F. Pinet and M. Schneider. Precise design of environmental data warehouses. *Operational Research*, 10(3), 2010.
- F. Pinet, A. Miralles, S. Bimonte, F. Vernier, N. Carluer, V. Gouy, and S. Bernard. The use of uml to design agricultural data warehouses. In *International Conference on Agricultural Engineering, AGENG*, 2010.
- E. Pourabbas. *Geographical Information Systems : Trends and Technologies*. CRC Press, 2014.
- N. Prat, I. Wattiau, and J. Akoka. Representation of aggregation knowledge in olap systems. In *The 18th European Conference on Information Systems*, 2010.
- F. Ravat, O. Teste, and Zurfluh G. Manipulation et fusion de données multidimensionnelles. In *RNTI-E-3 Revue des Nouvelles Technologies de l'Information, réd., Extraction et Gestion des Connaissances*, 1, 2005.
- R. Reix. *Systèmes d'information et management des organisations*. Sème édition. Paris : Vuibert, 2004.
- S. Rivard and J. Talbot. *Le développement de systèmes d'information*. 3ème édition. Presses de l'Université du Québec. Canada, 2001.
- S. Rivest. *Investigation des modes d'intégration physique entre un serveur de base de données multidimensionnelle et un SIG*. Rapport de DEA Informatique, Laval : Université Laval, Canada, 2000.

- S. Rivest, Y. Bédard, M.J Proulx, and M. Nadeau. Solap : a new type of user interface to support spatio-temporal multidimensional data exploration and analysis. In *Workshop ISPRS, October 2-3 Québec, Canada, 2003*.
- N. Roussat. Optimisation des stratégies de gestion du contenu « matières premières et énergie » des déchets produits et utilisés sur un territoire urbanisé, application aux déchets de démolition du grand lyon. thèse de doctorat en sciences de l'environnement industriel et urbain. ecole doctorale de chimie de lyon. france, 2007.
- B. Roy. Méthodologie multicritères d'aide à la décision, collection gestion, série production et techniques quantitatives appliquées à la gestion. *Economica. Paris*, 1985.
- B. Roy. Science de la décision ou science de l'aide à la décision. *Revue internationale de Systémique*, 6(5), 1993.
- T.L. Saaty. *L'aide multicritères à la décision*. Décider face à la complexité : une approche analytique multicritère d'aide à la décision (traduit par Lionel Dahan). Paris : Entreprise Moderne d'Édition, 1984.
- M. Salehi. Developing a model and a language to identify and specify the integrity constraints in spatial datacubes. thèse de doctorat, faculté des études supérieures de l'université laval, canada, 2009.
- M. Salehi, Y. Bédard, M. Mostafavi, and J. Brodeur. On languages for the specification of integrity constraints in spatial conceptual models. *Advances in Conceptual Modeling – Foundations and Applications. J.-L. Hainaut, E. Rundensteiner, M. Kirchberger et al, Springer Berlin - Heidelberg*, 4802 :388–397, 2007.
- M. Salehi, Y. Bédard, , and S. Rivest. Formal conceptual model and definition framework for spatial data cubes. *Geomatica*, 64 :313–326, 2010.
- C. Sapia. On modeling and predicting query behavior in olap systems. In *1st international workshop on design and management of data warehouse. Heidelberg. germany*, pages 1–10, 1999.
- T. Sboui, M. Salehi, Y. Bédard, and S. Rivest. Catégorisation des problèmes d'intégration des modèles des cubes de données spatiales. In *8emes Journées Francophones Extraction et Gestion des Connaissances, Sophia Antipolis*, 2008.
- H.A. Simon. *The new science of management decision*. Englewood Cliffs, NJ : Prentice-Hall, 1960.
- H.A. Simon. *Administration et processus de décision*. 1er Ed, Economica, Paris, 1983.
- U. Simon, R. Bruggmann, and S. Buden. Aspects of decision support in water management –example berlin and postdam (germany) i, spacially differentiated evaluation. *water research*, 38(7) :1809–1816, 2004.
- J. Simos. L'évaluation environnementale : Un processus cognitif négocié. phd thesis, 1990.

- T. Siqueira, R.C. Mateus, R.R. Ciferri, V.C. Times, and C.D.A. Ciferri. Querying vague spatial information in geographic data warehouses advancing geoinformation science for a changing world. *Lecture Notes in Geoinformation and Cartography*, 1 :379–397, 2011.
- H. Sivakumar, C. Matt, M. Sethu, Z. Robert, and G.Y.L. Denny. *Administration et processus de décision*. Professional Microsoft SQL Server Analysis Services 2008 with MDX, Indianapolis, Indiana, USA. Wiley Publishing Inc Press, 2009.
- 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*, 12(6) :938–958, 2000.
- R. Sugumaran and J. DeGroot. *Spatial Decision Support systems, Principles and Practices*. CRC Press, New York, USA, 2011.
- O.A. Taher and M. Maryvonne. *Multidimensional Structures Dedicated to Continuous Spatiotemporal Phenomena*. BNCOD, LNCS 3567, Springer-Verlag Berlin Heidelberg, 2005.
- O. Teste. Modélisation et manipulation d’entrepôts de données complexes et historisées. thèse de doctorat en informatique, université paul sabatier - toulouse iii, france, 2000.
- R. Torlone. Conceptual multidimensional models. *Multidimensional databases, IGI Publishing*, 5 :69–90, 2003.
- J. Trahand. *Aide à la décision*. In R. Le Duff (Ed.), *Encyclopédie de la gestion et du management*. Paris, 1999.
- T. Ubeda. contrôle de la qualité spatiale des bases de données géographiques : cohérence topologique et corrections d’erreurs. thèse de doctorat en informatique. institut national des sciences appliquées de lyon. france, 1997.
- R. Villarroel, E. Fernández-Medina, M. Piattini, and J. Trujillo. A uml 2.0/ocl extension for designing secure data warehouses. *Journal of Research and Practice in Information Technology*, 38(1) :31–43, 2006.
- Ph. Vincke. *L’aide multicritères à la décision*. Editions de l’Université Libre de Bruxelles, 1989.
- L. Wang and Q. Cheng. Web-based collaborative decision support services : Concept, challenges and application. In *ISPRS Technical Commission II Symposium, Vienna*, 2006.
- Y. Wang and D.K. Moskovits. Tracking fragmentation of natural communities and changes in land cover : applications of landsat data for conservation in urban landscape (chicago wilderness). *Conservation Biology*, 15(4) :835–843, 2001.
- L. Warnecke, R.V. Nanni, Z. Nedovic-Budic, and W. Stiteler. *L’aide multicritères à la décision*. Remote Sensing and Geographic Information Technology in the Nation’s 50

- State Forestry Organizations." GeoManagement Associates Inc., Syracuse, New York, 2002.
- R. Weibel and G. Dutton. Generalizing spatial data and dealing with multiple representations. In : Longley, P., Goodchild, M., Maguire, D. and Rhind, D. *Geographic Information Systems and Science*, New York : John Wiley and Sons, pages 125–155, 2001.
- J.A. Wilson and K.W. Lowe. Planning for the restoration of native biodiversity within the goulburn broken catchments, victoria using spatial modelling. *Ecological Management and Restoration*, 4(3) :212, 2003.
- C. Xi-Qian, C. Zhong-Xian, and C. Xiu-Kun. Applying dp to etl of spatial data warehouse. In *International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826)*, 2004.
- F.Z. Younsi, D. Hamdadou, and A. Benyettou. Proposition d'un modèle décisionnel en aménagement du territoire par utilisation des sig et réseaux de neurones. In *Colloque sur l'Optimisation et les Systèmes d'Information (COSI' 07)*, Oran, Algérie, 2007.
- F.Z. Younsi, D. Hamdadou, and B. Beldjilali. Proposition d'un système interactif d'aide à la décision spatiale : Télédétection, sig et analyse multicritère. In *Conférence Internationale sur l'Informatique et ses Applications (CIIA'09)*, Saida, Algérie, 2009.
- M. Zaamoune. Intégration et optimisation des grilles régulières de points dans une architecture solap relationnelle. thèse de doctorat en informatique, université de blaise pascal – clermont ii, 2015.
- F.A. Zemri and D. Hamdadou. Integration of multi criteria analysis methods to a spatio temporal decision support system for epidemiological monitoring. In *The International Conference on Advanced Aspects of Software Engineering (ICAASE'2014)*, Constantine Algérie, 2014.
- Z. Zhang, R. Sherman, Z. Yang, R. Wu, W. Wang, M. Yin, G. Yang, and X. Ou. Integrating a participatory process with a gis-based multi-criteria decision analysis for protected area zoning in china. *Journal for Nature Conservation*, 21 :225–240, 2013.
- E. ZIYATI. Optimisation de requêtes olap en entrepôts de données approche basée sur la fragmentation génétique, thèse de doctorat en informatique et télécommunications, université mohammed v – agdal, maroc, 2010.

Spatial OLAP and Multicriteria Integrated Approach for Decision Support System: Application in Agroforestry Management

Charef Abdallah Bensalloua, University of Oran 1 Ahmed Benbella, Oran, Algeria

Djamila Hamdadou, University of Oran 1 Ahmed Benbella, Oran, Algeria

ABSTRACT

This article describes how Spatial On-Line Analytical Processing, such a decisional technology, offers the possibilities of spatial and multidimensional analysis of data stored in multidimensional structure namely spatial data warehouse. However, this technology is limited in the quality aspect of the decision related to the multicriteria consideration. In the current article, the objective is to propose a Spatial Decision Support System namely “Silvicultura” for facilitating decision making in complex situations. This approach is based on integrating multicriteria analysis with SOLAP in order to enrich the spatial and multidimensional analysis with the contribution of MCA tools for mitigating conflict situations. The authors have based their proposal modeling on Unified Modeling language, since it is a well-known standard modeling language and can be easily extended for multidimensional modeling. Finally, in order to validate their proposal, the authors present a case study to show how to use it in the agroforestry management.

KEYWORDS

Agroforestry, Multicriteria Analysis, Multidimensional Analysis, SOLAP, Spatial Analysis, Spatial Data Warehouse, Spatial Decision Support System, UML Modeling

1. INTRODUCTION

Organizations in all governmental, private, economic or social sectors, etc. need data analysis tools allowing them to have indicators of the evolution of their activities.

These tools must be capable for storing and analyzing large volume of data from multiple sources taking advantage of the service offered by modern applications and technologies. They must also be able to provide a clear vision of the activity of an organization and its environment in order to develop its strategy of production, preservation of the existing patrimony and the regularization of its overall functioning.

Computerized information systems are typically used providing data storage and analysis mechanisms and are adapted to decision analysis queries (Kimball & Ross, 2003). They are called Decision Support Systems (DSS) with the aim of assisting decision-makers in their job.

DOI: 10.4018/IJDSST.2018070101

In this context, on one hand, Data Warehouse (DW) systems provide storage, integration and logging space for all the data needed for decision analysis. On another hand, On-Line Analytical Processing (OLAP) defines intuitive tools for a simple and interactive exploration of this data.

Furthermore, a Spatial Decision Support System (SDSS) is a specific tool for problem solving in management process. Basing on a spatial data infrastructure, this system is used to facilitate the evolution of actions which have spatial nature.

Nowadays, in Algeria, we note the increasing need to these systems in various domains namely the agroforestry area management. It includes forests, forest vocation areas, wetlands ... etc. These spaces are defined in the forests general regime¹.

In the management and preservation of forestry resources, efforts must allow to:

- Preserve, regenerate and develop the existing properties through silvicultural activities and management plans;
- Maintain and strengthen protected areas through the development and implementation of a management plan;
- Maintain and develop wetlands by their identifications, classification and protection of priority areas of economic and ecological interest;
- Rehabilitate the natural space by restoring recreational forests and green spaces.

In addition, in the fight against erosion and mountain agriculture, forestry policy aims to:

- Preserve the coastline by the stabilization based on a natural plant;
- Maintain the natural balance of ecosystems by the stabilization of the dune covered;
- Protect watersheds threatened by erosion, by the integrated development planning.

As such, the services responsible for managing the forest area have a large volume of complex and multi-source data. Hence the need for effective systems to access and analyze data for using them in planning purposes. This need is summarized by their role in the calculation and diffusion of statistical results that impact on the curve of economic and social development.

In this context, the development of an interactive tool for the decision-making to manage agroforestry areas at several observation axes, such as: cadastral properties, management of wildlife, silvicultural activities, etc., involves intensive use of databases, taking advantage of the spatial applications. This situation involves manipulating multisource data, where spatial data management is the task of the Geographic Information System (GIS) which allows the spatial analysis.

A GIS can be defined as a "...data management system designed to input, store, retrieve, manipulate, analyze, and display spatial data for the purposes of research and decision-making..." (DeMers, 1997). Indeed, GIS is used to manipulate spatially referenced data of objects and phenomena in geographic space. GIS is widely used in land-use planning and natural resources management.

However, GIS is based on transactional processes and is unable to perform multidimensional analysis offered by On-Line Analytical Processing (OLAP) technologies. That's why a new technology has emerged known as the Spatial-OLAP (SOLAP). In fact, this last, allows studying the geographical distribution of phenomenon and effectuating comparison across geographic granularity by the visualization of the analysis results on tables, diagrams and maps.

Nevertheless, SOLAP technology alone still suffers from several drawbacks largely due to a lack of capacity for supporting spatial problems.

To evolve into a true decision support tool, the solution, in our proposal, is to combine it with Multicriteria Analysis (MCA) which offers several advantages when we must take into account conflicting interests.

Indeed, MCA is a decision-making tool which allows integrating methods to give solutions for complex problems with different groups of interest. It evaluates the relative importance of the criteria and reflects their importance in the final result (Belton & Stewart, 2002). MCA allows the treatment of both quantitative and qualitative data and can be used when the problem to be treated is well defined (Kilgour & Eden, 2010).

Within the context of agroforestry policy, a SDSS should consider economic, environmental and social criteria. MCA is one of the approaches that are frequently used in agroforestry planning. It considers several criteria to help explore important decisions, such as group decisions or situations where there are many parties involved (Belton & Stewart, 2002).

In the current study, we develop an interactive spatial decision support system by integrating MCA engine and SOLAP. Indeed, the integration SOLAP-AMC in a decision model may be able to help decision maker for effective management of various activities related to agroforestry areas.

The present paper is organized as follows: In Section 2, we explore examples of related work, while our contribution is explained in Section 3. The proposed spatial decision support system will be presented in Section 4. In Section 5, the adopted decisional model is presented. In Section 6, the process adopted by “*Silvicultura*” is explained while its UML modeling is presented in Section 7. The multidimensional modeling of the agroforestry spatial data warehouse is explained in Section 8. A case study is presented for the experimentation in Section 9. Finally, in Section 10, our main conclusions will be stated and our future work will be introduced.

2. RELATED WORK

There are several examples of the application of multicriteria methodologies, including forest research, such as the one carried out to compare alternatives for land use by (Fontana et al. 2013). In this study, authors obtained a classification of the alternatives through the PROMETHEE II method. Another example is that presented in (Nordström et al., 2010). In this study, the Analytic Hierarchy Process (AHP) was used for planning the urban forest in Sweden by determining the strength of preference for one criterion or alternative over another. AHP was also used in the research presented in (Ananda, 2007) and (Diaz-Balteiro & Romero, 2008). A critical review of multicriteria decision making methods with special reference to forest management and planning is presented in (Ananda & Herath, 2009). Authors in (De Steiguer, J.E. et al. 2003) have annotated bibliography of multicriteria decision models for forestry and natural resources management. Participatory multicriteria assessment of forest planning policies in conflicting situations is presented in (Acosta, M. & Corral, S. 2015). Authors in (Fontana et al., 2013) have used the ecosystem services concept to define a multi-criteria decision analysis.

Moreover, GIS is being used by various disciplines to perform a variety of spatial analysis, including overlaying combinations of features, analyzing networks; buffer zoning, etc. The fact that in agroforestry spatial characteristics are widely used, several agroforestry-specific GIS studies have been launched. For example, in the case presented in (Bentrup et al., 2000), GIS is presented to facilitate conservation buffer planning in the central USA. In (Greene et al., 2010) an approach to integrate GIS and multicriteria decision analysis that integrates exploration and evaluation phases in the study of a forest-dominated landscape. Integrating a participatory process with a GIS and multicriteria decision analysis for protected area zoning in China is presented by authors in (Zhang et al., 2013).

Besides, OLAP tools are deployed to access, visualize, and analyze integrated, aggregated and summarized data for supporting analytical processes. A combination of OLAP with GIS concepts spawned the Spatial OLAP which can be defined as “...a visual platform built especially to support rapid and easy spatio-temporal analysis and exploration of data following a multidimensional approach...” (Rivest et al., 2003).

In the literature, we find some authors that address issues related to the use of SOLAP systems in forestry. We classify their studies into three categories, depending on the factors affecting the

development and the use of such a system namely: the designer, the data manipulated and the end user. This categorization can be summarized as follows:

1. **Studies addressing the designer:** This category refers to methods of modeling and the development of the spatial data cube structure. As such, in forestry, authors in (Bakillah et al., 2006) discuss the issue of changes in the structure of the data cube at semantic and geometric level. Indeed, the proposed model is based on restoring links between versions of the cube. Semantically, their approach restores the links using a semantic similarity function based on ontology and reflects the finest level of definition of concepts. For the geometric level, this approach is based on the Quadtree indexing method to form a matrix of spatial correspondence between the geometries of different epoch;
2. **Studies addressing data manipulated in the system:** In some cases, users use more than one spatial data cube to ensure decision-making needs. In (Sboui et al., 2008), authors take an example in the analysis of the risk of wildfire, where users have two spatial data cubes. Since these cubes are usually modeled differently from a designer to another, the authors proposed a categorization of heterogeneity problems taking into account the different elements of spatial data cubes. Therefore, they defined four categories including: the heterogeneity Cube to Cube, the heterogeneity dimension to dimension, the heterogeneity level to level and heterogeneity measure to measure. For each category, they considered the different components of the data cubes model, including the schema and metadata. In the same context, another study presented in (Taher & Maryvonne, 2005) for the dimensions in the hypercube. Indeed, the multidimensional approach proposed by the authors, is based on the concept of facts seen in different dimensions. These dimensions are considered as analysis axes forming a space for identifying the facts by a set of coordinates. The dimensions therefore, have discrete values. However, these values are not sufficient when analyzing continuous natural phenomena, such as weather or pollution. The authors introduced mechanisms based on interpolation in the spatial and temporal dimensions, which give the user the impression of navigating a continuous hypercube. In (Maryvonne et al, 2002), the authors propose solutions for designing multidimensional structures when the data sources are heterogeneous in temporal, spatial and semantics point of view. Their goal is to enable the extraction of geographical knowledge by exploring detailed data associated to an epoch;
3. **Studies addressing the end-users:** In order to propose solutions for the interactive processing, where the user poses a spatial action in order to visualize quickly the consequences, we argue that the user experience plays an important role in the decision process.

In fact, Julien (2004) proposes solution to design an interactive legend for SOLAP. In this context, he studied graphic semiotics and its applicability to the multidimensional analysis.

Another issue regarding the use of SOLAP system is that treated in (Guimond, 2005). As such, the author poses the problem that some users know their needs in terms of decision support after a period of use. For this reason, the author developed scenarios for integrating layout design in a process of developing SOLAP applications.

In agroforestry, various studies have treated the problems in the design, data integration or the use of a SOLAP system. They report the complexity of an effective management of natural areas with ecological, economic and social purposes. However, in most studies, the discussed problems are related to the heterogeneity of manipulated data, or development methods. So, few studies have addressed issues related to the end users.

In the field of agroforestry, the same territory is usually, under the control of several actors associated to different disciplines including: civil engineering, silvicultural activities, territorial management, exploiting of forest products, and even mountain farming. This requires an integrated management of the forest areas.

However, the use of SOLAP tool by different actors independently allows them to extract decisional indicators which are heterogeneous and lead to conflict. This does not allow an integrated management taking into account technical, ecological and social considerations. As a result, decision-making process will not benefit from any experience between different actors. Indeed, the integration SOLAP-AMC may help decision maker for effective management of agroforestry areas taking profit from the spatial analysis of GIS, the multidimensional analysis of OLAP and AMC which allows analysis of complex problems with different requirements.

3. CONTRIBUTION

The main objective of this study involves the design and development of a spatial decision support system and its application in the agroforestry. The system should take into account all the complex requirements of this domain related to the spatial dimension of data on the one hand and different economic, ecologic and social interests on the other hand.

Throughout the current study, other objectives are achieved. Indeed, our proposal is based on evaluation of management criteria, defined by the interveners using Multicriteria Analysis (MCA) in one hand, and the extraction of indicators for the analysis of the situation through SOLAP on the other hand. The results of this analysis should enable actors in the agroforestry management defining the set of candidate actions. MCA then, allows ranking these actions to be implemented. The final ranked list of actions is the proposed solution in case of the presence of potential risk, which can threaten the functioning of agroforestry ecosystems, or a preventive measure solution otherwise.

In fact, the following tasks are carried out:

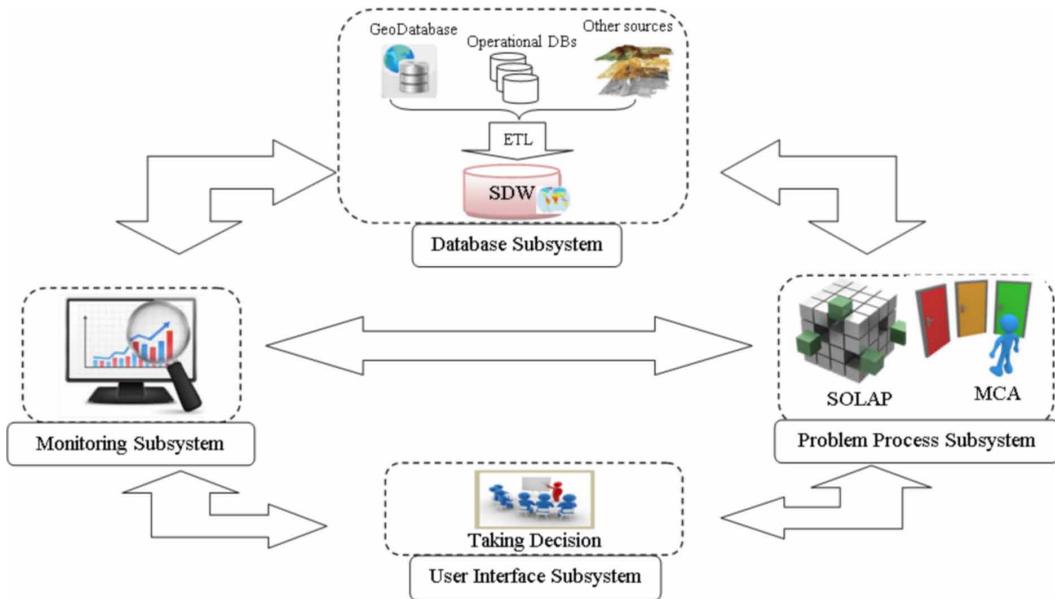
- The proposal of a categorization of studies in the field of forestry and spatial OLAP, depending on the factors interfering namely the designer, the end user and data;
- The exploitation of Unified Modeling Language (UML)-based modeling which provides powerful constructs to represent static and dynamic aspects of the system “*Silvicultura*” such as activity, use cases and sequence diagrams;
- The proposal of an extension of the UML profile presented in (Boulil et al, 2015). This profile allows us to consider the main properties of Multi-Dimensional (MD) modeling allowing to build a data cube with several fact tables;
- The implementation of our profile to support “constellation schema” of the agroforestry Spatial Data Warehouse (SDW);
- The proposal of a tool for multidimensional analysis with Multi-Dimensional eXpressions (MDX) language using SOLAP architecture namely “*Sylvicole*”. This allows a better temporal exploration of data and spatial analysis with GIS mechanisms integrated in “*Sylvicole*”;
- The use of MCA methods such as ELECTRE I and ELECTRE III. This allows sharing experience among the interveners in decision-making process, and assumes that the solution of the problem is at least well-argued.

4. THE PROPOSED SPATIAL DECISION SUPPORT SYSTEM “SILVICULTURA”

A spatial decision support system (SDSS) is an interactive system designed to assist in decision making. It is developed for use with spatial database and/or when the solution will have spatial dimension.

The SDSS we propose namely “*Silvicultura*” is based on the determination of SOLAP and MCA role in a decision-making process for the agroforestry ecosystems management as shown in Figure 1. It is designed to bring the whole of the knowledge provided by experts in this domain.

Figure 1. The spatial decision support system “Silvicultura”. An overall view.



The development of a spatial decision support system for agroforestry management will depend on combining together four subsystems, namely: Database subsystem, Monitoring subsystem, Problem process subsystem and User interface subsystem.

4.1. Database Subsystem

The database subsystem is developed in order to build the spatial data warehouse (SDW) including all conventional and spatial data related to the agroforestry management. Data source are issued from Operational process, GeoDatabase related to GIS applications and other sources such as remote sensing. ETL process is applied for extracting, transforming and loading data in the SDW.

4.2. Monitoring Subsystem

The monitoring subsystem allows benefiting from expert knowledge. The job of the forest experts is to propose management criteria. Authors in (Guillermo & Phil, 2000) see that management criteria are the meaning of application of general principle in the forest management policy.

E.g.: for the sustainable forest management, ecosystem integrity is maintained. For this aim we can consider the management criteria: The main processes and functions of the forest ecosystem are maintained.

Subjective parameters are also proposed for the applying of MCA to reduce the list of management criteria (MC) in problem process subsystem. Experts will use the reduced list of management criteria to generate indicators which represent variables that characterize the state of the criteria; and also to generate verifiers, which are data that reinforce the facility of indicator evaluation (Guillermo & Phil, 2000). These latter will be transformed into multidimensional expressions (MDX). After obtaining result of the spatial multidimensional analysis, monitoring process is started. Indeed, according to their experience, experts can see if there is a threat for agroforestry ecosystem, then they will propose a list of actions in order to remove the risk. Else, they will affine the management criteria in order to enhance ecosystem balance.

4.3. Problem Process Subsystem

The problem process subsystem contains methods and algorithms to be applied in “*Silvicultura*”. Indeed, spatial multidimensional analysis is applied thanks to SOLAP, the core of the system. This allows responding to the MDX queries sent by monitoring subsystem. Also, MCA is applied twice: first, it is used for choosing a reduced list from management criteria (MC) proposed by monitoring subsystem to be taken into consideration at future steps of “*Silvicultura*” process. Second, MCA is used for ranking the list of actions proposed by experts for removing the risk enhancing agroforestry ecosystem.

4.4. User Interface Subsystem

The user interface subsystem is a human-machine interface used mainly by decision makers and experts for introducing subjective parameters and criteria for MCA. It is used also for introducing MDX queries for spatial MD analysis. Results are then displayed in tabular, diagram and cartographic form. This manipulation is automated thanks to the graphical interface which facilitates applying different SOLAP aggregation operators.

5. THE ADOPTED DECISIONAL MODEL

The proposed model we propose is adapted from the one proposed in (Hamdadou & Thérèse, 2011). It highlights all the elements of forest management, namely the management units (cantons), experts, and decision-maker on the one hand; and operations affecting its spaces on the other hand, by allowing them to express their characteristics and influences on each other.

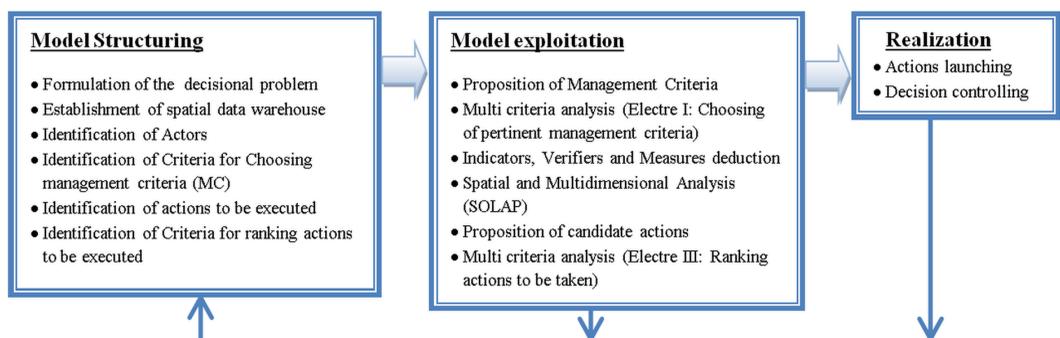
The proposed decision model is iterative and it is structured in three main phases (Pictet, 1996; Florent, 1997):

1. **Model Structuring:** In this phase, we define all the elements of the model; and also, we discuss the issues to keep in mind during the process of agroforestry management;
2. **Model exploitation:** In this phase, we bring together all the elements of the model for the various processes of the model;
3. **Realization:** This phase involves the implementation of the actions and control.

We note that we will not take into account the third phase which concerns the actions launching and the decision controlling.

The Figure 2 illustrates the different components of our decisional model.

Figure 2. Decisional model of “*Silvicultura*”



5.1. Model Structuring

In this section, we define different elements of the decision support model.

5.1.1. Formulation of the Decision Problem

The decision problem must be formulated in this step. Indeed, the question to be asked is the first component to be defined. In our study, the decision maker asks the question: “which actions to be executed and where?” This allows determination of intervenient actors and other components of the decision process.

5.1.2. Establishment of Spatial Data Warehouse

First, we must start with a preliminary step of the analysis of data sources and identifying user needs. Data extraction, transformation and loading process is then started. Also, logical and physical structure of the spatial data warehouse must be considered. (Kimball & Ross, 2003; Michael et al, 2008).

5.1.3. Identification of Actors

The identification of actors is one of the most important tasks when modeling a decision support system. Indeed, the absence of an actor will necessarily alter the result of the process that will generate inconsistent recommendations. We classify actors into three main classes:

1. **The Decision maker:** The main responsible for the management of forest areas;
2. **The Experts:** They are directly involved in the overall strategy of forests management;
3. **The Analyst:** He helps for implementing methods of decision support model.

5.1.4. Identification of Criteria for Choosing Management Criteria

This includes criteria on which a MCA will be applied to choose a reduced list of management criteria proposed by experts. Authors in (Guillermo & Phil, 2000) propose that management criteria evaluation must be ensured by MCA since it must cover every actor in forest management; and it must take into account quantitative and qualitative data.

5.1.5. Identification of Actions to be Executed

The list of actions is proposed by experts according to the expertise of results of spatial and multidimensional analysis with SOLAP. Actions are related to different axes namely: protection of fauna and flora species, regeneration after fire or overhunting, protection of wetlands, exploiting of forest resources, etc.

5.1.6. Identification of Criteria for Ranking Actions to be Executed

The list of actions must be subject of MCA for ranking according to a list of criteria related to execution time, amount, ecological interest, etc. and other parameters fixed by decision maker and experts.

5.2. Model Exploitation

Our approach focuses on achieving the following points.

5.2.1. Proposition of Management Criteria

In the current study, we consider that experts are responsible for providing the list of management criteria related to their goals to defend. Management criteria can be related to:

1. Legal affairs of all events that occur in agro-forestry areas;
2. Preserving wetlands, protected areas and natural species;

3. Protecting watersheds, maintaining the natural equilibrium, etc.;
4. The animation of mountain farming and other socioeconomic purpose.

5.2.2. The Used Multicriteria Analysis Methods

The multicriteria analysis is a tool for decision support in the event of a situation involving one or more possible actions. It allows:

1. To avoid conflicts by providing an acceptable basis for dialogue by all actors;
2. To help decision makers form and/or change their preferences, or to take a decision in accordance with their goals;
3. To reduce the complexity of multi-makers and multi action situations (Maystre et al., 1994).

ELECTRE methods have been developed by Bernard Roy in early 1970. He has initiated a variety of methods, called upgrade, based on comparisons of actions in pairs. They require little information to be implemented. This information is more readily available to the decision maker (Hamdadou, 2008).

According to the data provided and the model sensitivity to different parameters, different algorithms have been developed, creating different methods of ELECTRE.

ELECTRE I - Choosing of pertinent management criteria: It reveals the problem of choice, called alpha (α). It is to look in all the envisaged actions, a subset containing the “best” action or, the “most satisfactory” actions. (Roy, 1968) In this study, we use this method to choose a subset from all management criteria proposed by experts. This allows reducing their number by choosing the most satisfactory from them. Our choice to use this method is that ELECTRE I is very simple and the result contains only the best actions.

ELECTRE III - Ranking actions to be executed: It addresses the issue of ranking. However, ELECTRE III seeks not only two kinds of outranking, strong and weak, but examines a family that goes from totally hard to totally weak or nonexistent, through all the nuances that allow continuous scale between these two extremes (Roy, 1968; Gilliams et al., 2005).

The indifference threshold (q) and the preference threshold value (p) are defined for a certain criteria g as:

$$|g(a) - g(b)| \leq q \text{ (alternative } a \text{ is indifferent to alternative } b \text{ and } b \text{ to } a \text{ for criterion } g)$$
$$g(a) - g(b) > p \text{ (alternative } a \text{ is preferred to alternative } b \text{ for criterion } g)$$

A third important threshold value for this technique is the veto threshold (v) and is defined as:

$$g(b) > g(a) + v \text{ (alternative } a \text{ is not as good as alternative } b \text{ for criterion } g)$$

In our context, we use this method for ranking the list of actions proposed by experts according to a list of criteria.

The use of this method presents the following advantages:

- Mix Potential use of both quantitative and / or qualitative criteria;
- Acceptance and integration of the concept of incomparability of alternatives in the whole procedure of classification or ranking;
- Treatment of non-comparability, with two approximations, so as to focus on the alternatives that exhibit special characteristics;

- Simplicity of comparisons and consequentially understanding of the results;
- Exploitation of the subjective parameters (weight, indifference, preference and veto thresholds) attributing a high level of effectiveness to the decision support process.

5.2.3. Indicators and Verifiers Deduction

After applying MCA (ELECTRE I) for choosing the reduced list of management criteria, indicators are deducted. And then, verifiers are identified for every indicator. For these verifiers, related measures are identified.

5.2.4. Spatial and Multidimensional Analysis (SOLAP)

Verifiers are similar to measures in SOLAP context. So, for each measure, MDX queries are launched using SOLAP graphical interface for spatial and multidimensional analysis.

5.2.5. Proposition of Candidate Actions

The candidate actions are proposed by experts for solving eventual risk threatening the agroforestry ecosystems. They are related to the nature of the risk and available human, financial and natural resources. For example: definition plots for the use of wood at determined cantons; definition plots reserved for hunting at determined cantons, etc.

6. THE PROCESS ADOPTED BY “SILVICULTURA”

Each expert, along to its axis of interest, and basing on a set of parameters related to the management unit (canton), offers a set of management criteria.

After defining all management criteria, MCA ELECTRE I is then applied to choose a reduced list from those initially proposed by experts to be taken into consideration. This has for objective to choose the most important criteria in order to facilitate the next steps of the process.

After that, each management criteria from the final list has a set of indicators. From these indicators, they are deducted the verifiers which are similar to measures in SOLAP concept. They are subject of SOLAP analysis. The experts can then, determine the presence or not of natural hazards, basing on their knowledge. If there was no risk, a reformulation of the management criteria is launched to improve the current situation respecting the fundamental principles and the overall strategy of forest management. Here we note the iterative form of the process. Else, there is presence of risk identified by the experts. In this case, a set of actions will be proposed to remove or reduce the risk.

MCA (ELECTRE III) is applied to rank these actions, taking into account a list of criteria. The final list represents the result of the process, and will be subject of the realization phase. Iterative algorithm of the process can be written as shown in Algorithm 1.

7. UML MODELING FOR “SILVICULTURA”

The definition of the system components allows defining all interactions within “*Silvicultura*”. In fact, this can be easily modeled thanks to Unified Modeling Language (UML). This later has been widely used as a modeling language for designing various systems and software. In our approach, we have used it since it becomes a standard for modeling using object-oriented concept.

7.1. Activity Diagram of “Silvicultura”

The activity diagram is used for modeling the activity flow of “*Silvicultura*”. An activity diagram is drawn from a very high level. So, it gives high level view of the system. This high-level view is mainly

Algorithm 1. Silvicultura_process

```

Input: C = list of cantons
Output: Ac = list of actions
Begin
    Variable Risk: Boolean;
    MC, Cr, SP1, I, V, MDX, Results: list;
    {Management criteria, criteria for ELECTRE I, subjective parameters for ELECTRE I, indicators list, verifiers list,
    multidimensional expressions list}
    Get_cantons_characteristics ();
    Risk ← false;
    While (risk=false) do
    Begin
        MC ← Propose_management_criteria();
        Cr ← propose_criteria_for_choosing_MC();
        SP1 ← Propose_subjective_parameters();
        MC ← MCA1(); {MC list is reduced}
        I ← get_indicators(MC); {indicators are deducted}
        V ← get_verifiers(I); {verifiers are deducted}
        MDX ← deduct_MDX(V); {MDX list are deducted}
        Results ← SOLAP(MDX); {spatial_multidimensional analysis launching}
        Risk ← Expertise(results); {check risk presence}
        If (risk=true) then
        Begin
            Ac ← propose_list_actions(); {proposing actions}
            C ← propose_criteria_for_ranking_Ac();{propose criteria}
            SP2 ← Propose_subjective_parameters();{proposition subjective parameters for actions ranking }
            Ac ← MCA2(); {ranking the list of actions by ELECTRE III}
        End If
    End While
    ToPronounce (Ac);
    Launching (Ac);
    Controlling (Ac);
END Algorithm.
    
```

for business users who are not a technical person. This diagram is used to model the activities which are effectuated for business requirements. Figure 3 illustrates the Activity diagram of “Silvicultura”.

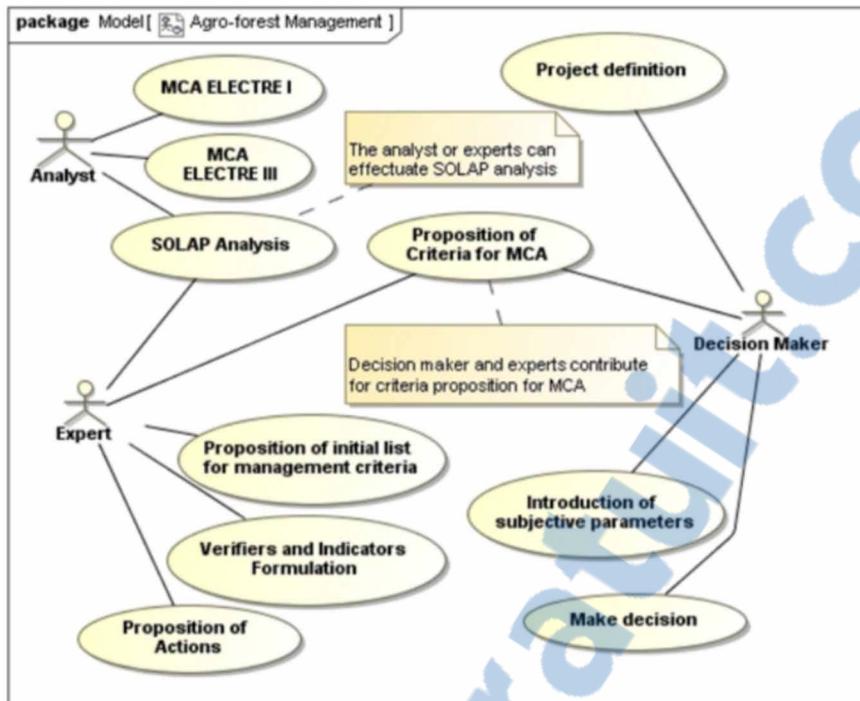
7.2. Use Cases Diagram of “Silvicultura”

The use cases diagram allows defining the actors’ interactions. It is a set of sequences of actions that are performed by the system actors. Figure 4 represents the uses case diagram of the decisional process of “Silvicultura”, in which we have considered three actors and ten use cases. The actors are: (a) *the decision maker* who achieves the *Project definition, introduces the subjective parameters, contributes with Experts to propose criteria for MCA and Makes decision*; (b) *Expert* who performs *the proposition of initial list of management criteria, Verifiers and Indicators Formulation, propose criteria for MCA with the Decision maker contribution, Proposes actions to be executed and effectuate SOLAP analysis*; (c) finally, the *Analyst* who effectuates *MCA Electre I, MCA Electre III and SOLAP analysis*.

7.3. Sequence Diagram of “Silvicultura”

The sequence diagram is a dynamic modeling technique. It models the flow of logic and focuses on identifying the interactions within the system. Figure 5 depicts the UML sequence diagram for “Silvicultura” approach where the interactions between the actors are shown. The diagram shows three actors decision-maker, expert and analyst. The vertical axis shows the messages, which are being sent from the sender to the receiver (Grady et al, 2005).

Figure 4. Use cases diagram of “Silvicultura”

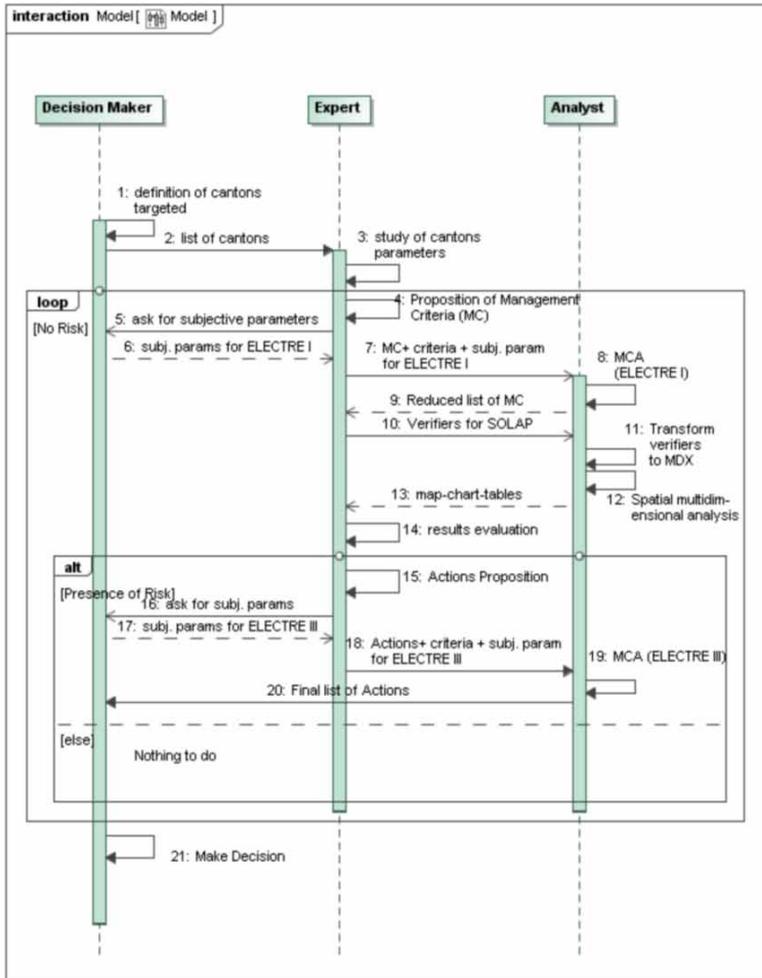


1. Unlike the traditional OLAP world, Unified Dimensional Model (UDM)² allows to have multiple fact tables within a cube (Sivakumar et al, 2009). So, basing on this principle, we propose that the cardinality in the composition relationship between hypercube package stereotype and the one of fact class must be “one-to-many” as shown in Figure 6;
2. In addition to the composition relationship between the hypercube and dimension stereotypes, we propose an association between fact and dimension stereotypes since a hypercube may contain several fact tables. This is mandatory to define which facts table concerns which dimensions in the same hypercube;
3. An association relationship is created between fact and *AggLevel* stereotypes to facilitate the dimensioning;
4. An association relationship is also created between measure and hierarchy stereotypes. This can define which aggregation function is applicable on which hierarchy (Nicolas et al., 2006).

To prevent incorrect modeling, authors in (Boulil et al., 2015) consider that a spatial hypercube must contain at least one spatial dimension or at least one spatial measure, however, in our proposal we define an OCL constraint in the context of the hypercube stereotype which considers that a spatial hypercube must contain at least one spatial dimension and at least one spatial measure in every fact table (see Figure 7). Indeed, in our proposal, we have to effectuate spatial analysis with every spatial measure, such as localization, buffer zone, etc.

For the UML profile implementation, we use MagicDraw³ the UML-based tool that enables the checking of OCL constraints of the profile. Indeed, constraints are checked by MagicDraw at the conceptual abstraction level which prevents incorrect modeling.

Figure 5. Sequence diagram of the decisional process



8.2. Constellation Schema for the Agroforestry SDW

SDW must provide answers for all queries of the experts. We must ensure the modeling of all the measures and dimensions required in the analysis process. Moreover, we must take into account the modeling of hierarchies which are used for data aggregations (Joel et al, 2010).

In order to select the dimensions and measures basing on the collected data, we must take into account the objectives of the analysis related to agroforestry management:

1. **Dimensions:** Facts are analyzed according to a set of dimensions describing three categories:
 - a. **Temporal Dimension:** To effectuate business analysis at the monthly, quarterly, or yearly level, a time dimension is required;
 - b. **Spatial Dimensions:** We use two spatial dimensions which are administrative and ecological;
 - c. **Thematic dimensions:** These dimensions are numerous and are related to business processes.

In Figure 8, dimensions and hierarchies are shown.

Figure 6. SDW profile stereotypes for “Silvicultura”

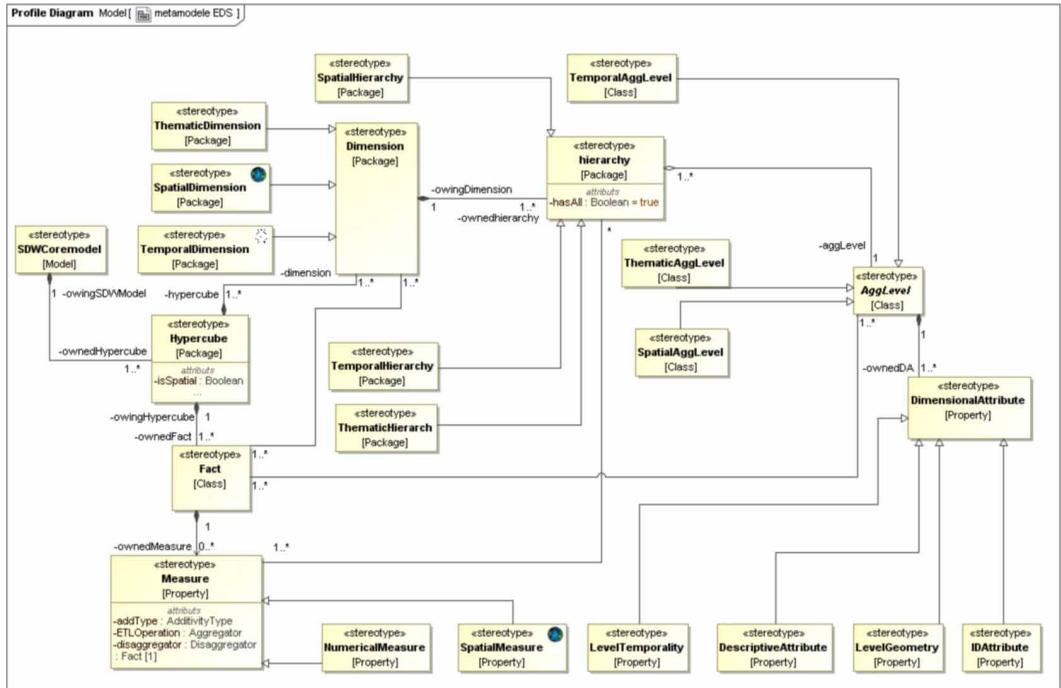


Figure 7. OCL constraint in the context of the hypercube stereotype

```

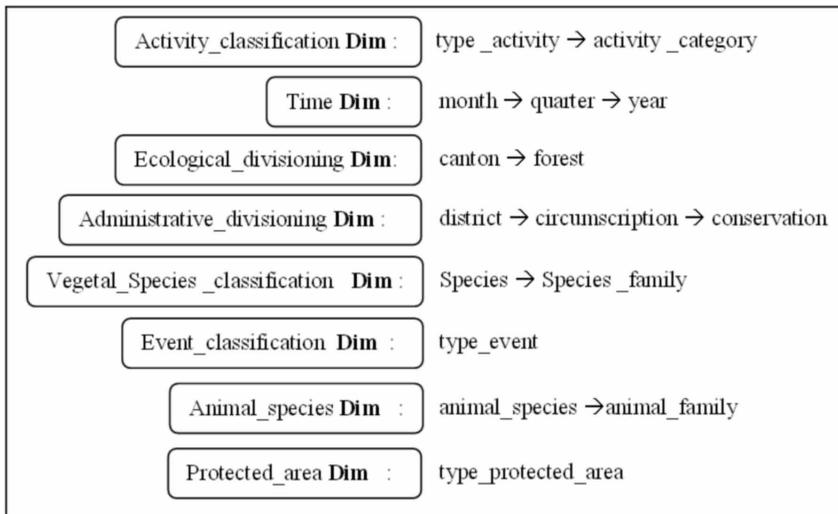
context Hypercube inv AtLeastOneSpatialDimensionAndOneSpatialMeasure:
isSpatial=true implies (
ownedMember->select (m|m.oclIsTypeOf(SpatialDimension)) ->size()>0
and ownedMember->exists (m|m.oclIsTypeOf(Fact) and
m.ownedAttribute->select (a|
a.oclIsTypeOf(SpatialMeasure)) ->size()>0))
    
```

The dimension “*Activity_classification*” records the classification of silvicultural activities at two levels of aggregations. These aggregation levels are formed by grouping dimensional attributes into disjoint subsets according to the analysis requirements (Boulil et al., 2015). For example, the spatial dimension “*Ecological_divisioning*” consists of one hierarchy which is composed of two aggregation levels, while the other spatial dimension “*Administrative_divisioning*” is composed of three levels.

2. **Measures:** They are similar to verifiers indicated above. They are obtained from queries to be formulated by the experts in relation to their management criteria. In our analysis context, measures can be numerical or spatial.

In our case study, experts are interested in the analysis of five aspects of their job, which are: silvicultural⁴ activities, events control, dendrometry⁵, wild life and protected areas. These aspects are defined using a hypercube with five fact tables sharing some dimension tables. In our proposal, the spatial hypercube is not linked to one activity but a set of activities. Indeed, we can store data

Figure 8. Dimensions and hierarchies



from multiple fact tables within the same cube (Sivakumar et al., 2009). This allows drawing the constellation schema.

The class diagram for the agroforestry SDW is shown in Figure 9. It is considered as spatial because it has at least one spatial measure in addition of spatial dimensions.

As shown in this Figure 9, “*silvicultural_act*” facts table is described using three measures where one is spatial (*localisation*) and two numerical (*volume_fin* and *superficie*); as well as “*control_event*” facts table. “*control_dendrometry*”’s facts are described using three numerical measures (*cir*, *haut*, and *age*), in addition of one spatial measure (*localise*). “*protected_area*” facts tables are described using one spatial measure (*localisation*). In addition, “*wild_life*” facts are described using two numerical measures and one spatial measure as depicted in Figure 9.

After we define the SDW model, we process to its implementation. We use Relational OLAP architecture composed of a relational database manager system (DBMS) which provides OLAP operators such as: *Roll up*, *Drill down*, *Slice*, *Dice*, etc.; and *GIS* software which allows applying spatial analysis.

The relational OLAP architecture is composed of four tiers: Extract-Transform-Load tier (ETL), data storage tier, OLAP server and OLAP client. However, the complex ETL techniques used in this phase are beyond the scope of this paper.

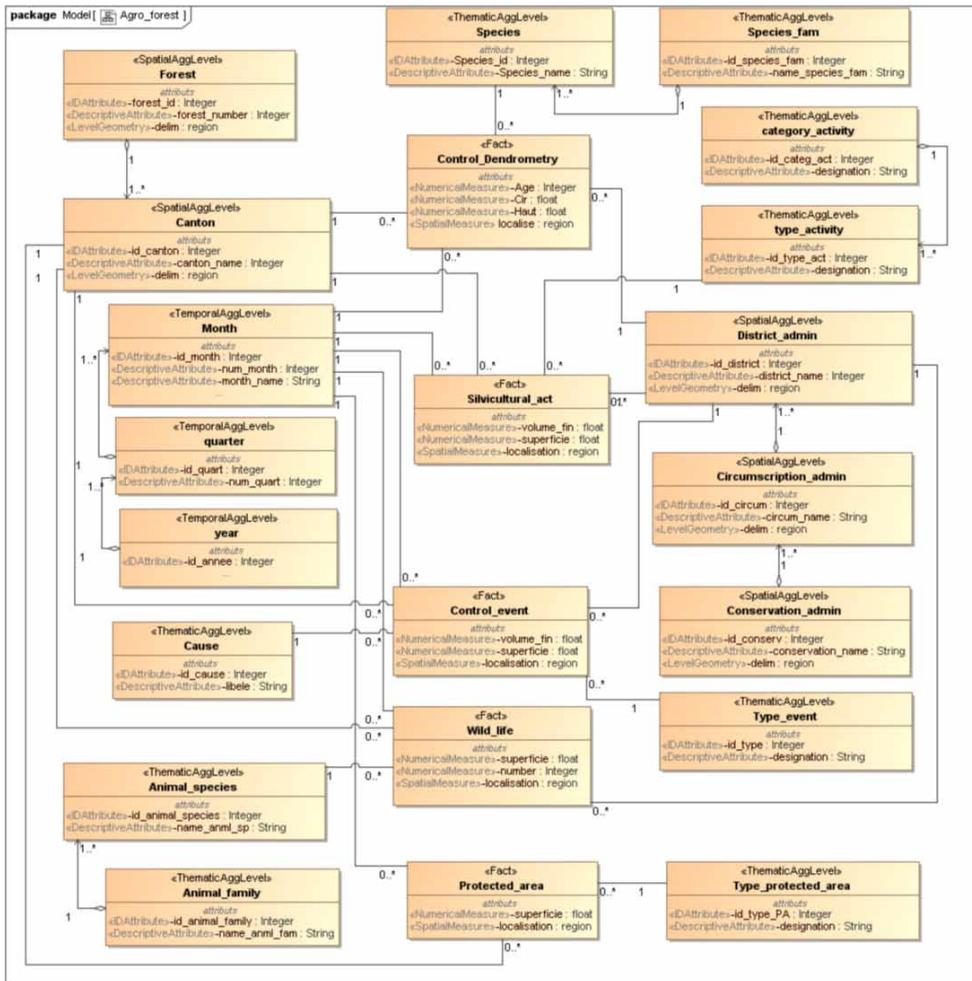
SOLAP client namely “*Sylvicole*” is developed with JAVA Eclipse, the object-oriented programming language. In fact, the solution provides cartographic synchronization with tabular and diagram displays as shown in Figure 11. The analyst and experts can use the SOLAP client for effectuating spatial and multidimensional analysis. The results allow dressing the list of the candidate actions, subject of the following phase of the decisional process.

9. CASE STUDY: RESULTS AND DISCUSSION

Periodically, the heads of the forest sector in Algeria carry out global studies of different activities and effectuate an inventory of natural resources. The purpose of these studies is to plan activities for future programs.

In the following we use the approach of “*Silvicultura*” our proposal, with a dataset concerning the Mediterranean forests of “*Mostaganem*” department situated in north western of Algeria. A part

Figure 9. Class diagram of Agro-forest SDW



of this dataset is property of forest administration of Mostaganem⁶ and some details are confidential. They are generally related to names of persons or other parameters which we did not get but we consider that they are not very interesting in the current study.

We must indicate that we have solicited the experience of forestry experts in different phase of this experimentation.

Firstly, we note that the decision maker asks the question: “what activities should be carried out through a set of management units (cantons)?”

Then, experts are consulted to provide all the technical characteristics of the cantons relating to pedology⁷, agricultural traditions, fauna and flora...etc. On the basis of these data, they propose, the set of management criteria (MC) specific to the targeted cantons and relating to their different axes of reflection.

After that, the analyst effectuates the MCA ELECTRE I for choosing a reduced list of management criteria (MC) to be taken in consideration. This analysis aims to consider all the experts requirements. The subjective parameters namely weights of the identified criteria, are provided by the decision maker.

9.1. MCA (ELECTRE I) for Choosing Management Criteria

In order to define the list of management criteria to be taken into account, we realize a MCA (ELECTRE I). The initial list of management criteria (MC) is initially as following:

- MC1:** Preservation of the existing patrimony.
- MC2:** Regeneration of the existing patrimony.
- MC3:** Development of the existing patrimony.
- MC4:** Rehabilitation of the natural environment.
- MC5:** Preservation of wetlands.
- MC6:** Development of wetlands.
- MC7:** Conservation of protected areas.
- MC8:** Consolidation of protected areas.
- MC9:** Preservation of the coastal fringe.
- MC10:** Maintaining the natural balance of ecosystems.
- MC11:** Protection of watersheds.
- MC12:** Animation of mountain farming.

The participation of expert in the decision-support process allow to define the following list of criteria ($Cr_i, i=1..12$) for analyzing the management criteria ($MC_p, i=1..5$):

- Cr1:** Protection of flora.
- Cr2:** Protection of fauna.
- Cr3:** Protection of wet areas.
- Cr4:** Extension in Forest.
- Cr5:** Socio-economic needs.

The weighs of these criteria may vary according to the iterative process of “*Silvicultura*”. We associate a weigh for each criterion. We propose a qualitative scale (important=3; middle=2and weak=1). So weights for ELECTRE I, are as follows: $Cr_1=3$; $Cr_2=3$; $Cr_3=2$; $Cr_4=1$; $Cr_5=1$.

The generation of performance matrix is made in collaboration of all the experts involved in the definition of the initial list of criteria. We applied an application developed in our research team at the “LIO” laboratory for obtaining the results of ELECTRE I process. In Figure 10 are presented the Performances, Concordances and discordances matrix, and finally the solution.

Using these parameters, the solution is: MC_4, MC_7, MC_{10} and MC_{12} . We note that the choice of MCA parameters affects this result. This solution signifies that we will be interested by these management criteria. As noted in Figure 10, these criteria will be transformed into indicators and later into verifiers.

In Table 1, are depicted some examples of the transformation of management criteria (MC) into indicators and then verifiers.

9.2. Spatial On-Line Analytical Processing

Spatial and multidimensional analysis is the core of such a system. Indeed, thanks to the spatial-OLAP application “*sylvicole*” developed for this aim, the analyst can launch analytical queries (MDX). These queries are automated in a graphical interface. This earlier allows visualizing results in tabular form, maps or charts.

In Figure 11, an example of visualization with “*sylvicole*” application. By exploiting the fact table “*Control dendrometry*”, we can see age (“*age*” measure) of forest species (*thematic dimension*) dominant in the forests of “*Mostaganem*” by “*Canton*” (*spatial dimension*). We can see the spatial measure “*delimitation*” that displays corresponding dendrometric plots.

Figure 10. Result of ELECTRE I

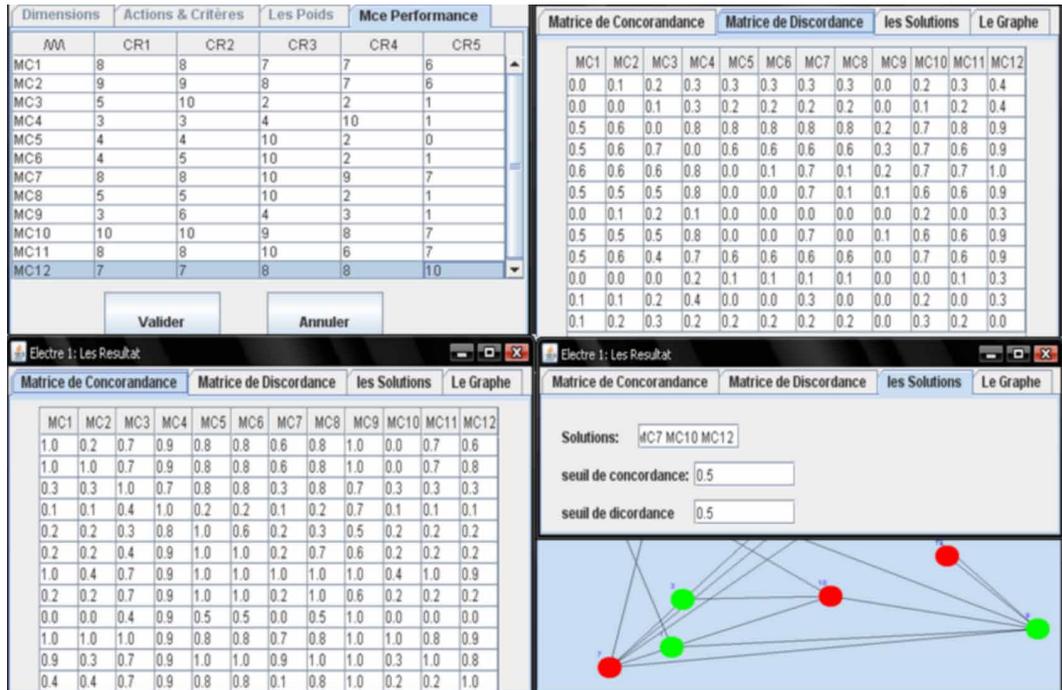


Table 1. Transformation management-criteria, indicators, verifiers

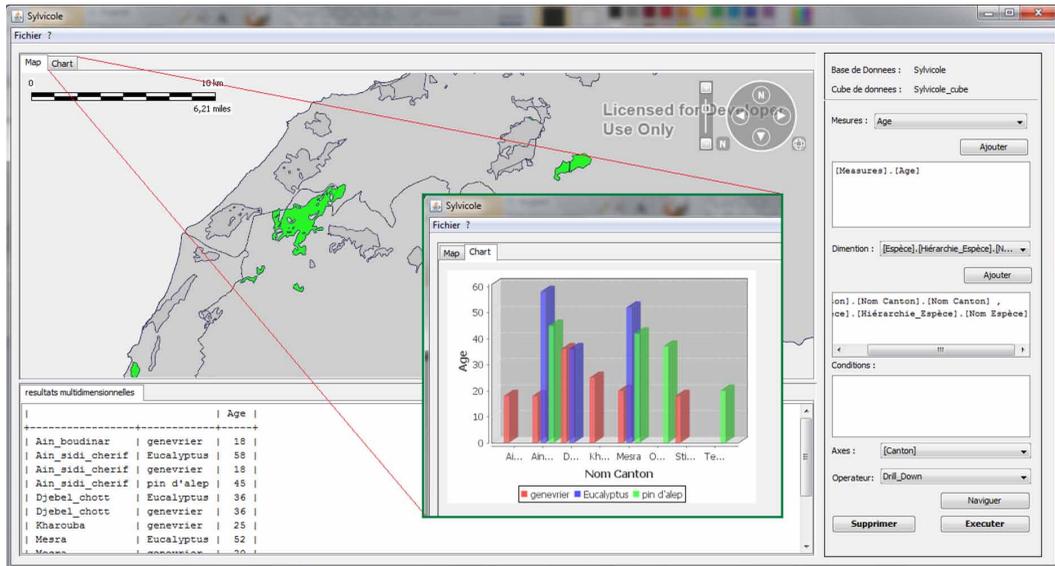
Management Criteria	Indicators	Verifiers
MC 4	<ul style="list-style-type: none"> Investigation the rehabilitation activities in the forest 	<ul style="list-style-type: none"> To measure the financial volume of rehabilitation activities performed in the forest To measure the area of rehabilitation activities performed in the forest
MC 7	<ul style="list-style-type: none"> Investigating the situation of protected areas 	<ul style="list-style-type: none"> To identify protected areas To count infractions in protected areas
MC 10	<ul style="list-style-type: none"> Investigating the state of the fauna and flora species 	<ul style="list-style-type: none"> To identify the overhunting areas To make census of the flora species area and to measure dendrometry
MC 12	<ul style="list-style-type: none"> Investigating the situation of mountains agriculture 	<ul style="list-style-type: none"> To make census of exploiting forest resources authorizations To identify areas of exploiting forest resources authorizations

In order to lunch an operation of “Regeneration”, such an analysis can examine what are the groves containing the oldest species and this for each management unit.

In the example shown in Figure 11, we can easily see that the “Eucalyptus” in “Ain-Sidi-Cherif” and “Mesra” are older, followed by the “pin d’Alep” (Aleppo pine) in two cantons implying a need for “Regeneration”. On the other hand, the absence of this species is remarkable in the cantons “Ain-Boudinar”, “Kharouba” and “Stidia.” This may qualify a “Reforestation” of this species as a candidate action in the third phase of the decision support process.

Similarly, a second subject of analysis can be started with the fact table “Control_event”. Indeed, analyzes of measures: “financial-volume”, “area” and “localization” can be done in different

Figure 11. Sample visualization with the SOLAP client “Sylvicole”



dimensions. This lets we know what kind of event happened to any site, at what time and what type (forest fires, illegal wood logging, attack by insects or parasites ... etc.). The result of these analyzes offer some operations such as: reforestation, opening of firewall trench, intensify control brigade, etc. according to soil type, distance, elevation, climate ... etc. These operations are proposed by experts and qualified for the third phase of the decision support process.

The third fact table “*silvicultural work*” can also be used to perform multidimensional analysis of its measures according to different thematic, temporal and spatial dimensions. Indeed, through the application SOLAP for example we can know which parts of the forest and since when were happened “timber harvesting”? This allows knowing if there is need to launch such an operation. This allows the renewal of the forest on one hand, and following an agro-forest policy, allows neighbors of the forest to benefit from natural resources on the other hand.

Measures of the fourth fact table “*Wild_life*”, can be analyzed in different dimensions. We can know if the number of some faunal species in a region represents a danger to the plant species. In this case, to keep the natural balance, experts can propose operations like “hunting” or “encourage more natural enemies”.

The fifth subject of analysis can be started analyzing the measures of the last fact table “*protected areas*”. The multidimensional analysis of these measures in different dimensions allows controlling the ecological changes in these areas. For example, we can know the evolution of the area (numerical measurement) if there is an expansion or an annual degradation. Spatial analysis of the measure “localisation” can examine the presence of the road infrastructure or urban expansion nearby. This can be considered as a threat. Experts can offer “gabionade operations” or “production of fencing” to protect these areas.

9.3. MCA (ELECTRE III) for Ranking Actions to be Executed

The list of actions regarding the MCA is not known in advance. Indeed, it will be fixed by the actors (experts) only after SOLAP analysis. Basing on the results of SOLAP queries, the initial list of actions is the following:

- Ac1:** Construction lookout stations at high points with defined coordinates.
- Ac2:** Realization of water points at defined coordinates.
- Ac3:** Diffusion brigades control at sites well located.
- Ac4:** Launching silvicultural activities (tending, sanitation, etc.).
- Ac5:** Launching a reforestation project to well-defined plots.
- Ac6:** Launching of development or opening forest tracks.
- Ac7:** Launch of development work or trenching firewall.

These actions are evaluated taking into account this list of criteria ($C_i, i=1..5$):

- C1:** Execution time
- C2:** Amount
- C3:** Workforce
- C4:** Ecological interest
- C5:** Generation of industrial products

Expert can then, evaluate actions through different criteria. Also, Weights, Preference sense, Indifference, Preference and Veto indexes are given as shown in Table 2.

After applying an application developed in our research team, the results can be illustrated in two phases namely aggregation and exploitation.

9.4. Aggregation

Figure 12 shows results of aggregation. It presents the Global Concordance, Discordances and Credibility Matrix.

9.5. Exploitation

The second phase is exploiting results for ranking.

Figure 13 shows results of Exploitation. It presents the Distillation process and the final ranking.

Table 2. Performances and parameters for ELECTRE III

Actions	Criteria				
	C1	C2	C3	C4	C5
Ac1	1	2	2	8	0
Ac2	2	3	3	8	0
Ac3	1	2	1	7	0
Ac4	7	5	5	7	6
Ac5	8	9	9	10	0
Ac6	8	8	9	7	7
Ac7	7	7	9	7	9
Weight	2	2	2	3	1
Preference sens	↓	↓	↑	↑	↑
Indifference	200	100	1	2	3
Preference	400	200	2	2.8	6
Veto	800	400	4	5	9

Figure 12. Aggregation results of ELECTRE III

The screenshot shows three windows from the 'Resultats Agregation' application. Each window has tabs for 'Matrices Concordances', 'Concordance Globale', 'Matrices Discordances', and 'Matrice Crédibilité'. The 'Concordance Globale' window is selected in all three.

Window 1 (Top): Shows a 7x7 matrix of concordance values for actions 1 through 7.

	Action 1	Action 2	Action 3	Action 4	Action 5	Action 6	Action 7
Action 1	1,0000	1,0000	1,0000	0,7000	0,8000	0,7000	0,7000
Action 2	1,0000	1,0000	1,0000	0,7000	0,8000	0,7000	0,7000
Action 3	1,0000	0,8000	1,0000	0,7000	0,5000	0,7000	0,7000
Action 4	1,0000	1,0000	1,0000	1,0000	0,5000	0,8000	0,8000
Action 5	1,0000	1,0000	1,0000	0,9000	1,0000	0,9000	0,9000
Action 6	1,0000	1,0000	1,0000	1,0000	0,7000	1,0000	1,0000
Action 7	1,0000	1,0000	1,0000	1,0000	0,7000	1,0000	1,0000

Window 2 (Middle): Shows a 7x7 matrix of discordance values for actions 1 through 7. A dropdown menu on the left is set to 'Cr3'.

	Action 1	Action 2	Action 3	Action 4	Action 5	Action 6	Action 7
Action 1	0,0000	0,0000	0,0000	0,5000	1,0000	1,0000	1,0000
Action 2	0,0000	0,0000	0,0000	0,0000	1,0000	1,0000	1,0000
Action 3	0,0000	0,0000	0,0000	1,0000	1,0000	1,0000	1,0000
Action 4	0,0000	0,0000	0,0000	0,0000	1,0000	1,0000	1,0000
Action 5	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Action 6	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Action 7	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Window 3 (Bottom): Shows a 7x7 matrix of credibility values for actions 1 through 7.

	Action 1	Action 2	Action 3	Action 4	Action 5	Action 6	Action 7
Action 1	1,0000	1,0000	1,0000	0,7000	0,8000	0,7000	0,0000
Action 2	1,0000	1,0000	1,0000	0,7000	0,8000	0,7000	0,0000
Action 3	1,0000	0,8000	1,0000	0,7000	0,5000	0,7000	0,0000
Action 4	1,0000	1,0000	1,0000	1,0000	0,5000	0,8000	0,8000
Action 5	1,0000	1,0000	1,0000	0,9000	1,0000	0,9000	0,0000
Action 6	1,0000	1,0000	1,0000	1,0000	0,7000	1,0000	1,0000
Action 7	1,0000	1,0000	1,0000	1,0000	0,7000	1,0000	1,0000

Figure 13. Exploitation results of ELECTRE III

The screenshot shows the 'Fichier' application with tabs for 'Matrice de crédibilité', 'Distillation', 'Rangement final', and 'Etapas'. The 'Rangement final' tab is selected.

Ascendante (Ascending):

RANG	ACTION
7	Action 3
5	Action 1
5	Action 2
4	Action 4
3	Action 6
2	Action 5
1	Action 7

Descendante (Descending):

RANG	ACTION
1	Action 5
3	Action 6
3	Action 7
4	Action 4
5	Action 2
7	Action 1
7	Action 3

Final Classification (Classement):

- Classement 1 ==> Action 5 ; Action 7 ;
- Classement 2 ==> Action 6 ;
- Classement 3 ==> Action 4 ;
- Classement 4 ==> Action 2 ;
- Classement 5 ==> Action 1 ;
- Classement 6 ==> Action 3 ;

The final ranking of actions to be executed represents the final step of the “*Silvicultura*” process. We note that realization and control of this solution represent the following phases. However, this exceeds the limits of the current study.

10. CONCLUSION AND FUTURE WORK

SOLAP combines OLAP analysis with cartographic visualizations of GIS systems, allowing understanding of the geographical distributions of analytical data. SOLAP extends OLAP with spatial OLAP operators (e.g., Spatial Roll-up) to allow interactive spatial multidimensional explorations of analytical data.

However, SOLAP technology alone cannot solve some situation related to spatial problems. The solution, in our proposal, is to combine this technology with the multi criteria analysis (MCA) tools. Indeed, the MCA offers several advantages in terms of decision-making when we must take into account conflicting interests.

The main objective of this study is to develop a Multicriteria Spatial Decision Support System (SDSS) namely “*Silvicultura*” by integrating a multicriteria analysis engine and SOLAP. This integration can help decision makers for effective management of agroforestry areas and their activities.

“*Silvicultura*” is composed of four subsystems, namely Database subsystem, Monitoring subsystem, Problem Process subsystem and User Interface Subsystem. These subsystems are interconnected by data and queries flow for making decision to be announced in the user interface. We have based our proposal in UML for two main reasons: (1) UML is a well-known standard modeling language known by most database designers, and (2) UML can be easily extended so that it can be adapted for a specific domain such as the multidimensional modeling for spatial data warehouses. Our decisional model is implemented using Relational OLAP architecture composed of a relational database manager system (DBMS) that can implement complex indicators by means of MultiDimensional eXpression (MDX) in one hand. In addition, GIS software allows spatial analysis and spatial data exploration, on the other hand. The user interface of SOLAP is developed by Java Eclipse programming language. Our proposal is tested with data related to the agroforestry area in *Mostaganem* (Algeria), with consideration of the current management policy in this country.

Our future works concern the extension of “*Silvicultura*” by including spatial knowledge database system. It will be an intelligent mechanism with the ability to link spatial data warehouse with the monitoring subsystem, in one hand.

On the other hand, In the Database Subsystem, the SDW will be enhanced by enriching and optimizing our UML profile by introducing spatial integrity constraints (IC) and the integration of mechanisms checking data quality in heterogeneous spatial data cubes. In addition, we think for implementing other MCA methods that can ameliorate or solve other spatial conflict situation in a decisional process.

REFERENCES

- Acosta, M., & Corral, S. (2015). Participatory Multi-Criteria Assessment of Forest Planning Policies in Conflicting Situations: The Case of Tenerife. *Forests*, 6(12), 3946–3969. doi:10.3390/f6113946
- Ananda, J. (2007). Implementing participatory decision making in forest planning. *Environmental Management*, 39(4), 534–544. doi:10.1007/s00267-006-0031-2 PMID:17318695
- Ananda, J., & Herath, G. (2009). A critical review of multi-criteria decision making methods with special reference to forest management and planning. *Ecological Economics*, 68(10), 2535–2548. doi:10.1016/j.ecolecon.2009.05.010
- Bakillah, M., Mostafavi, M. A., & Bédard, Y. (2006, October). Développement d'une approche géosémantique intégrée pour ajuster les résultats des requêtes spatiotemporelles dans le domaine forestier. *Paper presented at the Géomatique – Au cœur des processus colloquy*, Montreal, Canada.
- Belton, V., & Stewart, T. J. (2002). *Multiple Criteria Decision Analysis - An Integrated Approach*. Boston: Kluwer Academic Publishers. doi:10.1007/978-1-4615-1495-4
- Bentrup, G., & Leininger, T. (2002). Agroforestry mapping: The way GIS. *Journal of Soil and Water Conservation*, 57, 148–153.
- Bouilil, K., Florence, L. B., Sandro, B., Corinne, G., & Flavie, C. (2014). Multidimensional modeling and analysis of large and complex watercourse data: An OLAP-based solution. *Ecological Informatics*, 24, 90–106. doi:10.1016/j.ecoinf.2014.07.001
- Bouilil, K., Sandro, B., & Francois, P. (2015). Conceptual model for spatial data cubes: A UML profile and its automatic implementation. *Computer Standards & Interfaces*, 38, 113–132. doi:10.1016/j.csi.2014.06.004
- De Steiguer, J. E., Liberti, L., Schuler, A., & Hansen, B. (2003). *Multi-Criteria Decision Models for Forestry and Natural Resources Management: An Annotated Bibliography*. United States Department of Agriculture, Forest Service, Northeastern Forest Experiment Station, Upper Darby, PA, USA.
- DeMers, M. N. (1997). *Fundamental of Geographic Information Systems*. New York: Wiley.
- Díaz-Balteiro, L., & Romero, C. (2008). Making forestry decisions with multiple criteria: A review and an assessment. *Forest Ecology and Management*, 255(8-9), 3222–3241. doi:10.1016/j.foreco.2008.01.038
- Florent, J. (1997). *Décider sur le territoire, proposition d'une approche par utilisation de SIG et de méthode d'analyse multicritère* [Doctoral dissertation]. Federal Polytechnic School of Lausanne, Switzerland.
- Fontana, V., Radtke, A., Fedrigotti, V. B., Tappeiner, U., Tasser, E., Zerbe, S., & Buchholz, T. (2013). Comparing land-use alternatives: Using the ecosystem services concept to define a multi-criteria decision analysis. *Ecological Economics*, 93, 128–136. doi:10.1016/j.ecolecon.2013.05.007
- Gilliams, S., Raymaekers, D., Muys, B., & Orshoven, J. V. (2005, October). Comparing multiple criteria decision methods to extend a geographical information system on afforestation. *Computers and Electronics in Agriculture*, 49(1), 142–158. doi:10.1016/j.compag.2005.02.011
- Grady, B., James, R., & Ivar, J. (2005). *The Unified Modeling Language: reference manual*, (2nd ed.). USA: Addison-Wesley Professional press.
- Greene, R., Luther, J. E., Devillers, R., & Eddy, B. (2010). An approach to GIS-based multiple criteria decision analysis that integrates exploration and evaluation phases: Case study in a forest-dominated landscape. *Forest Ecology and Management*, 260(12), 2102–2114. doi:10.1016/j.foreco.2010.08.052
- Guillermo, A. M., & Phil, M. (2000). *Application de l'analyse multicritère à l'évaluation des critères et indicateurs- Manuels de critères et indicateurs pour la gestion durable des forêts (Tech. Rep.) (French vers.)*. Montpellier, France: CIFOR.
- Guimond, L. E. (2005, January). *Conception d'un environnement de découverte des besoins pour le développement de solutions solap* [Master's thesis]. Master of Science, Laval University, Canada.

Hamdadou, D. (2008, June). Un Modèle de Prise de Décision pour l'Aménagement de Territoire, Une Approche Multicritère et une Approche de Négociation [Doctoral dissertation]. University of Oran I, Algeria.

Hamdadou, D., & Thérèse, L. (2011). A Multicriteria Group Decision Support System for Industrial Diagnosis, INFOCOMP. *Journal of Computational Science*, 10(3), 12–24.

Joel, D. S., Anjolina, G. O., Robson, N. F., Ana, C. S., & Valéria, C. T. (2010). Modeling and querying geographical data warehouses. *Information Systems*, 35(5), 592–614. doi:10.1016/j.is.2009.10.005

Juan, T., Emilio, S., Eduardo, F. M., & Mario, P. (2009). A UML 2.0 profile to define security requirements for Data Warehouses. *Computer Standards & Interfaces*, 31(5), 969–983. doi:10.1016/j.csi.2008.09.040

Julien, P. (2004, September). Conception d'une légende interactive et forable pour le SOLAP. Master's thesis presented at high studies faculty [Masters]. Master of Science, Laval University, Quebec, Canada.

Kilgour, D. M., & Eden, C. (2010). *Handbook of Group Decision and Negotiation* (Vol. 4). New York, NY, USA: Springer Science & Business Media. doi:10.1007/978-90-481-9097-3

Kimball, R., & Ross, M. (2003). *Entrepôts de données, Guide pratique de modélisation dimensionnelle*. Paris, France: Vuibert Informatique Press.

Maryvonne, M., Bédard, Y. & Alexandre, B. (2002). Conception d'entrepôts de données géospatiales à partir de sources hétérogènes, exemple d'application en foresterie. *Ingénierie des Systèmes d'information*, 7(3), 89–111.

Maystre, L. Y., Pictet, J., & Simos, J. (1994). *Méthodes multicritères Electre*. Lausanne, Switzerland: Romandes Polytechnics and Universities Press.

Michael, M. G., Aryya, G., Anita, K., & Christopher, S. (2008). A user-centered design for a spatial data warehouse for data exploration in environmental research. *Ecological Informatics*, 3(4-5), 273–285. doi:10.1016/j.ecoinf.2008.08.002

Montserrat, A. & Serafín, C. (2017) Multicriteria Decision Analysis and Participatory Decision Support Systems in Forest Management. *Forests review*, 8, 116

Nicolas, P., Jacky, A., & Isabelle, C. W. (2006). A UML-based data warehouse design method. *Decision Support Systems*, 42(3), 1449–1473. doi:10.1016/j.dss.2005.12.001

Nordström, E., Eriksson, L. O., & Öhman, K. (2010). Integrating multiple criteria decision analysis in participatory forest planning: Experience from a case study in northern Sweden. *Forest Policy and Economics*, 12(8), 562–574. doi:10.1016/j.forpol.2010.07.006

Octavio, G., Mazón, J. N., Irene, G., & Trujillo, J. (2012). A personalization process for spatial data warehouse development. *Decision Support Systems*, 52(4), 884–898. doi:10.1016/j.dss.2011.11.010

Pictet, J. (1996). *Dépasser l'évaluation environnementale. Procédure d'étude et insertion dans la décision globale*. Lausanne, Switzerland: Romandes Polytechnics and Universities Press.

Rivest, S., Bédard, Y., Proulx, M. J., & Nadeau, M. (2003, October 2-3). SOLAP: a new type of user interface to support spatio-temporal multidimensional data exploration and analysis. In *Workshop ISPRS*, Québec, Canada.

Roy, B. (1968). Classement et choix en présence de points de vue multiples (la méthode ELECTRE). *French review of Computer Science and Operational Research*, 2(V1), 57–75.

Roy, B. (1996). *Multicriteria Methodology for Decision Aiding*. Dordrecht, Netherlands: Kluwer Academic Press. doi:10.1007/978-1-4757-2500-1

Salehi, M., Bédard, Y., & Rivest, S. (2010). Formal conceptual model and definition framework for spatial data cubes. *Geomatica*, 64, 313–326.

Sboui, T., Salehi, M., Bédard, Y., & Rivest, S. (2008 January 29). Catégorisation des problèmes d'intégration des modèles des cubes de données spatiales. *Paper presented at the meeting of 8emes Journées Francophones Extraction et Gestion des Connaissances*, Sophia Antipolis.

Sivakumar, H., Matt, C., Sethu, M., Robert, Z., & Denny, G. Y. L. (2009). *Professional Microsoft SQL Server Analysis Services 2008 with MDX*. Indianapolis, Indiana, USA: Wiley Publishing Inc Press.

Taher O.A., & Maryvonne, M. (2005). Multidimensional Structures Dedicated to Continuous Spatiotemporal Phenomena. In *British National Conference on Databases, LNCS (Vol. 3567, pp. 29-40)*. Springer-Verlag.

Zhang, Z., Sherman, R., Yang, Z., Wu, R., Wang, W., Yin, M., & Ou, X. et al. (2013). Integrating a participatory process with a GIS-based multi-criteria decision analysis for protected area zoning in China. *Journal for Nature Conservation, 21(4)*, 225–240. doi:10.1016/j.jnc.2012.12.006

ENDNOTES

- 1 Law No. 84-12 of 23 June 1984, as amended and supplemented.
- 2 The model for data in a Microsoft SQL Server Analysis Services (SSAS).
- 3 <http://www.magicdraw.com>
- 4 The branch of forestry dealing with the development and care of forests.
- 5 The branch of botany that is concerned with the measurement of the various dimensions of trees, such as their diameter, size, shape, age, overall volume, thickness of the bark, etc.
- 6 <http://www.dgf.org.dz/fr/structure/conservation-des-forêts-de-la-wilaya-de-mostaganem>
- 7 Study of soil, in particular its chemical, physical and biological characteristics

Charef Abdallah Bensalloua specialises in spatial techniques, geographical information systems (GIS), remote sensing and geo-decisional systems. He is a computer science professor at the University of Mostaganem in Algeria.

Djamila Hamdadou specialises in decision support systems, multicriteria analysis and collaborative and spatial decisional systems. She leads a research team at the LIO laboratory and is a computer science professor at the University of Oran1 in Algeria.

Users Integrity Constraints in SOLAP Systems. Application in Agroforestry

Abdallah Bensalloua Charef*, Hamdadou Djamilia

Computer Science Laboratory of Oran (LIO) - University of Oran 1 Ahmed Benbella, Oran (Algeria)

Received 6 August 2017 | Accepted 16 December 2017 | Published 26 January 2018



ABSTRACT

SpatialData Warehouse and Spatial On-Line Analytical Processing are decision support technologies which offer the spatial and multidimensional analysis of data stored in multidimensional structure. They are aimed also at supporting geographic knowledge discovery to help decision-maker in his job related to make the appropriate decision. However, if we don't consider data quality in the spatial hypercubes and how it is explored, it may provide unreliable results. In this paper, we propose a system for the implementation of user integrity constraints in SOLAP namely "UIC-SOLAP". It corresponds to a methodology for guaranteeing results quality in an analytical process effectuated by different users exploiting several facts tables within the same hypercube. We integrate users Integrity Constraints (IC) by specifying visualization ICs according to their preferences and we define inter-facts ICs in this case. In order to validate our proposition, we propose the multidimensional modeling by UML profile to support constellation schema of a hypercube with several fact tables related to subjects of analysis in forestry management. Then, we propose implementation of some ICs related to users of such a system.

KEYWORDS

Agroforestry, Integrity Constraints, SOLAP, Spatial Data Warehouse, Spatial Decision Support System, Spatial Multidimensional Analysis, UML Multidimensional Modeling.

DOI: 10.9781/ijimai.2018.01.003

I. INTRODUCTION

Spatial On-Line Analytical Processing (SOLAP) and Spatial Data Warehouses (SDWs) are decision support technologies that allow spatial and multidimensional analysis of multisource data. Basing on the multidimensional model, these technologies allow a multidimensional representation of data in hypercubes (datacubes). The hypercubes are essentially based on facts and dimensions concepts. The facts tables are analysis subjects which are described by numerical, textual or spatial measures. The dimensions define the analyzing axis. They can be organized in hierarchies of several levels. This allows the visualization of the measures at different levels of detail.

The data quality in the spatial hypercubes is important since this data is used as a basis for decision-making. In fact, poor data quality could lead to poor decision making. Without considering data quality in spatial hypercubes, it may provide unreliable results. As a result, integrity constraints (ICs) become very important to improve the logical consistency of any database, which increase data quality.

In transactional spatial databases, spatial integrity constraints are defined along the database conceptual models to preserve spatial data quality [1], [2]. However, in analytical process other kind of integrity constraints must be considered for maintaining consistent analysis using the hypercube. It is related to the manner of navigating and querying data with consideration of aggregation rules, hierarchies and other concepts of the multidimensional models.

In the database community, some research works study integrity

constraints for non-spatial datacubes [3], [4]. However, for studying integrity constraints for spatial datacube, its specific characteristic related especially to spatial data features should be considered. Therefore, the objective of this paper is to propose a system for the implementation of user integrity constraints in SOLAP namely "UIC-SOLAP". It takes into account fundamental considerations about users querying and aggregating in spatial OLAP. In addition, other data integrity constraints are implemented in order to obtain efficient analysis within the decision-making process.

Section 2 reviews a state of the art related to quality management in SDW and SOLAP in one hand. In the other hand, some researches in the field of forestry management are cited. In Section 3 our contribution is stated. Section 4 discusses the integrity constraints from spatial databases to spatial datacubes. In Section 5 the proposed system namely "UIC-SOLAP" for user integrity constraints in SOLAP is depicted while the process adopted by it, is presented in Section 6. Section 7 explains the case study related to an implementation of Spatial Data Warehouse for forest management. In Section 8, obtained results are discussed related to the application of the "UIC-SOLAP" system with SOLAP prototype developed for forestry management. Finally, Section 9 concludes and draws more research directions for the implementation of the integrity constraints and data quality in spatial hypercubes.

II. STATE OF THE ART

Quality management in SDW and SOLAP systems is an important research issue. Indeed, precision in SDW was addressed in [5] and [6] which provide logic models and an indexing technique for storing and querying vague spatial data.

* Corresponding author.

E-mail address: charef_bensalloua@yahoo.fr

In [7] authors have studied the problem of completeness in classical DW. They proposed solution taking into account missing values in hierarchies. Inconsistencies which refer to the existence of logical contradictions in SDW can be also controlled by ICs. As mentioned in [8], ICs expression on conceptual models is essential for taking into account all quality rules. For this aim, [9] proposes ad-hoc multidimensional conceptual models allowing the expression of some data ICs using logic predicates. In [10], authors propose an extension of the Entity-Relationship (ER) model for the design of spatio-temporal data warehouses. They define a set of ad hoc pictograms to express ICs on spatial data (topological relations between spatial members). Authors in [11] propose a UML profile for DWs, but they consider only a very small number of data ICs. Aggregation problems are presented in [12] by defining simple schema constraints with UML multiplicities. In [13], the author concludes that several languages for the specification of ICs in spatial databases have been proposed but they are not efficient for defining ICs in spatial hypercubes. Indeed, the author presents a formal language for the specification of integrity constraints in the conceptual model of spatial hypercubes. This language is based on controlled natural languages as well as natural hybrid languages with pictograms. In [14], authors study complex structural aggregation constraints. Basing on a UML(Unified Modeling Language) model, [8] shows that the Spatial OCL (Object Constraints Language) which is an OCL extension for spatial data, allows the definition of a large number of ICs on spatial data. Data modeling languages can express only a very limited number of integrity constraints as for example the cardinalities of the relations between classes of objects. For the expression of IC, Many works use non-standard languages, namely logical, natural, visual or hybrid. Others prefer OCL because it integrates easily with UML. Therefore, it is necessary to use languages dedicated for the ICs specification that are interoperable with the language of definition of the data structures. In [15], the author presents a method for the conceptual modeling of the SOLAP ICs after having introduced the main concepts on which this method is based namely the UML profiles and Spatial OCL.

In the field of natural resources specifically in the forest management and preserving, researches published are relevant for consideration in forest environments as well as in social and economic anxieties. In [16], the author implements a participatory decision making in forest planning. Also, authors in [17], propose a Decision Support for forest management. Authors in [18] treat the problem of forest land use policies. In [19], authors integrate multiple criteria decision analysis in participatory forest planning. Authors in [20] propose a participatory multi-criteria assessment of forest planning policies in conflicting situations. An empirical study on voting power in participatory forest planning is proposed in [21]. For forest protection, authors in [22], propose software for forest fire detection and extinction.

III. CONTRIBUTION

In forestry, we notice the complexity of effective management of natural areas for ecological, economic and social purposes. Indeed, the use of SOLAP application for decision-making allows analyzing this domain in various axes of reflection. However, different kinds of heterogeneity may exist between the data sources (remote sensing, GIS applications ... etc.). As a result, several problems of inconsistency may emerge. In most studies, the discussed problems are related to the heterogeneity of manipulated data. So, few studies have addressed issues related to the end users when decision-making is related to many sub-domains.

We notice also, that, the same territory is usually, under the control of several actors with different anxieties namely territorial

management, exploiting of forest products and even local collectivities with social anxieties. However, the use of SOLAP tool by different actors independently allows them to extract decisional indicators which are heterogeneous. In fact, different user's ICs will generate unsteadiness in the whole decision system.

To the best of our knowledge, no study has treated the formulation of users ICs when using a hypercube with multiple facts tables. Indeed, defining multiple user-defined ICs may generate semantic inconsistencies which affect the analytical quality of such a system.

The main objective of the current study is to propose a system for the implementation of user integrity constraints in SOLAP namely "UIC-SOLAP".

Our proposal integrates all defined Integrity Constraints which are related to the users of SOLAP system along different stages. Indeed, properties of objects in spatial data sources may be adjusted according to every kind of final user. Also, we must verify user queries and aggregations when requesting the SOLAP by the user.

In addition, Integrity Constraints must be defined to respect particular information of a user, so that only concerned data is accessible. These are concerned by the respect of Integrity Constraints and they are all integrated in our proposal namely "UIC-SOLAP".

Secondary objectives are also achieved in this paper. Indeed, we propose a classification of ICs related to the users of SOLAP systems. After that, we propose a multidimensional model for forest spatial data warehouse (SDW). For this aim, we propose a UML profile which allows considering the main properties of spatial and Multi Dimensional (MD) modeling. We integrate also, users ICs definition with formalization of users profiles for specifying visualization ICs according to their preferences. Also, we define inter-facts ICs within the same hypercube. Indeed, unlike the traditional OLAP world, Unified Dimensional Model (UDM), the model for data in Microsoft SQL Server Analysis Services (SSAS), allows to have multiple fact tables within a cube [23].

Then, we propose the exploitation of the UML profile to support "constellation schema" of the forestry Spatial Data Warehouse (SDW). This will help to build a hypercube with several fact tables related to subjects of analysis in forestry management.

Finally, we test our proposal using data concerning forest areas of Mostaganem in Algeria to apply our methodology for verifying the user's integrity constraints in such a system.

IV. INTEGRITY CONSTRAINTS FROM SPATIAL DATABASES TO SPATIAL DATACUBES

Spatial Data Warehouses (SDWs) integrate, organize in a multidimensional way and store very large volumes of spatial and non-spatial data from multiple sources to support the decision-making process within an organization [24].

Spatial OLAP (SOLAP) systems are a category of software tools that allow interactive exploration based on a spatio-multidimensional approach at several levels of detail of the SDW [25]. They extend and enrich the OLAP systems by new concepts and operators which allow effectuating spatial analyzes. Consequently, they allow the visualization of the results of analytical queries in the form of tables, charts and maps.

SOLAP systems are based on data stored in SDWs. As a result, SOLAP analysis quality includes spatial data quality which is the characteristic that distinguishes spatial datacubes from traditional ones.

Here, we must take into account the internal quality defined by its geometric and semantic accuracy; and the external quality which is related to the user needs or quality in the context of use.

In the Spatial Databases, [26] classifies spatial ICs into three categories:

1. Topological ICs which include all geometric properties and relationships on data;
2. Semantic ICs which concern the meaning of geographical features and
3. User-defined ICs which are business rules in spatial databases.

However, in [13], the author notes some limitations when trying to apply these ICs to real applications.

In the Temporal Databases, a formal classification for temporal ICs is introduced in [27]. Indeed, ICs are categorized according to the transaction time (when a fact is stored in the database), valid time (when a fact was true in the reality), and both.

In the Spatiotemporal Databases static and transition ICs is integrated in [26] with the categories of spatial ICs. Here, other temporal ICs are involved. They restrict the possible lifecycle of objects like the IC that specifies the period of applicability of a rule.

The quality of SOLAP decision-making depends on the quality of the data stored and on the way in which the data is explored. Integrity constraints (ICs) are defined in [28] as assertions, typically defined in the conceptual model of an application, that are aimed at preventing the appearance of incorrect data in a database. In SOLAP system context, IC can be used for:

1. identifying erroneous data stored,
2. identifying incorrect analytical queries and
3. defining correct aggregation rules.

In analytical processing, a multidimensional model is based on concepts like dimension, hierarchy, level, member, fact, and measure. As a result, an integrity constraint specification must hold specific semantics supporting these concepts. Hence, it is more efficient to express spatial multidimensional integrity constraints referring to multidimensional elements. At the conceptual level, SDW ICs define conditions that data, metadata or analytical expressions are expected to satisfy. In [8], authors see that the definition of ICs at a conceptual level makes it possible to consider quality problems in the early stages of development. Indeed, they propose two new classifications of SDW ICs: (1) a classification oriented SDW concepts (Fact, Dimension, Aggregation, etc.) and (2) an implementation-oriented classification. The first category categorizes ICs according to the multidimensional elements that these ICs imply. The second group categorizes ICs following the implementation levels in the SOLAP architecture. Indeed, the conceptual expression of any IC depends on the concerned SDW concepts; when its physical translation depends on the SOLAP tier where it is implemented. In fact, he reveals that ICs can be specified using OCL and Spatial OCL at the conceptual level. However, Spatial OCL and UML do not allow specifying all IC proposed in that classification. This is related especially to metadata IC, the aggregation ICs specific to an application field and also visualization ICs. The author in [15], proposes an extension of the precedent classification by introducing a new class related to "Query ICs". This aims to avoid misinterpretations of the results as shown in Fig. 1.

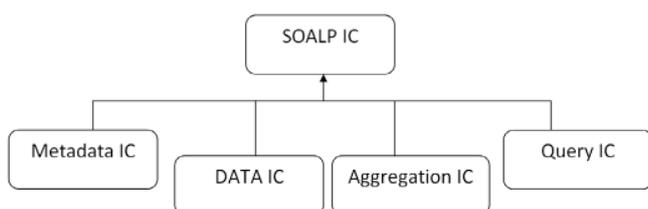


Fig. 1. Classification of SOLAP ICs.

However, the solution can be by adding in the UML profile:

1. more tagged values representing the main metadata,
2. more stereotypes allowing users to define their own aggregation ICs and also,
3. formalization of user profile for specifying visualization ICs according to their preferences.

A. New facts ICs in Constellation Schema

According to [13], a hyper-cell consists of a pair (L, MS) where L is a finite set of dimension levels and MS is a finite set of measures of a hypercube schema. The author classifies fact ICs into three categories. Indeed, a fact can be defined for one fact of a hyper-cell (f-Inter0), for several facts of the same hyper-cell (f-Inter1), and for several facts of several hyper-cells (f-Inter2) as shown in Fig. 2.

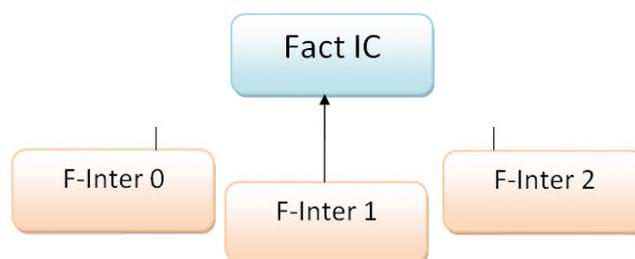


Fig. 2. Categories of fact ICs.

In this classification, the author considers ICs facts of one or several hyper-cells corresponding to the same facts table.

In our proposal, we consider new ICs for measures in several facts tables within the same hypercube as shown in Fig. 3. This situation can occur when considering a constellation schema of a hypercube with several fact tables. In our case study, these fact tables are related to subjects of analysis in forestry management.

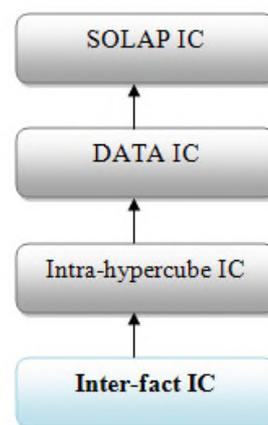


Fig. 3. Inter-fact ICs within a hypercube.

B. Users ICs in SOLAP

ICs related to the users of SOLAP system are numerous. In our proposal, when we use a hypercube with a constellation schema with several users, we classify these ICs into four categories as shown in Fig. 4. This classification may not to be exhaustive but we have focused on these parts that we consider directly linked to the user of such a system.

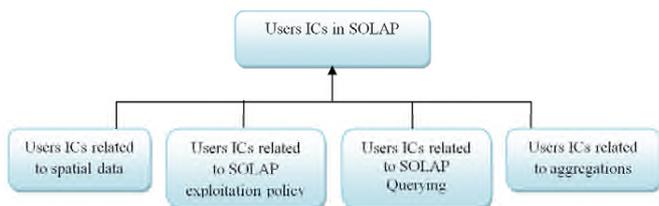


Fig. 4. Users ICs in SOLAP.

1) Users ICs related to Spatial Data

Integrity constraints apply to the database, a set of states that are valid by virtue of properties of objects that need to be stored according to every kind of final user of SOLAP system.

Geometric properties relationships of spatial data are related to «topological», «metric» and «ordering» relationships [30]. Here, semantic ICs concern the meaning of geographical features. The user can define integrity constraints which allow database consistency to be maintained according to his need [26].

After definition of the integrity constraints, we need to check if they are consistent. Indeed, before checking if a database satisfies a set of spatial semantic constraints, we need to make sure that the constraints themselves must not be in conflict [31].

For example, a user rule must be activated to locate *oil station* at a given distance from the *forest*. That must not be in contradictory with another user-defined IC.

2) Users ICs related to SOLAP Exploitation Policy

At the modeling stage, the UML profile may be enriched with a stereotype called “UserProfile”, which allows specifying constraints depending on particular information of a user or a group of users [32]. As a result, we customize information depending on the characteristics of the user that is requesting that information.

The Fig. 5. illustrates user ICs for an example of secure multidimensional modeling adapted from [32]. It shows a part of the Secure Model, named ‘Hospital’ which is based on a typical health-care system. In this example authors notice the following considerations related to the user of the system:

- Patients may access their own information as patients.
- deny access to data of patients who have been treated before the date of initial contract.
- deny access to admission information to users whose working area is different than the area of a particular admission instance.

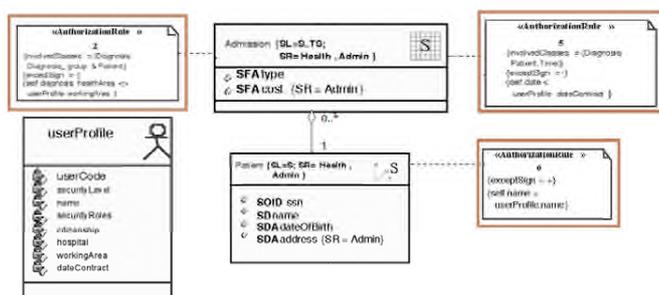


Fig. 5. Example of user ICs adapted from [32].

3) Users ICs related to SOLAP Querying

A SOLAP query is a combination of measures and members of different dimensions. A querying IC verifies the validity of this combination that may be effectuated by any user.

For example, it does not make sense to request geographic distribution of silviculturals effectuated by “*Mostaganem*” forest administration “*conservation*”, in 1985 in “*Zemoura*” forest.

Here, according to the multidimensional model represented with a constellation schema proposed in Fig. 10, we have an analytical query which combines:

- Fact table: [“*silviculturals*”].
- SpatialMeasure: [“*localisation*”].
- Spatial aggregation level: [*forest*= “*Zemoura*”].
- Spatialaggregation level: [*conservation*= “*Mostaganem*”].
- Timedimension: [*year*= “*1985*”].

The result will demonstrate that there is no silvicultural activities; while, according to the administrative division after February, 1984, “*Zemoura*” forest became under the direction of another conservation namely “*Relizane*” (western of Algeria). This can produce unreliable decision. So a query integrity constraint is able to prohibit this combination of parameters.

4) Users ICs related to Aggregations

The author in [15] has implemented aggregation ICs in the UML profile to avoid the implementation of semantically and structurally incorrect models of SDW.

For example, in order to force the user not to aggregate non-additive measures using the aggregate function “sum”, the author defines the OCL statement which forbids this as shown in Fig. 6.

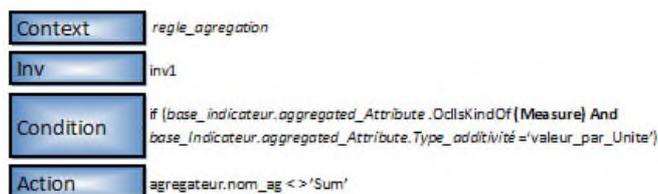


Fig. 6. Example of IC related to aggregations.

Generally, this kind of aggregation rules is valuable for every user. However, any user need to use a limit number of measures related to a specific facts table with group of dimension levels. Hence, we can attach the aggregation rules to the “*UserProfile*” stereotype defined in Case Study section.

V. THE PROPOSED SYSTEM FOR USER INTEGRITY CONSTRAINTS IN SOLAP “UIC-SOLAP”

In the current study we propose a system namely “*UIC-SOLAP*” for checking integrity constraints which regulate the interactions between the user and the SOLAP process as shown in Fig. 7.

This system integrates all defined Integrity Constraints which are related to the users of SOLAP system along different stages of the latter. Indeed, in data sources which can be spatially referenced data, properties of objects may need to be stored according to every kind of final user.

In the interaction with SOLAP Server, the user uses combination of measures and members in analytical queries and he uses aggregation rules. These are concerned by the respect of Integrity Constraints.

Also, when using the user-interface of a SOLAP application, Integrity Constraints must be defined to respect particular information of a user, so that only concerned data is accessible.

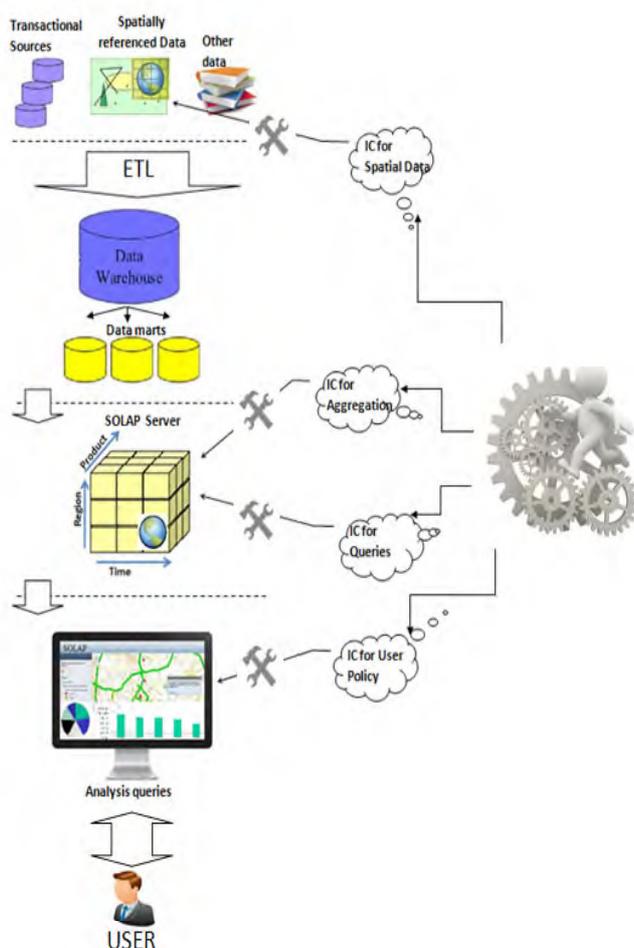


Fig. 7. "UIC-SOLAP" User Integrity Constraints in SOLAP.

VI. THE PROCESS ADOPTED BY "UIC-SOLAP"

Throughout the SOLAP building and using, the process adopted by the proposed system namely "UIC-SOLAP", can be described as following:

- Firstly, the users can make different assertions in the spatial database according to their needs. Along this task, we must take care that these assertions are not in conflict between them.
- Secondly, to identify incorrect analytical queries, we must verify the validity of the combination of measures and members.
- Then, for defining correct aggregation rules, we must define in the modeling stage, stereotypes allowing users to define their own aggregation ICs.
- And finally, for the respect of user policy, formalization of user_profile concept is assured allowing specifying constraints depending on particular information of a user.

The operating algorithm for applying UIC_SOLAP can be as follows:

Algorithm

```
// check data user_integrity_constraints when analyzing
data, before applying "ETL" process
S ← ∅; // set of spatial data properties
P → ∅; // final list of properties initially empty
For (every kind of final_user of SOLAP)
```

Begin

```
S ← Read (database_properties); //user will enter
//spatial properties according to him
```

```
Verify_contradictory (P,S); //verify if entered properties
are contradictory with the others
```

```
While (not ok) // While S is in contradictory with at
least one of database_properties
```

Begin

```
redefine_properties(S) // S will be redefined
```

```
Verify_contradictory (P,S);
```

End while

```
P ← P+S; // add S to final list of properties
```

end for;

```
// (checking exploitation policy user integrity constraints)
When connecting to SOLAP
```

```
Read (user_id)
```

```
Apply_profile (user_id) //applying a set of parameters
allowing user to see only information related to his
specialty
```

```
// ... When querying the SOLAP
```

```
While (user is connected)
```

Begin

```
//checking querying user integrity constraints
```

```
Read(query)
```

```
Verify_combin ({dimension_members}, measure)
```

```
If (not OK) then show_message("inconsistent query will
generate inconsistent results");
```

```
//checking querying user integrity constraints
```

```
Read (aggregation)
```

```
Verify_aggregation(query, aggregation_model)
```

```
//to Verify if requested aggregation function is
allowed to be applied
```

```
If (not OK) then show_message("not allowed aggregation
will generate inconsistent results")
```

End while

End Algorithm.

For the modeling, in our approach we use Unified Modeling Language (UML) since it becomes a standard using object-oriented concept and is widely used for designing various systems and software. UML can be also easily extended with profile to adapt it to specific domain namely multidimensional modeling.

The activity diagram of the proposed system namely "UIC-SOLAP" is shown in Fig. 8.

VII. CASE STUDY

The case study used to illustrate our proposal concerns the multidimensional (MD) analysis in a Spatial Decision Support System (SDSS) for forest management.

We use an extension of UML for MD modeling with three mechanisms: stereotypes, tagged values and constraints in the UML profile [33]. We use also Object Constraint language (OCL) to formalize IC constraints which refine the definitions of stereotypes and tagged values [34].

The forest SDW metamodel we propose is adapted from [35]. It

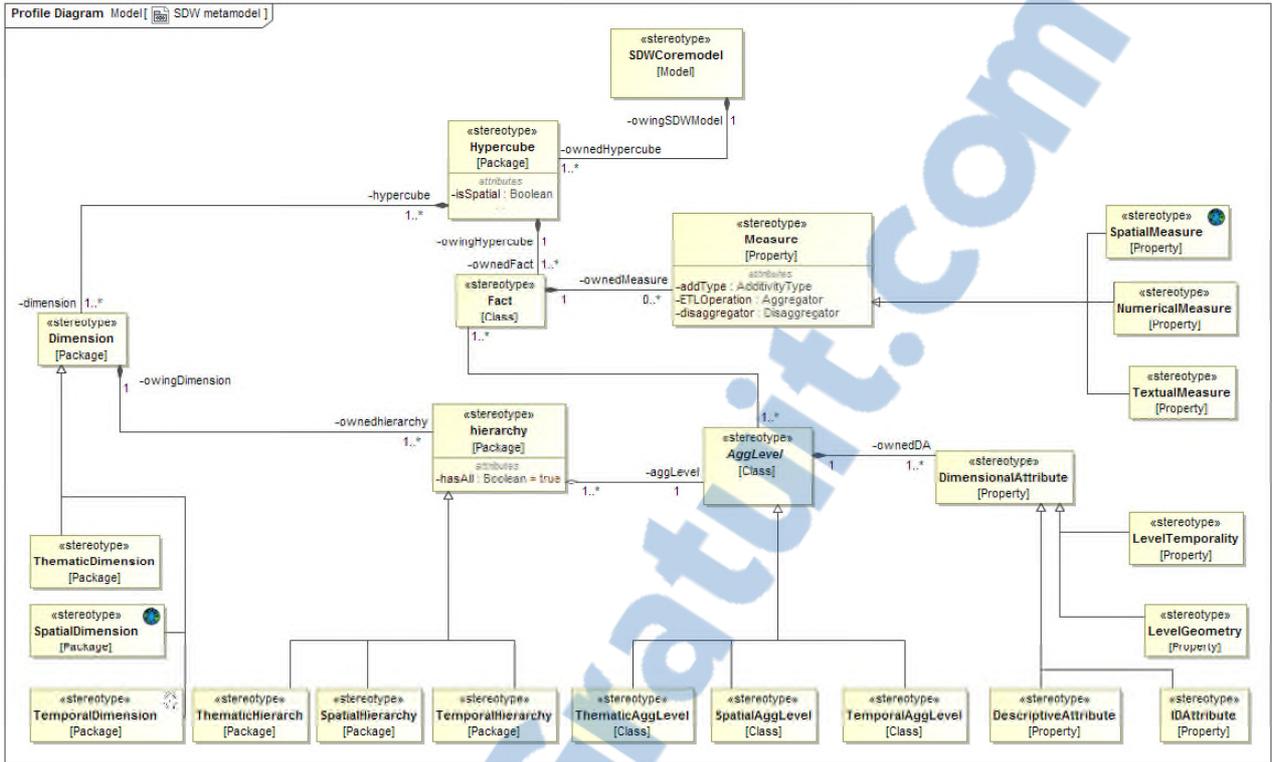


Fig. 9. Forest SDW metamodel.

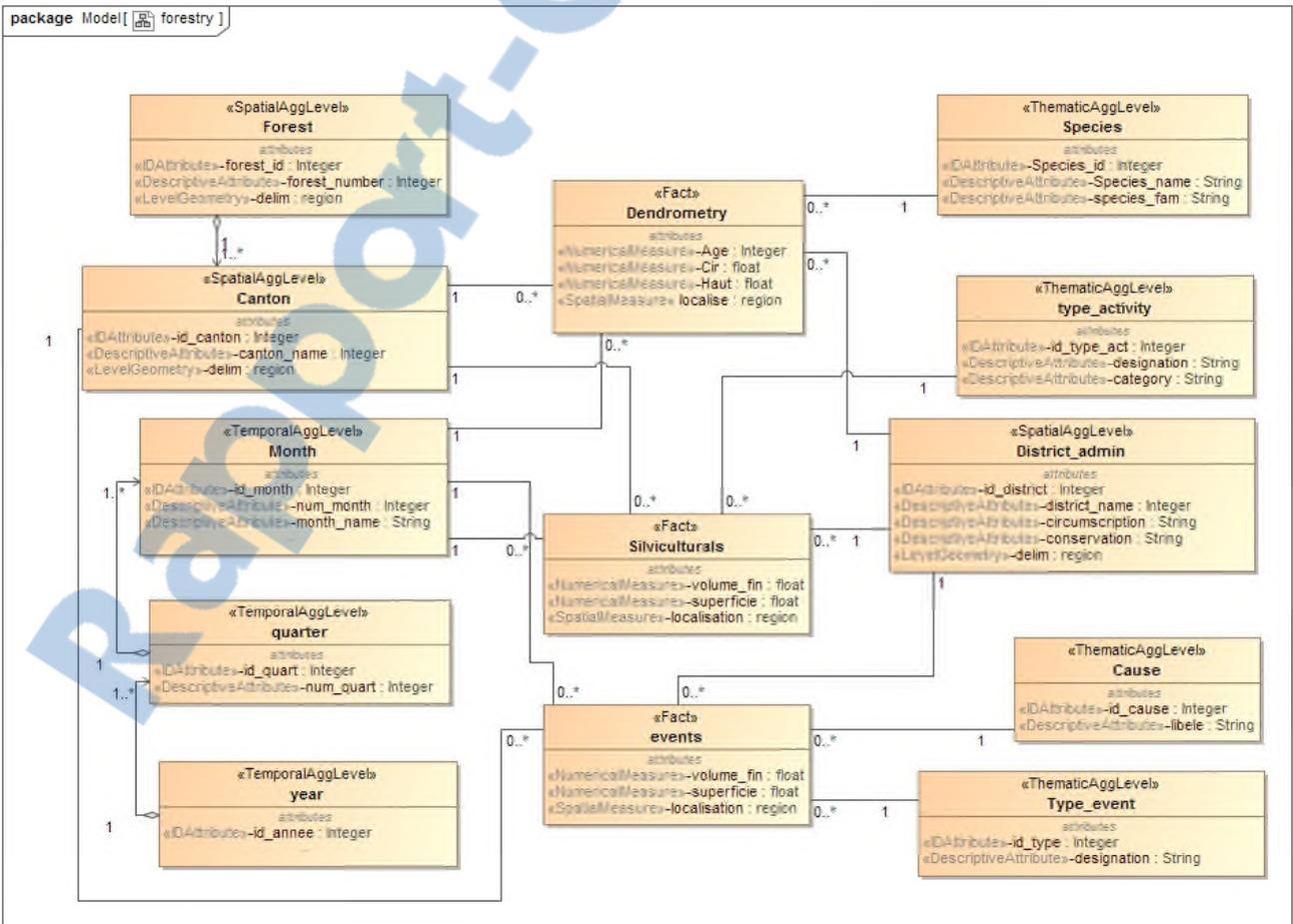


Fig. 10. Constellation schema of forest SDW.

- «Dendrometry» facts table related to the measurements of various characteristics of trees, such as their circumference, *high*, *age*, etc. It is described using three numerical measures (*cir*, *haut*, and *age*), and one spatial measure (*localise*).

Dimensions can be:

- Temporal, to effectuate business analysis at three levels: month, quarter, or year;
- Spatial, to effectuate business analysis at geo-referenced areas. We have two spatial dimensions: administrative with levels (district, circumscription and conservation); and ecologic with levels (canton and forest);
- Thematic, which is related to the business processes. We have four thematic dimensions: *Species*, *Type_activity*, *Cause* and *Type_event*.

We define the *UserProfile* stereotype for specifying IC depending on particular information of the user. In the class diagram, we have just one class of this stereotype; that has no association to other classes. It is used only for formulating IC as shown in Fig. 11.



Fig. 11. UML Stereotype and Class for UserProfile.

For the new type of inter-facts IC within one hypercube cited in section 4.1, in the MD schema of forest SDW we introduce the assertion that verifies prospective conditions between two or more facts tables.

Fig. 12 shows an inter-facts IC between two facts tables *Silviculturals* and *Events* of the same hypercube. It specifies that superficies of reforestation must be greater than superficies destroyed by fires. We note that specification languages of ICs are explained in [13]. However, they go beyond the objective of the current study.

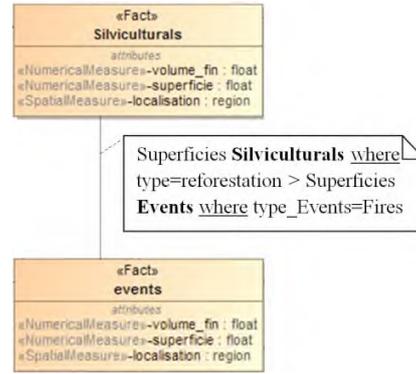


Fig. 12. Inter-facts IC.

VIII. DISCUSSION OF RESULTS

We have developed a SOLAP application with JAVA Eclipse, the object-oriented programming language. With the integration of ArcGis runtime SDK, the solution provides cartographic synchronization with tabular and diagram displays as shown in Fig. 13. Different users may pass their analytical queries using buttons and combo boxes available in the interface. They can also formulate their queries directly in MultiDimensional eXpression (MDX) language to communicate with the “Silvi” hypercube. The prototype is tested with a dataset concerning the Mediterranean forests of Mostaganem department situated in north western of Algeria.

Different users with different aspects of interest can manipulate the system with different manners. This can generate inconsistent analytical results. Hence, the formulation of ICs at different stages can avoid integrity problems.

In the following examples, we will focus on some ICs related to users of SOLAP application.

- Firstly, for the Users ICs related to spatial data, we take into consideration “user-defined” ICs. They are business rules which control the integrity of data stored and results. Here we have three examples related to users of each one of facts tables of the “Silvi” hypercube:

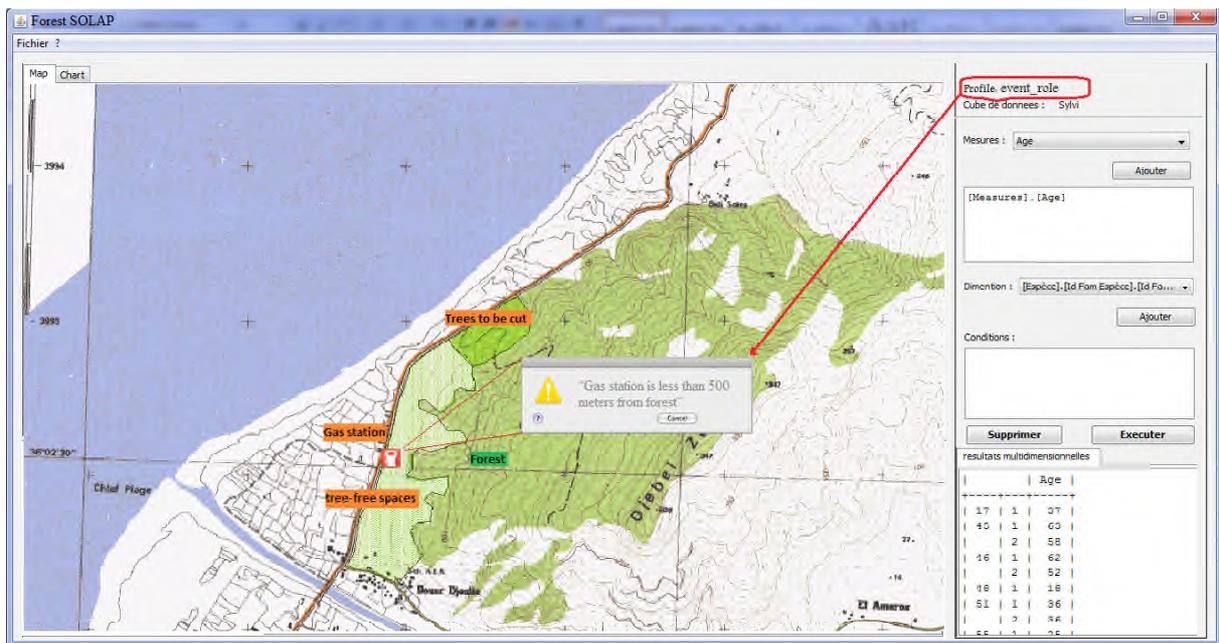


Fig. 13. Data User ICs in Forest SOLAP.

1. "The distance between a forest and a gas station shall be more than 500 meters".
2. "Afforestation can be effectuated at forest spaces tree-free".
3. "it is authorized to cut down a tree when its circumference $>= 4m$; height $>= 20m$; age $>= 50$ and it is near a road".

These ICs may be contradictory since the users are heterogeneous and data will be used in different facts tables within the same hypercube.

In our proposal we try to remove these contradictory issues. Since our prototype may be used by any kind of users, they may query any one of facts tables. Their interests and rules may be different; hence, one user may not accept afforestation near a gas station even it is forest space tree-free; It can be dangerous for other spaces.

- For the Users ICs related to exploitation Policy, by the use of "UserProfile" concept, a "userrule" will be activated only if the concerned user is logging on. We create three profiles: *Event_role*, *Silvicultural_role* and *dendrometry_role* related to the users of Events, *Silviculturals* and *dendrometry* facts tables consecutive.

In this case, if the role is "Event_role" then only the related IC message appears "Gas station is less than 500meters from the forest"; despite "existing of very old trees near the road to be cut" and "existing of tree-free spaces within the forest to be reforested".

We note that the decisional problem can be solved by the application of collaborative group decision support system. However, it goes beyond the objective of the current study which is focused on user ICs in SOLAP application.

- For the Users ICs related to querying, considering the new administrative division, the system we propose verify the correctness of measures and dimension members combination in user queries. For the example above concerning combination of measures of *Silviculturals* facts table and ecological spatial dimension "forest=Zemoura" with administrative spatial dimension "conservation=Mostaganem" is false when time dimension member is after February 1984. Hence, the message in Fig. 14 appears.

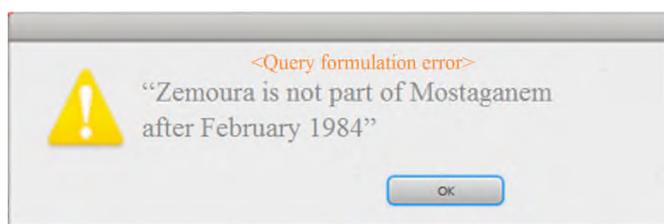


Fig. 14. Notification for non respected query IC.

- And finally, for the Users IC related to aggregation, it defines how user wants to aggregate measures for calculating different indicators. Here, we indicate that every user has a set of indicators for the multidimensional analysis related to his specialty. So we propose that in the aggregation model to be defined in the UML profile, for each indicator we must specify which function to be used for aggregating the measure. ICs are defined in the metamodel and can be controlled when validating the conceptual model.

In this case, since the IC is related to the indicator and the latter is associated to the user, hence we can associate an IC including the 'UserProfile specialty' to the facts table's name. Indeed, this association can enrich the analysis consistency, since every user is able to effectuate correct queries relating to his specialty.

We note that querying a data warehouse is the job of the analyst or decision maker. However, in the current study we focus on the case

when we use a hypercube with multiple facts tables. For example, in group decision systems, they may exist different actors (users). Each of them queries facts of his interest, and has no idea how to use other measures. This seems to be facultative, however it can be considered as a short cut and earlier control for measures aggregation. For example, concerning the *dendrometry* facts table, only users for whom *Userprofile.specialty* = "dendrometry", will be able to aggregating its measures.

IX. CONCLUSION AND FUTURE WORK

Spatial Data Warehouses (SDW) and SOLAP systems are used for discovering business information for Spatial Decision Support System (SDSS). In forestry management, the decision-making process requires the use of multidimensional model including multiple facts tables within the constellation schema.

As stated in studies presented in this paper, SOLAP analysis goodness depends on three quality aspects which concern data, aggregation and queries formulated. For addressing these issues, Integrity Constraints (ICs) have been acknowledged to be an excellent approach. Indeed, they define conditions that must be satisfied to improve the accuracy, consistency, and completeness of databases. ICs are often defined in conceptual models to allow handling these issues at the early stages of development.

The main objective of the current study is to highlight the ICs essentially related the users of SOLAP solution for the effective management of forestry areas and its activities. Indeed, we propose a UML modeling of the forest SDW profile. Next, we use it for the constellation schema modeling with three facts tables sharing spatial, temporal and thematic dimensions. Then, we propose a classification of ICs related the users of SOLAP systems. We propose also a new type of inter-facts ICs related to the use of hypercube with multiple facts tables. Finally, results of our proposal are shown using data concerning forest areas of Mostaganem in Algeria.

In our future works, we plan to extend our results by analyzing other interesting cases in which the SOLAP users ICs can be posed. This can be achieved by enriching our classification presented in this paper. We will address also, other aspects related to the quality in spatial data warehouses and verification of logical consistency in SOLAP Systems.

ACKNOWLEDGMENT

A part of this study was carried out in collaboration with experts from the Forestry Administration of Mostaganem (Algeria). Thereby we would like to acknowledge them with gratitude, in particular Mrs. "Touafek Narimann", the forest Conservator who left us forever and will remain in our hearts and our prayers as long as we live.

REFERENCES

- [1] M.A. Mostafavi, G. Edwards and R. Jeansoulin, "An Ontology-based Method for Quality Assessment of Spatial Data Bases", Proceedings of the 4th International Symposium on Spatial Data Quality, Vienna, Austria. 2004.
- [2] S. Vallières, J. Brodeur and D. Pilon, "Spatial Integrity Constraints: A Tool for Improving the Internal Quality of Spatial Data. Fundamentals of Spatial Data Quality, R. Devillers and R. Jeansoulin (Eds.), ISTE, pp. 161-177. 2006.
- [3] C. A. Hurtado, C. Gutierrez and A. O. Mendelzon, "Capturing summarizability with integrity constraints in OLAP", ACM Transactions on Database Systems, 30(3), pp. 854-886, 2005.
- [4] F. Ghazzi, F. Ravat, O. Teste and G. Zurfluh, "Contraintes pour modèle et langage multidimensionnels". Ingénierie des Systèmes d'Information 9(1), pp. 9-34, 2004.

- [5] T. L. L. Siqueira, R. C. Mateus, R. R. Ciferri, V. C. Times, and C. D. A. Ciferri, "Querying vague spatial information in geographic data warehouses advancing geoinformation science for a changing world" Volume 1 of Lecture Notes in Geoinformation and Cartography, pp. 379–397. Springer, 2011.
- [6] D. Perez, M. J. Somodevilla, and I. H. Pineda, "Fuzzy spatial data warehouse : A multidimensional model", Decision Support Systems Advances, pp. 3–9. IEEE Computer Society, 2007.
- [7] C. E. Dyreson, T. B. Pedersen, and C. S. Jensen, "Incomplete information in multidimensional databases", Multidimensional databases, pp. 282–309, IGI Publishing, 2003.
- [8] K. Boulil, S. Bimonte and F. Pinet, «Un modèle UML et des contraintes OCL pour les entrepôts de données spatiales. De la représentation conceptuelle à l'implémentation», Ingénierie des Systèmes d'Information. 16(6): 11-39, 2011.
- [9] F. Ghozzi, F. Ravat, O. Teste, and G. Zurfluh, "Modèle multidimensionnel à contraintes", Revue d'Intelligence Artificielle 17(1-3), 43–55, 2003.
- [10] E. Malinowski and E. Zimanyi, "Advanced data warehouse design. Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications", Data-Centric Systems and Applications, Volume. ISBN 978-3-540-74404-7. Springer, 2008.
- [11] O. Glorio, and J. Trujillo, "An MDA Approach for the Development of Spatial Data Warehouses", In I.-Y. Song, J. Eder, et T. Nguyen (Eds.), Data Warehousing and Knowledge Discovery, Volume 5182 of Lecture Notes in Computer Science, pp. 23–32. Springer. 2008.
- [12] J.-N. Mazon, J. Lechtenborger, and J. Trujillo, "A survey on summarizability issues in multidimensional modeling", Data & Knowledge Engineering, 68(12), 1452–1469. 2009.
- [13] M. Salehi, "Developing a model and a language to Identify and specify the integrity Constraints in spatial data cubes", PhD Thesis presented at the Faculty of high studies of LAVAL University, Quebec, 2009.
- [14] F. Pinet and M. Schneider, "A unified object constraint model for designing and implementing multidimensional systems". In Journal on Data Semantics XIII, Volume 5530 of Lecture Notes in Computer Science, pp. 37–71. Springer. 2009.
- [15] K. Boulil, "Une Approche Automatisée basée sur des Contraintes d'Intégrité définies en UML et OCL pour la Vérification de la Cohérence Logique dans les Systèmes SOLAP - Applications dans le domaine agri-environnemental", PHD thesis, University Blaise Pascal Clermont-Ferrand II, France, 2012.
- [16] J. Ananda, "Implementing participatory decision making in forest planning", Environment Managing, 39, 534–544, 2007.
- [17] A. Kangas, M. Kurttila, J. Kangas, T. Hujala and K. Eyvindson, "Decision Support for Forest Management", Springer: New York, NY, USA, 2015.
- [18] C. Prell, K. Hubacek and M. Reed, "Stakeholder analysis and social network analysis in natural resource management", Internet Society, 22(6), 501–518, 2009.
- [19] E. Nordström, L.O. Eriksson and K. Öhman, "Integrating multiple criteria decision analysis in participatory forest planning: Experience from a case study in northern Sweden. Forest Policy Econ. 2010, 12, 562–574.
- [20] M. Acosta and S. Corral, "Participatory Multi-Criteria Assessment of Forest Planning Policies in Conflicting Situations: The Case of Tenerife", Forests, 6, 3946–3969, 2015.
- [21] N. Vainikainen, A. Kangas and J. Kangas, "Empirical study on voting power in participatory forest planning". Journal of Environmental Management, 88, 173–180, 2008.
- [22] R. G. Crespo, V. H. M. Garcia, J. J. R. Escribano, E. T. Franco and J. M. C. Lovelle, "Forest fire detection and extinction software", In Proceedings of the 2009 International Conference on Artificial Intelligence, ICAI 2009 (Vol. 2, pp. 885-890).
- [23] H. Sivakumar, C. Matt, M. Sethu, Z. Robert and G. Y. L. Denny, "Professional Microsoft SQL Server Analysis Services 2008 with MDX", Indianapolis, Indiana, USA. Wiley Publishing Inc Press. 2009
- [24] 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, 12(6), 938–958, 2000.
- [25] Y. Bédard, S. Rivest and M.J. Proulx, "Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective. Data Warehouses and OLAP: Concepts, Architectures and Solutions", R. Wrembel and C. Koncilia (Eds.), Idea Group Publishing, p. 298-319, 2007.
- [26] S. Cockcroft, "A Taxonomy of Spatial Data Integrity Constraints", Geoinformatica, 1(4), pp. 327-343, 1997.
- [27] M.H. Bohlen, "Valid Time Integrity Constraints", Technical Report 94-30, Department of Computer Science, University of Arizona, Arizona, USA, 22 pages, 1994.
- [28] K. Boulil, S. Bimonte and F. Pinet, "Un cadre conceptuel basé sur UML et Spatial OCL pour la définition des contraintes d'intégrité dans les systèmes SOLAP". Actes des 8èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 2012). Bordeaux, France, 2012.
- [29] M. Duboisset, Un système de contraintes d'intégrité OCL pour les bases de données Spatiales – Application à un Système d'Information pour l'Épandage Agricole, PHD thesis, école doctorale pour l'Ingénieur (SPI), Clermont-Ferrand, France, 2007.
- [30] M.J. Egenhofer, "A Formal Definition of Binary Topological Relationships", Proceedings of the 3rd International Conference on Foundation of Data Organization and Algorithms, LNCS 716, Springer-Verlag, Paris, France, p. 457-472, 1989.
- [31] L. Bravo, and M. A. Rodriguez, "Semantic Integrity Constraints for Spatial Databases", Proceedings of the 3rd Alberto Mendelzon International Workshop on Foundations of Data Management, Arequipa, Peru, May 12-15, 2009.
- [32] R. Villarroel, E. Fernández-Medina, M. Piattini and J. Trujillo, "A UML 2.0/OCL Extension for Designing Secure Data Warehouses", Journal of Research and Practice in Information Technology, 38(1), 2006.
- [33] J. Trujillo, E. Soler, E. Fernández-Medina and M. Piattini, "A UML 2.0 profile to define security requirements for Data Warehouses", Computer Standards and Interfaces, 31(5), pp. 969–983, 2009.
- [34] K. Boulil, F. Le Ber, S. Bimonte, C. Grag and F. Cernesson, "Multidimensional modeling and analysis of large and complex watercourse data: an OLAP-based solution", Ecological Informatics, 24, 90–106, 2014.
- [35] K. Boulil, S. Bimonte, F. Pinet, "Conceptual model for spatial data cubes: A UML profile and its automatic implementation". Computer Standards & Interfaces, 38, pp 113–132, 2015.



Abdallah Bensalloua Charef

Specialised in software engineering, Spatial Techniques, Geographical Information Systems (GIS), Remote Sensing and Geo-decisional Systems. He is computer science professor at the University of Mostaganem in Algeria.



Hamdadou Djamilia

Specialised in Decision Support Systems, Multicriteria Analysis and Collaborative and spatial decisional Systems. She is professor and leads a research team at the laboratory of computer science of Oran (LIO) at the University of Oran1 in Algeria.

Sharing and Security of Spatially Referenced Data in SOLAP: An Application to Agro-Forestry Management

ABDALLAH BENSALLOUA Charef¹, HAMDADOU Djamila² and BELDJILLALI bouziane³

Computing Laboratory,
University of Oran

P.O Box 1524, El M'Naouer, 31000 Oran, Algeria

{¹ Charef_bensalloua, ² dzhamdadoud, ³ bouzianebeldjilali}@yahoo.fr

Abstract – This paper presents the necessity of the use of a Spatial On-Line Analytical Processing (SOLAP) in some professional domains, in order to be enabled to navigate through spatial data, and also as a security oriented system. This domain is agro-forestry management, which uses and produces a large amount of spatial data to be analyzed. This data has a specific and complex nature since it can be in the form of texts, numbers, vectors or images. This tool can be developed for interactive and rapid analysis; for example introduce spatial data type in data layer, display spatial data in human readable manner, ... etc. so that the results of the analyses are the basis of strategic business decisions in the world of agro-forestry management. So, Remote-sensing data structure which concern different domains (climatic, hydraulic, geologic ... etc) necessitates a good management of data warehouse, to be exploitable for analyze, visualization and integration in a management and decision support in the agro-forestry sector. Data collected in a study of different tasks of this domain, is used for the conception of different database models, to implement its internal structure in the 'ORACLE' Database Manager System (DBMS). Furthermore, different data layers are prepared by *MapInfo* Geographic Information Systems (GIS) software, in order to be imported by the use of *MapInfo MapX* library, in the prototype interface built by *Delphi*.

Keywords: SOLAP, data warehouse, spatially referenced data, data sharing, data security, agro-forestry management.

I. INTRODUCTION

With the aim to manage their work, various land management organisms collect a large amount of spatial data. These data are usually stored in transactional systems called On-Line Transactional Processing (OLTP) to provide security and manage concurrent access to databases. However, transactional systems are not planned to support decision making.

For this purpose, new information systems are developed, including analytical systems, called Business Intelligent (BI) solution. The most used tools in this area are the On-Line Analytical Processing (OLAP) which give the possibility of an

interactive exploration of the data warehouse which is a central component of such system.

The implementation of a tool for the agro-forestry space management at different observation axes (cadastre ownership, fauna and flora management, forest works ...), result the intensive use of databases which benefit from spatial application service.

In order to explore and manipulate spatial data, we use a SOLAP tool which integrates concepts from two different worlds, namely geographic information systems (GIS) and OLAP.[1]

However, GIS is potential candidate to support decisional need, but it is designed neither to support decisional application nor to support highly interactive navigation through spatial data at different levels of aggregation and through different epochs; so an alternative solution which is SOLAP, can be used for analytical purpose [17].

The main objective of this study is to demonstrate the relevance of an application SOLAP as a strategic tool for effective management of natural areas, and the various activities related to agro-forestry, by allowing different actors in this field, work together to take consistent, integrated and sustainable decisions. And, in other side, this will be used as a security oriented system, to ensure data confidentiality.

In Section 2 we review some SOLAP solutions that exist in the literature. Section 3 is devoted to the description of our contribution. Section 4 presents general concepts of OLAP and SOLAP systems. Section 5 is devoted for analysis and requirements specification in the selected domain for this study which is the agro-forestry. Structuring and data modeling are described in Section 6. In Section 7, the implementation of the database and building the prototype are described. Section 8 describes in detail the procedure for using the proposed application "*forest_sys*". Finally, we conclude our discussion in section 9, by giving our perspectives.

II. RELATED WORKS

This section presents some works, focusing mainly on the developing techniques of SOLAP tools; in fact, in literature, there are few authors who address issues related to the integration of spatiotemporal data in a data warehouse, in Business Intelligent and security purpose.

In [9] and [10], the authors propose solutions to design multidimensional structures where data sources are heterogeneous in terms of temporal, spatial and semantic. The objective is to enable the extraction of geographical knowledge by exploring detailed data associated to an epoch.

Another study addressing the problem of data heterogeneity of forestry, in decision making, is that made by the authors in [7] and [8], in this study, the authors discuss the problem of the evolution of cubic structure of geospatial data in forestry. Indeed, the proposed model for the semantic level improves accuracy by evaluating the similarity not only in concepts, but also between the properties of concepts. At the geometric level, the approach is based on a Quadtree index method of the tool Oracle Spatial to form a matrix of spatial correspondences between the geometries of different epochs. In some cases, we use more than one data cube space to meet needs of decision-making.

In [2], the authors take an example in the analysis of the risk of wildfire, assuming that there are two spatial data cubes. And as these cubes are usually modeled differently from one organization to another or even from designer to another, the authors of this paper proposed a categorization of the problems of heterogeneity taking into account the different cubes spatial data. Therefore, they defined four categories, including the heterogeneity Cube to Cube, the heterogeneity Dimension to Dimension, heterogeneity level to level and heterogeneity measure to measure. For each category, they considered the various components of models of data cubes, including schema and metadata.

The authors of [11] have addressed another issue concerning the dimensions in hypercubes; in fact, the multidimensional approach proposed is based on the concept of facts seen in various dimensions. These dimensions are considered as analysis axes forming a space to identify the facts by a set of coordinates. Thus, dimensions have discrete values. However, these discrete values are not sufficient when analyzing natural phenomena such as continuous meteorological or pollution. In this paper, the authors have introduced mechanisms based on interpolation in the spatial and temporal dimensions that give the user the impression of navigating through a continuous hypercube, applying new aggregate functions for data processing.

In most works, the addressed problems are related to the heterogeneity of the data manipulated, or development methods. However, few studies have addressed issues related to users, who are the decision makers and can be related to independent organisms (actors). In fact, currently, in agro-forestry, a theme, such as "land use planning", is mostly treated by several structures (public works, cadastral ownership, forest management and the companies which use forest products, etc.). Therefore, the inconsistency of their work generates

products (maps, reports, statistics, etc..) which can not be aggregated or reused because their tools are strongly related to specific tasks. In addition, they often use different analytical tools providing not the same indicators, to make decisions affecting the same territory.[21]

To overcome this problem, an interactive decision support in agro-forestry is can be developed to enable data sharing and security.

III. OLAP AND SOLAP CONCEPTS

OLAP is defined as "a software category intended for the rapid exploration and analysis of data based on multidimensional approach with several aggregation levels" [17]. It includes dimensions which represent the analyses themes or axis. These contain members that are organized into levels of granularity or details. It include also measures which are numerical values analysed against the different dimensions. Facts are defined as the combination of dimension members and measures values. And then, we call Data cube, a set of measures aggregated according to a set of dimensions [17].

A combination of OLAP with GIS concepts gave birth of Spatial OLAP which can be defined as "a virtual platform built especially to support rapid and easy spatio-temporal analysis" [8].

As a system based on data warehouses, architectures of Spatial OLAP are at three levels, which can be summarized as follows:

- *Data Warehouse*

Data are extracted, formatted and integrated into the data warehouse, with the existence of metadata (information about data sources).

- *OLAP Server*

It allows to perform data analysis in accordance with the multidimensional paradigm, (eg: Oracle OLAP, Microsoft Analysis Services, etc.) and provides users with a multidimensional view of data that can be analyzed through a set of OLAP operators (Roll-Up-Down Drill, etc.).

- *Client OLAP*

Provides a user interface with interactive analysis tools (eg JPivot, Polaris... etc.). Its role is to make visible information, by extracting useful indicators to make decisions.

On-line analytical processing is also, defined as "the name given to the dynamic enterprise analysis required to create, manipulate, animate and synthesize information from exegetical contemplative and formulaic data analysis models. This includes the ability to discern new or unanticipated relationships between variables, the ability to identify the parameters necessary to handle large amounts of data, to create an unlimited number of dimensions, and to specify cross-dimensional conditions and expressions"[17]

OLAP system is based on multidimensional approach which gives the concept of: dimension, members, measures, facts and data cube. A dimension can be considered as an axe of analyze. They can be spatial such as location names. They are also organized hierarchically into levels of granularity or details.

A dimension is composed of members who can be aggregated to constitute the members of a higher level.

A measure is a numerical attribute analyzed by a system user by combination of a set of dimensions.

The combination of dimension members with the resulting of measures value for a particular aggregation level represents a fact [9]

A data cube is composed of a set of measures aggregated according to a set of dimensions.

Every OLAP system has an architecture which can be divided into three parts: the multidimensionally structured database, the OLAP server that manages the database, and the OLAP client that gives access to the database via the OLAP server “Fig. 1”. This can be by exploring and analyzing the data using different visualization methods and adapted operators such as drill-down, roll-up, drill-across and swap.

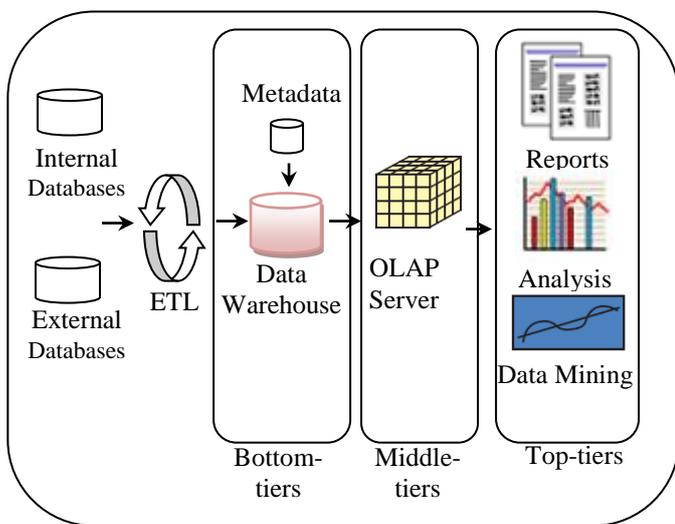


Fig 1. OLAP-based decisional system

OLAP application represents a good support for the usage of spatial dimensions in a multidimensional analysis process. However, this usage can be without consideration for the cartographic component of the data, which signifies that no spatial visualization or exploration of data will be permitted.[4],[6].

With the development of spatial data warehouses concepts, a new solution was then developed, which consists of combining the strengths of GIS, with the strengths of OLAP tools into a new type of application. This combination gave birth to Spatial OLAP applications which allows fully analyze of the geometric aspect of the spatial data.[8]

From these concepts a Spatial OLAP can be defined as “a visual platform built especially to support rapid and easy spatiotemporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic displays as well as in tabular and diagram displays” [2].

SOLAP tools support the multidimensional database structure described earlier. However, with the spatial data manipulation capabilities included, three types of spatial dimensions can be used: the non-geometric spatial dimensions, the geometric spatial dimensions and the mixed spatial dimensions. “Fig. 2” presents the three types of spatial dimensions.

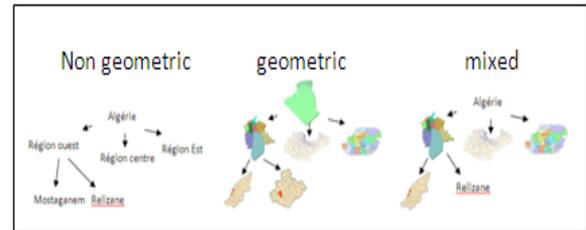


Fig 2. The three types of spatial dimensions supported by SOLAP tools

In the first case, which is used in conventional OLAP, no cartographic representation is associated with the members. The second case, in order to help in graphical visualization and queries, it include geometric shapes geo referenced to the map. The third case as its name indicates, it uses nominal data and geometric representation for the dimension members.[12],[18],[19].

A. Measures in SOLAP

In a SOLAP application, a measure can possess a geometric representation called then spatial measure, or only numerical one as in a conventional OLAP.

In a spatial measure we have two types: The first type consists of a set of coordinates, which requires a geometric operation, such as a spatial union, a spatial merge or a spatial intersection, to be computed. It is the set of all the geometries representing the spatial objects corresponding to a combination of dimension members. The second type of spatial measure results from the computation of spatial metric or topological operators. Ex: surface, distance number of neighbors.[22],[23].

Finally, the measure values (spatial and non-spatial) that result from the combinations of spatial and non-spatial dimension members are visualized using a SOLAP client.

A SOLAP client can be used with any type of SOLAP architecture: relational (ROLAP), multidimensional (MOLAP) or hybrid (HOLAP).[24],[25]

B. SOLAP operators

In the SOLAP application, different operators are used in order to take advantage of the multidimensional data structure. The most important operators are drill-down, roll-up, drill-across and swap.

A drill-down operation allows the end-user to navigate from a general level to a more detailed level inside a dimension

A rollup operation allows the user to navigate from a detailed level to a more general level inside a dimension

A drill-across allows to view different information but at the same level of detail.

A swap operation allows interchanging two dimensions in order to view the same information in a different way or in order to view different information.

If we manipulate the data on maps, these operators will be named spatial drilldown, spatial roll-up and spatial drill-across.

C. Categories of SOLAP

GIS and OLAP tool integration can be in different approaches: GIS-dominant, OLAP-dominant and total integration.

In the first approach we find full GIS capabilities, and practically no OLAP functionality except simplified access.

The second approach offers full OLAP capabilities but only cartographic GIS functionalities.

The third approach necessitates fully integration of GIS and OLAP capabilities. However, Implementing a SOLAP system in this case without using SOLAP technology requires major development.[11]

In order to design the spatial data warehouse architecture, a set of transformation rules from the conceptual model to logical design strategies is needed such as:

- Relational OLAP (ROLAP),
- Multidimensional OLAP (MOLAP)
- Hybrid OLAP (HOLAP)

D. Extract Transform Load process

ETL (Extract Transform Load) process is an important stage of the data warehousing development, and its design and implementation can be considered taking the half of the time for setting up a data warehouse.

In the literature some approaches were devised for conceptual modeling of the ETL process from the functional, the dynamic or the static points of view.

Nevertheless, issues such as the optimization of ETL logical schema are not very well understood. Besides, there is a need for techniques that automatically propagate changes occurred in the source schemas to the ETL process.

A variety of commercial ETL tools exist in the market. A lot of research efforts exist mostly targeting modeling and methodology issues. Some works are focused on the end-to-end methodology for the warehouse and ETL projects targeting the complete life cycle of the DW project, describing how to plan, design, build, and run the DW and its ETL backstage.

A data warehouse project consists of three main technical tasks: ETL, database design, and analysis tools.

In a data warehouse project, the most important step is the requirements collection. in order to individuate the best

approach to ETL, it is essential that requirements must be well defined.

In every data warehousing development, there are two main approaches to ETL: to resolve it implementing a set of scripts and ad-hoc programs or to buy a commercial ETL tool that simplifies the job of proposing a methodology, a graphical interface, and a language to solve specific problems.[3]

In “Fig. 3” it is described the general framework for ETL processes. On the left side, they are shown data sources that are involved in the overall process. Usually, data sources are relational databases and files. they are extracted by specialized routines or tools, which provide either complete snapshots or differentials of the data sources. Then, these data are transformed and cleaned before being loaded into the data warehouse.

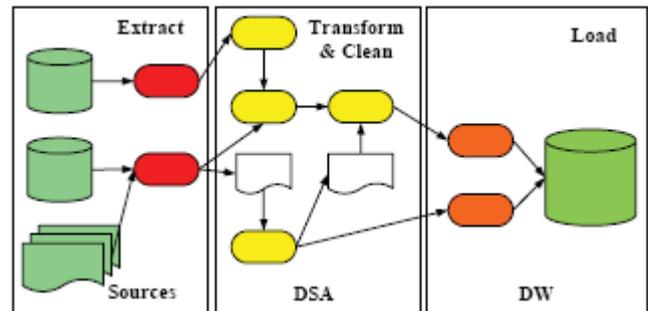


Fig3. The environment of extraction-transformation-loading processes (from: [3])

ETL processes constitute the major part of a data warehouse environment, resulting in the corresponding development effort and cost.

The following figure “Fig. 4” demonstrates that the life cycle of a data warehouse begins with an initial reverse engineering and requirements collection phase where the data sources are analyzed in order to comprehend their structure and contents.

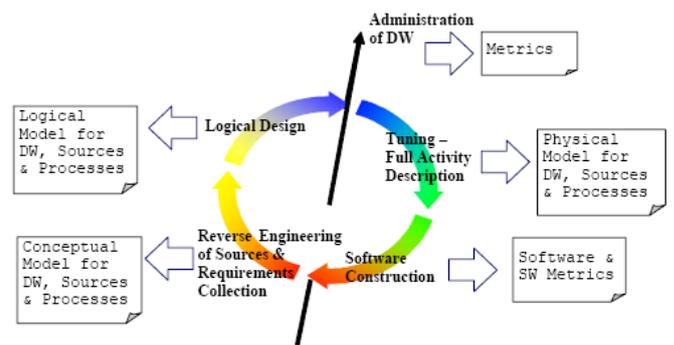


Fig4. The life cycle of a data warehouse and its ETL processes (from: [3])

IV. USING MAPS IN DECISIONAL PROCESS

The geographic information leads to knowledge objects and phenomena. These knowledge concern position or distribution in territories.

Maps are usually used for Visualizing geographic data for whatever objects or phenomena. This can help users to understand these phenomena and facilitate the extraction of

knowledge inside. Spatial relationships and distribution relative to adjacency, connectivity, inclusion, proximity, and exclusion are the most of this knowledge. In fact, visualizing map facilitates discovering correlations between phenomena. We can also view its evolution and relationships contained within its components. So, we can say that really, maps are tools for intuitive knowledge, and also can be considered as an open database rich of exploration indicators.

Geographic information systems (GIS) are used for gathering, storing, manipulating and displaying spatial data. These systems are based on data visualization by maps.

Since the map is considered as a database, sometimes we have to visualize data at different scales. And when the detail level becomes too important, map becomes overcrowded and unreadable if we need to have a more global cartographic view. Here, we talk about using map generalization processes. This process is considered, in [14] as "... responsible for reducing complexity in a map in a scale reduction process, emphasizing the essential while suppressing the unimportant, maintaining logical and unambiguous relations between map objects, and preserving aesthetic quality". During this process, categories of objects as well as individual objects are eliminated, others are replaced by a symbol of larger or smaller size, some are displaced, their shape is simplified, topological relationships may change, groups such as "building blocks" replace individual buildings where the density is too high. The following figure "Fig. 5" shows an example of a map generalization.[26]

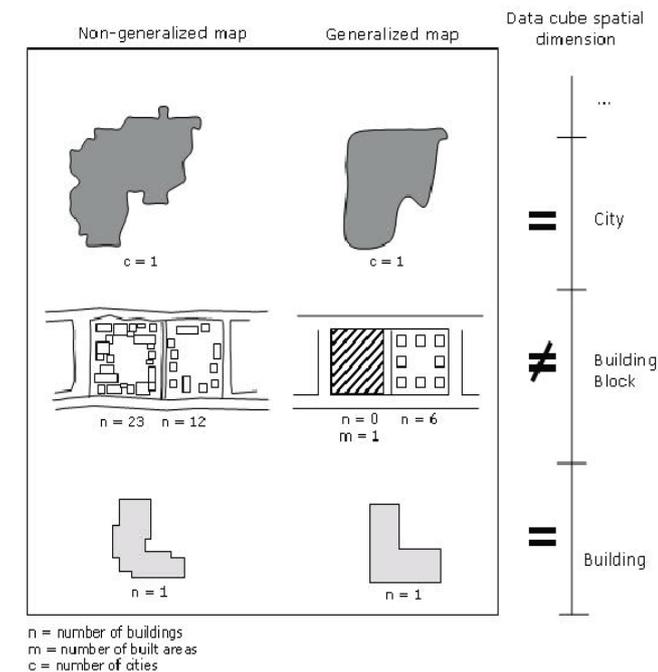


Fig5. Example of spatial aggregation-generalization (from: [26])

In [14], authors develop a list of motivations for generalization as follows:

- Develop a primary database
- Use resources economically
- Increase/ensure data robustness
- Derive data and maps for a range of purposes

- Optimize visual communication

V. SPECIFICATION AND ANALYSIS OF NEEDS

We begin by conducting an analytical review of key aspects of Agro-forestry World and fields of use of this system and data sharing.

A. Actors in the world of agro-forestry and data sharing

In the current state, in most of the time one theme is treated by several structures (public works, land registry, land management and business, etc.).

The inconsistency of their work generates results that can not be reused or aggregated. To overcome this problem, design a prototype information system for agro-forestry is necessary to enable the sharing of data and results [2].

B. Data sources of the system

The data used are heterogeneous in nature; so, they may be:

- Administrative data, such as restriction files and agro-forestry boundary properties, personnel forestry and agriculture files, information on organizations involved ...etc.
- Dendrometric data and statistics (flora and fauna), based on the achievement of the perpetual inventory of agro forestry, which are the descriptive characteristics of plots and trees, and the measurements made on trees.
- Map data on geology, soil,... etc.

C. Applications of the System

The application of this system is multiple; in fact, it can be used to estimate the amount of wood available in the medium term given the current state and a forestry scenario determined by the user; thus, for reflections on land use planning.

In addition, the purpose of part of this work is, using remote sensing techniques and data bank on forest fires, to allow these prejudices and obtaining information necessary for the services responsible for combating against fire.

VI. DATA MODELING AND STRUCTURING

After the synthesis of all concepts of the World Agro-forestry, we deduce the different tasks and work that must be made by stakeholders in the management of agro-forestry areas; from which we will deduce the various entities (classes) of the Database, and conclude relations which join it; this serves to trace the conceptual schema of the Database.

The analysis of each of these tasks will produce a set of classes that will, in the end be aggregated using the method of Hypergraph-Based Data Structure (HBDS), hence, classes bearing the same attributes and relationships can be grouped into a super class [11].

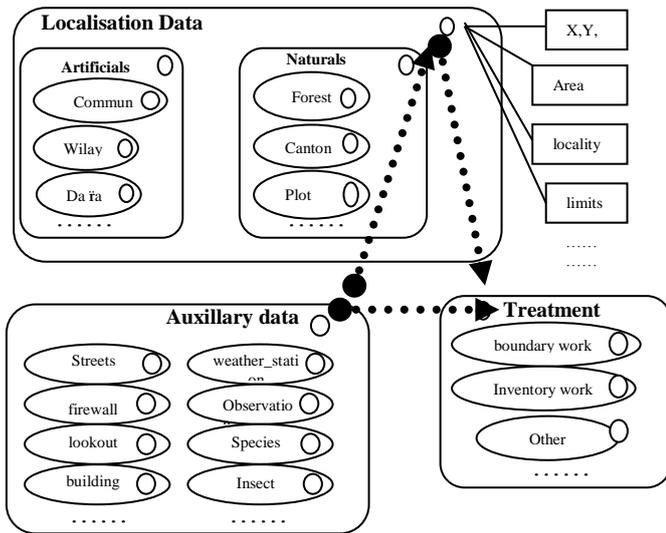


Fig 6. HBDS Global Schema of the Database

We have taken the following decision model “fig. 7”:

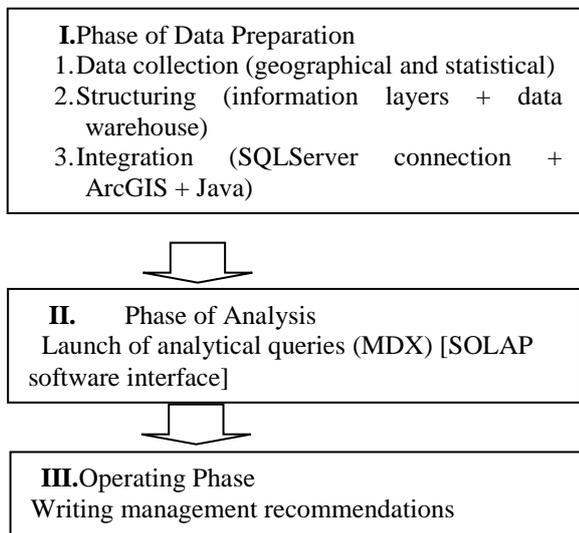


Fig 7: decisional model for natural hazards management

To construct SOLAP architecture the first step consists in building the multidimensional database. We have done it with a relational environment. It is composed of one table called the fact table containing the measures which are values of indicators, and a set of other tables called dimension tables representing the analysis axes.

The list of indicators will serve to define the features of the MDDB. In our prototype we have used a list of management indicators such as: number of buildings, number of residents, flood zones area ... and others. [13]

Each measure value is calculated for a combination of dimensions and aggregated for each hierarchical level of the selected dimensions.

Its value constitutes one cell of the hazard’s data cube: one cell is an intersection of all dimensions used in an analysis process.

To generate good indicators in this process the results will be from the combinations of geographical and non geographical information.

For the modeling developers usually focus on model called dimensional fact model. it is a graphical conceptual model. Basing on this model, user queries can be easily expressed. It also, supports specification of user requirements thanks to the dialogue with the designer

In our prototype, the dimensions used can be: spatial dimensions, temporal dimensions, nominative dimensions.

The nominative dimensions are the dimensions without spatial representation “Fig. 8”.

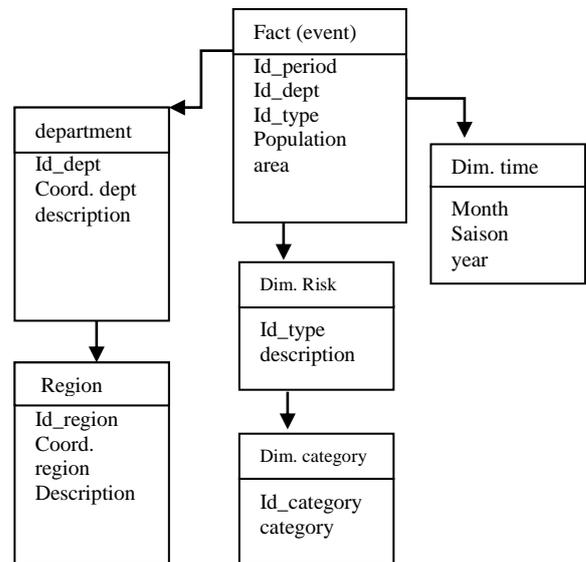


Fig 8: Structure of the data cube

In this model fact table contains all foreign keys of the dimension tables and all the measures. The schema is not normalized as becomes obvious, and this for facilitates analysis queries which implies some redundancies.

For small or medium-sized data volumes, such schemas have a sufficient performance because join operations are only necessary between the fact table and the related dimension tables. But for important sized data the snowflake schema is recommended: instead of modeling each dimension by one table, a table is created for each level of a hierarchical attribute.

VII. DATABASE AND PROTOTYPE IMPLEMENTATION

The design made in the previous section must pass onto physical modelling provided by the ORACLE DBMS, followed by the development of the client application, and preparation of map objects in MapInfo GIS software.

A. Use case diagram

In order to initiate the use structure of the database, it is essential to establish the use case diagram of the database, that records the users and their roles for handling different objects in the database [13].

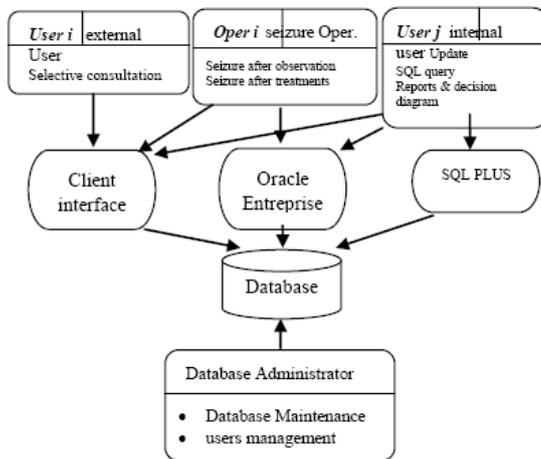


Fig 9. Use case Diagram of the Database

B. Database modeling and roles definition

It first starts by creating a user "oper2" who owns all the objects in the database, then we shall create the following:

- A user "admin" with the role of a database administrator, having therefore the right to manage users (create, delete, modify);
- Two "roles" (groups of privileges): "foret_select" which can be attributed for users in consultation, and "foret_tous_droits" which is a role assigning rights to update all "oper2" tables;
- Two users: "oper1" with all the fingers to access and change tables "oper2", and with the fingers "oper3" consultation "oper2" tables.

C. Preparation and filling the database

Filling the database can be assured by the seizure operator, and this, either directly using the application interface, after having prepared the data by pre-treatment, as in cartographic data. These data must undergo operations of scanning, geo-referencing and digitizing layers of information to be included in the database [16].

D. Creating the client application

The database created can be used directly in updating, or internal or external consultations. This is done directly through the tool called "Oracle Enterprise Manager" (OEM); in fact, it has all the tools to facilitate these operations [15].

However, an implementation of a Microsoft Windows prototype makes these operations more flexible; in fact, we used Borland Delphi, which provides integrated tools for connecting to databases servers, and create a rich interface in Object-oriented data [5].

Although the use of Oracle DBMS is justified through its various assets, the pattern of the user interface design is based on a global view of various features of the prototype, the interface can be modified or enhanced by involving different actors in agro-forestry management world, to develop a common model, so that this can ensure more sharing and security of data.

E. Security in SOLAP systems

Information systems security refers to the processes and methodologies involved with keeping information confidential, available, and assuring its integrity.

It refers to:

- Access controls, which prevent unauthorized personnel from entering or accessing a system. This can be by using authentication process to access the application
- Protecting information in a storage area. Even at the spatial data warehouse or the data cube
- The detection and remediation of security breaches, as well as documenting those events.

In every software engineering application as well as SOLAP there is a serious requirement for information security to be considered by application designer at every stage of development. This must be from the first stage which is analysis of requirement to implementation stage.

However, in SOLAP applications, this requirement is more interesting since they are used for decision making aims and data warehouses store strategic and confidential data

In literature we find some approaches showing for security in SOLAP applications and data warehouse modeling. But they don't treat security at all stages of data warehouse development.

In addition in this type of applications security must be focused on the main concepts of multidimensional modeling such as facts, dimensions, and measures

VIII. THE USE OF THE PROTOTYPE

The use of the prototype is controlled by the administrator, and access is only possible through a username (login name) and password (pass_wd).

The next section describes some windows displayed using the prototype in client-server mode on a machine running the kernel of the Oracle database "Forest" and the client application "forest_sys" that can be divided in three modules.

A. The module "Gestion des utilisateurs"

Fig. 7 demonstrates an interface to see all users, modify their settings, as well as creating new users. This module can be accessed if and only if the user has administrative privileges.

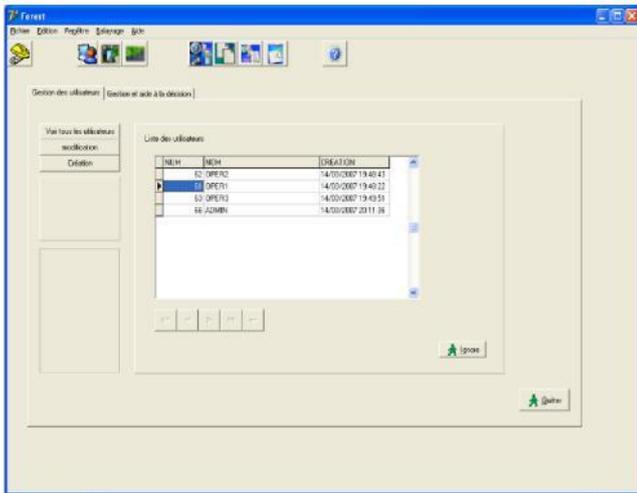


Fig 10. Users managing

B. The module "Gestion des espaces agro forestiers"

This module contains several components:

1) Decision support Space

This window is composed of three entities:

- **Pivot:** choose the "dimensions" representatives of the x-axis graph, and the "summary" representing the y-axis. It should be noted that a choice of several dimensions at once is possible for a single brief statement through the "Group By" in a SQL query.
- **Canvas chart:** to draw the graph according to the dimensions and the summary chosen.
- **Table Analysis:** shows the statistics for the graphic displayed.

The next window shows an analysis of silvicultural work costs, depending on the end date of execution. This analysis can be modified, or completely, replaced by another analysis with the pivot.

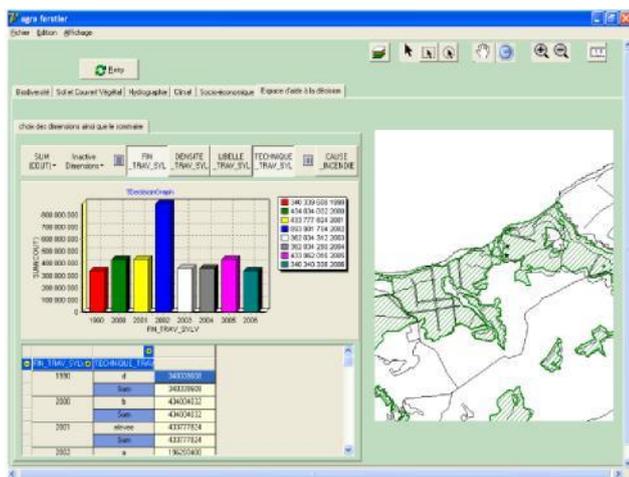


Fig 11. Silvicultural work Analysis

2) Socio-economy

This window is displayed as a series of questions may be asked by an agro-forestry manager.

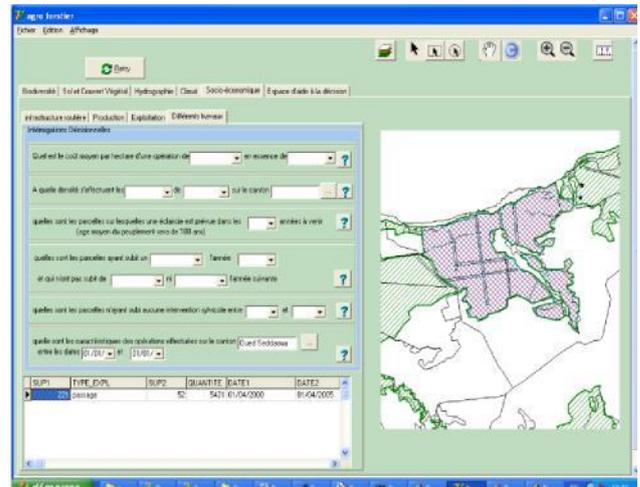


Fig 12. Decisional Interrogations

We note that the database developed can answer any other question, using the module of *Oracle SQL*Plus*. This question is launched by the administrator of the database, as SQL query [20].

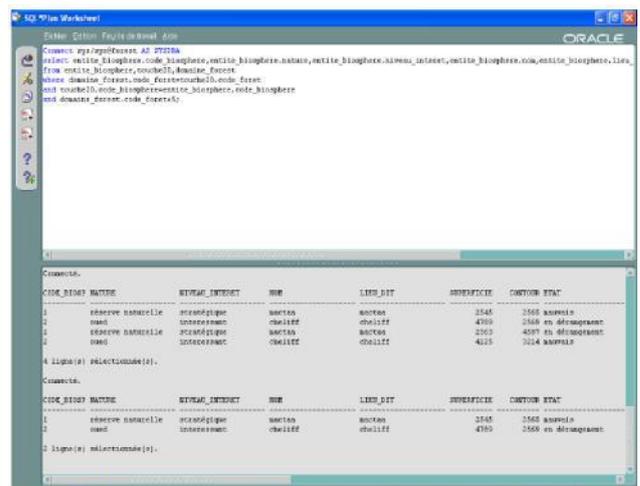


Fig 13. Launch applications from SQL*Plus

Here in "Fig. 13", an SQL query on "wetlands", in intersection with Forest Township Number 5 in the department of Mostaganem (Algeria), the result is displayed on the lower part of the window.

3) Fire Management

Here, you can start the search in two modes, either by specifying the area, or start searching between two dates.

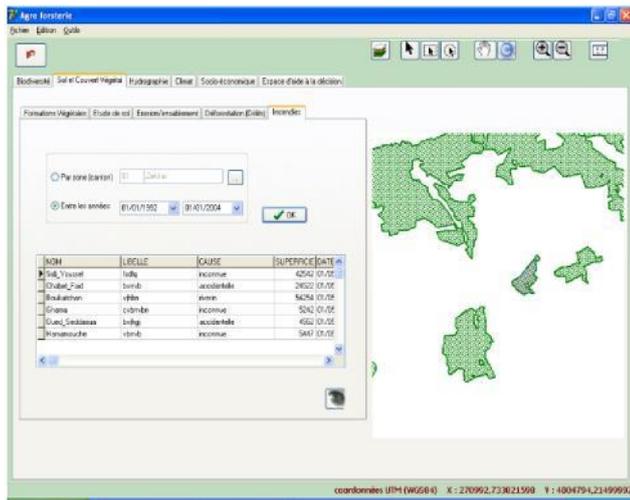


Fig 14. Fire management

C. The module “Handling cartographic compositions”

The model also offers the possibility of manipulation of cartographic objects that are in the form of digital georeferenced layers.

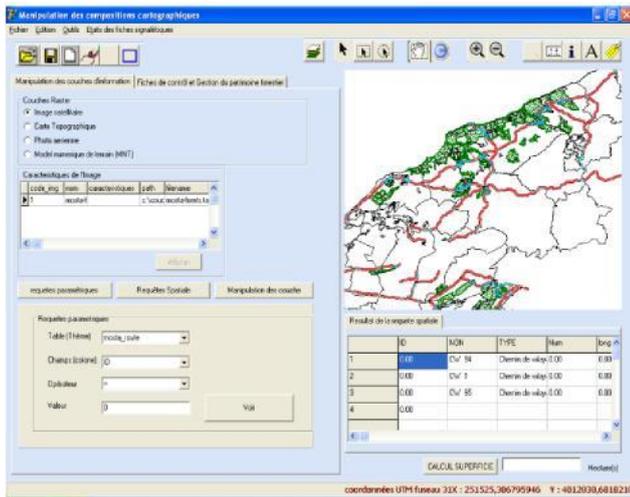


Fig 15. Cartographic Objects Manipulation

This interface allows the following tasks:

- Display results in map and alphanumeric of different spatial or parametric queries.
- Manipulation of raster information layers (topographic maps, aerial photography, satellite imagery, digital elevation model).
- See X and Y coordinate as the cursor on the map.
- Perform spatial queries, by selecting the subject of the application layer and the mode.



Fig 16. Spatial queries

IX. CONCLUSION

This paper presented spatial on line analytical processing as a platform built to support spatio-temporal analysis. So we have presented some actual research topics in this field, and then we have presented our prototype developed for agroforestry management purpose.

This work has allowed us to address the following:

- Improvement and optimization techniques for archiving of all data types;
- Demonstration of graphical objects (photos, maps, vector layers), awarded with the information available through the various application components, allows a good recovery from existing and preventing future spatial entities;
- A good user management of database and their privileges, and access rights through the security mechanisms of the Oracle DBMS.

It should be noted that a system based on the data warehouse, including spatial OLAP, based on the multidimensional paradigm, allows a multidimensional analysis of large amounts of data, thanks to the concepts of dimension, fact, measurement, hypercube and the various OLAP operators, in the aim of data sharing and security.

ACKNOWLEDGMENT

I am very grateful to all the members of LIO Laboratory (University of Oran) for the fruitful discussions about this type of research which has significantly contributed to improve the presentation of this work.

REFERENCES

[1] A. Brisebois. Analysis of the potential of extension SOLAP concept for the investigation of the three-dimensional spatial data (Analyse du potentiel d’extension du concept SOLAP pour l’exploration des données spatiales 3D). Master’s thesis, Laval University, Canada. 2004

[2] A. Polge, C. Barbe, P. Benoit : « Les systèmes d’information géographique (SIG) appliqués à la forêt : Les origines de la démarche régionale - Cohérence avec les réflexions et les acquis nationaux - Enjeux et perspectives », Regional Centre of the forester property of Bretagne, France, 2004.

- [3] A. SIMITSIS, «Modeling and Optimization of Extraction-Transformation-Loading (ETL) Processes in Data Warehouse Environments », Ph.D. Thesis, Athens, October 2004
- [4] A. Tchounikine, Miquel, M., Laurini, R., Ahmed, T., Bimonte, S., & Baillet, V.. Overview of works about spatio-temporal data integration into hyper-cubes (Panorama de travaux autour de l'intégration de données spatio-temporelles dans les hypercubes). *New Information Technologies Journal (Revue des nouvelles technologies de l'information)*, B-1, 21-33. 2005
- [5] C.A.Bensalloua, D. Hamdadou, B. Beldjillali, "SOLAP: A type of user interface and Security Oriented System for Agro-forestry management", 7th International Conference on Information Assurance and Security, Malaysia, 2011.
- [6] E. Bernier, and Y. Bédard, . "Using a data warehousing architecture to combine automatic generalization and multiple representation for Web-based on-demand mapping. In A. Ruas & W. McKinnis (Eds.), *Challenges in the portrayal of geographic information: Issues of generalisation and multi scale representation. (Forthcoming)*". 2005
- [7] M. Bakillah. « Développement d'une approche géosémantique intégrée pour ajuster les résultats des requêtes spatiotemporelles dans les bases de données géospatiales multidimensionnelles évolutives », Thesis to obtain the degree of Master of Science. Faculty of Forestry and Geomatics, Laval University, Quebec, 2007.
- [8] M. Bakillah, M.A. Mostafavi, Y. Bédard, « Développement d'une approche géosémantique intégrée pour ajuster les résultats des requêtes spatiotemporelles dans le domaine forestier ». In *Geomatics Symposium – «Au coeur des processus »*, October 25th-26th, Montreal, Canada, 2006.
- [9] M.Miquel, Y. Bédard, A. Brisebois, « Conception d'entrepôts de données géospatiales à partir de sources hétérogènes, exemple d'application en foresterie », *Journal of Information Syst èmes Ing énierie*, Vol. 7, No. 3, pp. 89-111, 2002.
- [10] M. Miquel, Y. Bédard, A. Brisebois, J. Pouliot, P. Marchand, J. Brodeur, «Modeling multidimensional spatio-temporal data warehouse in a context of evolving specifications», In *Proceedings of the theory, treatments and applications of geo referenced data*, Ottawa, 2002.
- [11] N. Dennouni : "Elaboration d'une base de données Géodésiques intégrant les nouvelles missions spatiales" Magister thesis in spatial techniques and applications, Spatial Techniques Centre, Algeria, 2004.
- [12] P. Marchand, . The spatio-temporal topological operator dimension, a hyperstructure for multidimensional spatio-temporal exploration and analysis. Doctoral thesis, Laval University, Canada. 2004
- [13] R. Chapis : "les bases de données Oracle8i développement, Administration et optimisation", Dunod, Paris, 2001.
- [14] R.Weibel, , & Dutton, G. (1999). Generalizing spatial data and dealing with multiple representations. In P. Longley, M. Goodchild, D. Maguire, & D. Rhind (Eds.), *Geographical information*.
- [15] S. C. Jason, "Oracle Certified Professional – DBO Certification Exam Guide", Oracle Press, USA, 2000.
- [16] S.O. Benoît: "Structures des données appliqués aux SIG - Notes de cours", Department of Geography, University of Quebec at Montreal, Canada, 2005.
- [17] S. Rivest, Y.Bédard, M.J. Proulx, M. Nadeau, F. Hubert and J. Pastor, "SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data". *ISPRS Journal of Photogrammetry & Remote Sensing*, 60, pp 17–33, 2005.
- [18] S. Rivest, Gignac, P., Charron, J., and Bédard, Y. Development of a spatio-temporal interactive data exploration system for the Information Data Bank of Ministry of Transportations, Quebec (Développement d'un système d'exploration spatio-temporelle interactive des données de la Banque d'information du ministère des Transports du Québec). Paper presented at Géomatique 2004, Canadian Institute of Geomatics, Montreal, Canada. 2004, October
- [19] S. Shekhar, Lu, C. T., Tan, X., Chawla, S., & Vatsavai, R. Map Cube: A visualization tool for spatial data warehouses. In H. Miller & J. Han (Eds.), *Geographic data mining and knowledge discovery* (pp. 74-109). London: Taylor & Francis. 2001
- [20] S. Urman, "oracle 8i Programmation avancé PL/SQL – construisez des applications PL/SQL puissantes et orientées web, Oracle Press, USA, 2001
- [21] V.Hernandez, Voss, A., and Gohring, W. Sustainable decision support by the use of multi-level and multi-criteria spatial analysis on the Nicaragua Development Gateway, From pharaohs to geoinformatics. In *Proceedings of the Fédération Internationale des Géomètres Working Week 2005 and GSDI-8* (pp. 16-21). 2005
- [22] Y.Bédard, (2005, May). Integrating GIS and OLAP: A new way to unlock geospatial data for decision-making. Paper presented at the Location Technology and Business Intelligence Conference, Philadelphia.
- [23] Y.Bédard, Gosselin, P., Rivest, S., Proulx, M. J., Nadeau M., Lebel, G. & Gagnon, M. F. (2003). Integrating GIS components with knowledge discovery technology for environmental health decision support. *International Journal of Medical Informatics*, 70(1), 79-94.
- [24] Y.Bédard, P. Gosselin, S.Rivest, M. Proulx, M. Nadeau, G. Lebel and M. F. Gagnon . Integrating GIS components with knowledge discovery technology for environmental health decision support. *International Journal of Medical Informatics*, 70(1), 79-94. 2003
- [25] Y. Bédard, Proulx, M. J., & Rivest, S. «OLAP improvement for geographic analysis: Examples of realizations and technological solutions (Enrichissement du OLAP pour l'analyse géographique: Exemples de réalisations et différentes possibilités technologiques) ». *New Information Technologies Journal (Revue des nouvelles technologies de l'information)*, B-1, 1-20. 2005
- [26] Y. Bédard, Sonia Rivest and Marie-Josée Proulx "Spatial.Online. Analytical.Processing. (SOLAP): Concepts, Architectures, and Solutions.from.a.Geomatics. Engineering.Perspective", Chapter.13 In Robert Wrembel and Christian Koncilia: "Data Warehouses and OLAP: Concepts, Architectures and Solutions", IRM Press 2007.

Author Biographies



ABDALLAH BENSALLOUA Charef, born in Mostaganem, 1979. Magister degree in Spatial Techniques and Applications -Remote Sensing at the National Centre of Spatial Techniques, 2008. PH.D Student in computer sciences at the University of Oran. Assistant Professor at the University of Mostaganem. Algeria.

Résumé

Spatial On-Line Analytical Processing (SOLAP) est une technologie qui offre des possibilités d'analyse spatiale et multidimensionnelle des données, produites par les techniques spatiales et les applications tels que les systèmes d'information géogra- phique (SIG), la télédétection, etc. Les données sont stockées dans des structures multidimensionnelles notamment les entrepôts de données spatiales (EDS), ce qui permet une organisation orientée sujet, intégrée, non volatile et variant dans le temps pour la prise de décision. Cependant, cette technologie est limitée dans l'aspect qualité de la décision relative aux considérations multicritères. A travers la présente thèse, nous proposons une approche méthodologique pour surmonter ces limites. Notre objectif est de proposer un système d'aide à la décision spatiale, à savoir Silvicultura, pour faciliter la prise de décision dans des situations complexes caractérisées par des données hétérogènes de sources multiples, et des objectifs souvent contradictoires. En effet, notre démarche décisionnelle est basée sur l'intégration de l'analyse multicritères (AMC) avec un système d'information décisionnel < Spatial Decisional Support System > (SDSS) notamment le SOLAP. Nous sommes, également, intéressés par la mise en place d'un système de contrôle de qualité d'analyse multidimensionnelle UIC-SOLAP à travers des contraintes d'intégrité relatives aux utilisateurs du SOLAP. La modélisation des deux systèmes est basée sur le langage de modélisation unifiée(UML). Ce dernier peut être facilement étendu pour la modélisation multidimensionnelle de l'entrepôt de données spatiales et pour traiter des aspects complexes spécifiques de tels systèmes. Afin de valider nos propositions, nous présentons une étude de cas réel pour examiner notre approche d'élaboration de système d'aide à la décision dans la gestion agroforestière en utilisant un jeu de données concernant les forêts de la wilaya de Mostaganem en Algérie.

Mots clés :

Aide à la Décision; Spatial On-Line Analytical Processing (SOLAP); Système d'Information Géographique (SIG); Entrepôt de Données Spatiales; Système d'Information Décisionnel Spatial; Analyse Spatiale; Analyse Multidimensionnelle; Analyse multicritères; Contrainte d'Intégrité; Modélisation UML; Gestion Agroforestière.