Vers des cubes matriciels supportant l'analyse spatiale à la volée dans un contexte décisionnel

Mémoire

Mathieu Plante

Maîtrise en sciences géomatiques
Maître ès sciences (M.Sc.)

Québec, Canada

© Mathieu Plante, 2014

Résumé

Depuis l'avènement du SOLAP, la problématique consistant à produire des analyses spatiales à la volée demeure entière. Les travaux précédents se sont tournés vers l'analyse visuelle et le calcul préalable afin d'obtenir des résultats en moins de 10 secondes. L'intégration des données matricielles dans les cubes SOLAP possède un potentiel inexploré pour le traitement à la volée des analyses spatiales. Cette recherche vise à explorer les avantages et les considérations à exploiter les cubes matriciels afin de produire des analyses spatiales à la volée dans un contexte décisionnel. Elle contribue à l'évolution du cadre théorique de l'intégration des données matricielles dans les cubes en ajoutant notamment la notion de couverture matricielle au cube afin de mieux supporter les analyses spatiales matricielles. Elle identifie des causes de la consommation excessive de ressources pour le traitement de ces analyses et propose des pistes d'optimisation basées sur l'exploitation des dimensions matricielles géométriques.

Table des matières

Liste des tableaux

Liste des figures

Remerciements

Ce projet de recherche a été possible grâce à la collaboration et au financement de la Chaire de recherche industrielle CRSNG en bases de données géospatiales décisionnelles, dirigée par le professeur Yvan Bédard. Je tiens à remercier l'organisme et ses membres pour leur soutien financier.

Je remercie mon directeur, Yvan Bédard, avec qui j'ai eu la chance de travailler durant ma maîtrise, pour son expertise, sa confiance, sa patience ainsi que toute la liberté qu'il m'a accordée pour réaliser et mener à terme ce projet. Grâce à cette recherche, mais surtout au cheminement qu'il sous-entend, j'ai élargi mes horizons et développé des qualités personnelles qui vont au-delà du cheminement académique.

Je veux aussi souligner mon appréciation aux professionnels de recherche de la Chaire et du département des sciences géomatiques pour leur assistance, leur soutien et leurs conseils. Également, un merci tout particulier à mes confrères étudiants entre autres à Eve, Véronique, Mathieu, Mojgan, Eugénie et Vincent sans qui cette recherche ne serait pas ce qu'elle est, car ils m'auront permis aux fils de discussions ouvertes et enrichissantes de faire évoluer ce projet sous sa forme actuelle. Un merci tout spécial à Eve, qui en plus d'avoir enduré l'exposition de mes élucubrations et de mes idées farfelues, est devenue une amie et une confidente qu'on chérit pour la vie.

Merci à mes parents, Serge et Lucie, à ma sœur Katherine, et à mon amoureuse, Nadia, pour avoir su trouver les mots justes pour m'encourager, et avoir traversé avec moi les moments forts des études graduées. Votre patience, votre compréhension et vos exhortations m'ont été d'un important support et ont été source de motivation.

Chapitre 1. Introduction

1.1 Mise en contexte

La structure de données OLAP est utilisée pour l'aide à la décision depuis le milieu des années 1990. Cette organisation de l'information est différente de l'architecture des bases de données traditionnelles. Elle vise à structurer les données selon des dimensions, représentant des axes d'analyse, possédant chacune une hiérarchie granulaire de l'information. Les thèmes analysés correspondent aux mesures (variables dépendantes) résultant du croisement d'une valeur ou de membres par dimension (variables indépendantes). Il en résulte ce qui est appelé des « faits », c.-à-d. les intersections uniques des membres et des mesures résultant de ces croisements. L'organisation des données autour d'une table de faits se nomme cube ou hypercube. Exploités initialement dans le domaine de l'informatique décisionnelle, plusieurs cubes de données OLAP possèdent maintenant la capacité de mettre à profit les données spatiales, ce qui permet d'obtenir des cubes spatiaux. L'application spatiale de l'OLAP, nommée SOLAP, fut développée pour exploiter les cubes spatiaux et se définit comme « un logiciel de navigation facile et rapide dans les bases de données spatiales qui offre plusieurs niveaux de granularité d'information, plusieurs thèmes, plusieurs époques et plusieurs modes de visualisation synchronisée ou non : cartes, tableaux et diagrammes » [Bédard, 2004]. Ce concept a évolué au point de devenir une spécialité avec son propre corpus théorique et sa communauté de chercheurs. L'exploration des données spatiales ainsi supportée permet de représenter visuellement les mesures OLAP sur une carte afin d'améliorer la compréhension de la distribution géographique des phénomènes. Or, la navigation simplifiée dans un logiciel SOLAP combinée à la structuration en dimensions des axes d'analyse offre la possibilité de valider plusieurs hypothèses avec une aisance remarquable. La technologie SOLAP est au cœur du domaine appelé « *Geographic Knowledge Discovery* ».

L'analyse spatiale visuelle n'est toutefois pas la seule façon d'obtenir de l'information statistique dans un espace géographique. De nombreux algorithmes d'analyse exploitent les données spatiales pour faire émerger les patrons et les tendances de distribution spatiale. La description statistique du territoire, les analyses locales ou globales des données et l'attribution d'une pondération à des attributs entrainent une analyse plus complète qu'une simple analyse visuelle. Il fut démontré, dans des travaux sur l'OLAP spatial, que cette structure de données supporte les relations spatiales [Rageul, 2007] et les opérateurs d'analyse spatiale [Brisebois, 2003]. Par ailleurs, il est possible de réaliser des analyses à la volée, en moins de 10 secondes, en conservant le fil de pensée des utilisateurs par l'utilisation d'une structure spatio-temporelle adaptée [Marchand, 2004]. De ce fait, nous considérons que SOLAP est un outil efficace et efficient pour l'aide à la décision par la combinaison des capacités de navigation simplifiées, des analyses spatiales visuelles et des opérateurs d'analyse à la volée.

1.2 Problématique

La principale force de la structure OLAP est de répondre rapidement à des interrogations multivariées. À elle seule, cette caractéristique explique en bonne partie l'intérêt de l'OLAP pour les applications d'aide à la décision. En effet, l'exécution à la volée ($t_{requête}$ < 10 secs pour rester à l'intérieur de la bande cognitive de Newell [Newell, 1990]) est une caractéristique *sine qua non* des outils logiciels visant à supporter sans interruption le fil de pensée de l'utilisateur et lui permettre de résoudre des problématiques de façon efficace [Marchand, 2004]. L'exploitation des cubes spatiaux OLAP n'échappe pas à cette condition. Ils doivent permettre l'interrogation ainsi que la représentation des données sous forme de cartes, de tableaux, de graphiques et de leurs différentes combinaisons conformément à ce prérequis de performance. Toutefois, lorsqu'il est nécessaire de produire des analyses spatiales avec des cubes spatiaux, il advient des situations où l'exécution à la volée n'est pas possible, c'est-à-dire qu'elles ne rencontrent pas la performance requise par la bande cognitive de Newell). Notamment, lorsque le volume de données à traiter est important en présence de nombreuses mesures et dimensions ou lorsque les algorithmes utilisés sont complexes (situation fréquente en analyse spatiale). Les travaux de Pierre Marchand ont démontré qu'il était impératif en 2004 de **calculer préalablement l'ensemble des croisements spatio-temporels possibles pour faire de l'exploration et de l'analyse spatio-temporelle (STEA) à la volée dans un contexte spatial vectoriel multidimensionnel**. Malgré l'évolution de la puissance des ordinateurs depuis cette époque, cette problématique existe toujours pour les analyses complexes de données non spatiales et est reconnue par l'ensemble de la communauté OLAP. Il va sans dire que **ce problème reste également entier pour les données spatiales**.

Afin de cerner cette problématique du calcul des analyses spatiales à la volée, il est nécessaire de mieux comprendre les processus de calcul. Ainsi, l'analyse spatiale est définie comme un « raisonnement qui permet de déduire les caractéristiques d'un phénomène en faisant intervenir des données à référence spatiale » [OQLF, 2004]. Ces processus analytiques de traitement des données spatiales utilisent des opérateurs pour produire des résultats. Les analyses les plus simples sont réalisées à partir d'opérateurs a) thématiques, b) métriques ou c) topologiques. D'autres opérateurs plus complexes sont en fait des compositions d'opérateurs plus simples. Or, ces **opérateurs, même parmi les moins compliqués, peuvent consommer une quantité importante de ressources lors de leur traitement sur des données vectorielles**. C'est le cas notamment pour les opérateurs vectoriels topologiques (voisinage immédiat ou sur des couches spatiales superposées). Toutefois, l'intégration des données spatiales dans des cubes de données ne se limite pas à utiliser des données géométriques vectorielles.

1.2.1 Intégration matricielle de l’aspect spatial

Les travaux sur le SOLAP matriciel [McHugh, 2008] ont démontré qu'il est possible d'intégrer l'aspect spatial dans les cubes via des données au format matriciel. L’analyse spatiale dans les données matricielles comporte des différences notoires par rapport aux opérations effectuées sur des données vectorielles. L’objet spatial de base, dans le paradigme vectoriel, est une représentation typiquement cartésienne finie (ex. polygone) d’une entité physique ou logique positionnée sur le territoire et possédant une définition sémantique pour l’utilisateur (ex. municipalité). L’analyse spatiale, sur ce type d’objet, applique une logique sémantique pour établir la nature ou mesurer la force des relations entre les objets (ex. rue *dans* municipalité). Le « quoi » prédomine le « où ». Or, l’objet spatial de base, dans le paradigme matriciel, constitue un élément unitaire du territoire divisé selon un maillage régulier (ex. cellule de 100 m x 100 m). Cette division du territoire est indépendante de l’interprétation sémantique des objets qui se trouvent sur le territoire (ex. municipalité, rue ou rivière). Le « où » prédomine le « quoi ». L’analyse spatiale sur la structure matricielle applique donc une logique géométrique ayant pour principale considération l’emplacement et la forme des cellules matricielles et leur emboîtement dans le cas de structure multi-granulaire. Les algorithmes d’analyse spatiale utilisés sur les structures matricielles sont conséquemment plus simples et rapides que ceux utilisés en structures vectorielles.

L’application des données matricielles aux cubes de données présente certains avantages. À priori, les structures OLAP permettent d’exploiter des niveaux de détail variables en fonction de l’emplacement des membres dans les hiérarchies des dimensions. Les membres situés près de la racine des arborescences des dimensions constituent une vue agrégée des sujets exploités. À l’opposé, les membres situés près des feuilles présentent un niveau plus détaillé de cette information. Appliqués aux dimensions spatiales, les membres près de la racine représentent une vue générale (grossière) du territoire et les membres près des feuilles une vue détaillée. Or, il est envisageable d’exploiter la hiérarchie spatiale dans ce type de dimension pour interroger *a priori* les niveaux les plus grossiers pour ensuite traiter l’information détaillée. Il suffit alors d’exploiter les relations topologiques d’inclusion des membres de la dimension pour interroger les membres agrégés préalablement aux membres détaillés, ou vice-versa. Les données matricielles possèdent une topologie beaucoup plus simple par la nature même de leur structure organisée en lignes et en colonnes ainsi qu’une nature multi-granulaire implicite.

L’autre avantage des données matricielles pour les cubes de données est lié à la visualisation dans les outils SOLAP. Il est reconnu [De Smith et *al.*, 2007] que les données matricielles sont volumineuses et exigeantes à traiter dans le contexte géospatial, et ce malgré leur simplicité relative comparativement aux données vectorielles. Évidemment, cette lourdeur compromet l’intérêt d’utiliser des données matricielles pour faire du traitement à la volée (grand volume de données vs simplicité de traitement). L’**application de structures**

matricielles au contexte géodécisionnel permet toutefois de réduire considérablement le volume de données d'intérêt pour les analyses. En effet, la carte géographique occupe une place de premier plan dans l'exploitation des données à l'aide d'outil SOLAP. Dans ce contexte où la visualisation de la carte est prépondérante, il est concevable que les analyses spatiales à la volée s'exécutent uniquement sur la partie visible (*cf.* fenêtre de visualisation) ou sélectionnée (*cf.* un ou quelques membres spatiaux) du territoire. Un changement de niveau dans une dimension spatiale ou une modification du contexte visible (ex. changement d'échelle ou déplacement) force l'exécution d'une nouvelle requête d'analyse spatiale. Les ressources nécessaires aux traitements des analyses s'en trouvent réduit puisque la zone visible limite l'étendue spatiale de l'analyse.

Considérant le contexte géodécisionnel et la structuration des données matricielles, **nous croyons que cette structure matricielle de l'information, appliquée dans des cubes de données, poserait un avantage pour la production d'analyses spatiales à la volée exploitant les hiérarchies spatiales des dimensions.** À cet égard, le traitement des analyses spatiales à la volée, dans le contexte des structures matricielles intégrées aux cubes de données, est envisageable et mérite d'être exploré plus attentivement.

1.2.2 Absence de comparaison vectorielle versus matriciel

Les concepts d'exploration et d'analyse des données dans des outils SOLAP furent étudiés à maintes reprises et la compatibilité de la structure multidimensionnelle avec les analyses spatiales fut démontrée [Brisebois, 2003]. L'exploitation à la volée, des résultats d'analyse spatiale, faisant suite au calcul préalable de l'ensemble des croisements spatio-temporels a également été réalisé [Marchand, 2004] ainsi que suivant la saisie explicite des relations spatiales [Rageul, 2007]. La presque totalité des travaux antérieurs ou en cours, également ceux concernant les agrégations spatiales à la volée, matérialise l'aspect spatial dans les cubes sous forme descriptive ou vectorielle. Dans tous les cas, l'intégration des géométries dans les cubes spatiaux fut concrétisée par l'utilisation de données vectorielles au niveau des mesures ou des dimensions spatiales. Cette observation vaut également pour les travaux récents portant sur les cubes matriciels [McHugh, 2008; Bernier, 2006; Brisebois, 2003] ayant simulé une tessellation régulière à partir de cellules vectorielles. Aucune étude décrivant si et comment les cubes matriciels peuvent être utilisés de manière complémentaire aux cubes vectoriels pour minimiser le calcul préalable des analyses spatiales n'est connue à ce jour. Ceci vaut pour l'ensemble de la littérature rencontrée sur les SOLAP et non seulement pour les travaux pionniers faits au Centre de Recherche en Géomatique (CRG). De ce fait, l**'incertitude concernant la capacité des cubes matriciels à répondre à l'exigence de performance d'exécution est très grande et difficile à quantifier**.

1.3 Objectifs

Le but de la présente recherche est d'explorer les facteurs qui, une fois réunis, pourraient offrir une alternative aux analyses spatiales vectorielles dans les structures de données SOLAP. Actuellement, l'exploration à la volée nécessite le calcul préalable des croisements spatio-temporels tel que fait par Pierre Marchand [Marchand, 2004]. L'analyse spatiale matricielle constitue possiblement une alternative au calcul préalable et ce fut une des motivations ayant déclenché la recherche théorique de M.Sc de Rosemarie McHugh [McHugh, 2008] sur la nature des cubes géospatiaux matriciels. Encore faut-il identifier les conditions spécifiques favorisant l'exécution à la volée des analyses matricielles dans une structure OLAP spatiale, *c.-à-d.* favorisant la rencontre de la limite cognitive de 10 secondes que nous considérons *sine qua non* à la conservation du flux de pensée lors d'exploration interactive de données. Une fois ces conditions identifiées, il sera possible de développer et tester de telles solutions, mais ceci va au-delà de l'objectif du présent mémoire de MSc.

La recherche en cours vise donc à **explorer les avantages et les considérations à exploiter les cubes spatiaux matriciels afin de produire des analyses spatiales à la volée** dans un contexte décisionnel. Plus spécifiquement, **deux sous-objectifs** cherchent à guider le processus de création de cubes spatiaux matriciels :

- Identifier des causes de la consommation de ressources informatiques par des analyses spatiales matricielles;
- Identifier des moyens d'optimiser les ressources pour le traitement des analyses spatiales matricielles.

L'atteinte de ces objectifs s'inscrit dans l'objectif global d'accroitre la base de connaissances spécifique à l'analyse des problématiques matricielles décisionnelles. La compréhension accrue des caractéristiques de l'exécution à la volée devrait permettre de proposer des pistes de solution visant à optimiser les ressources nécessaires au traitement des analyses spatiales matricielles dans le contexte posé.

1.4 Méthodologie

La réalisation de ce projet, l'état actuel des connaissances et l'atteinte des objectifs de cette recherche sollicitent l'application d'une méthode exploratoire. Cette méthode possède comme principal intérêt de ne pas limiter la recherche à des contraintes précises tout en évoluant à l'intérieur d'un cadre défini. L'incertitude concernant l'avancement de la formalisation des notions de cubes matriciels explique le choix de cette méthode, de même que la nature ouverte des sous-objectifs fixés. Rappelons que cette orientation de recherche en est à son tout début et qu'à notre connaissance, aucune autre équipe de recherche au monde ne s'y est attaquée. Il apparaît plus souhaitable que cette recherche s'attarde à identifier les causes de la consommation de ressources informatiques et les moyens de les optimiser pour le traitement des analyses spatiales matricielles dans des cubes OLAP spatiaux. À ce stade-ci, cette approche semble plus profitable que de s'efforcer de démontrer qu'il est possible avec un algorithme et des données choisies de réaliser un temps de réponse suffisamment court pour atteindre les exigences de performance. Cette réflexion est à la base de la présente recherche et oriente la démarche entreprise pour l'atteinte des objectifs du projet. Malgré notre optimisme, il n'était pas possible *a priori* d'évaluer si les concepts théoriques concernant les cubes matriciels, développés dans des travaux précédents, étaient suffisamment avancés pour se lancer dans l'implantation d'analyses spatiales matricielles à la volée. Conséquemment, il fallait dans un premier temps bien comprendre les enjeux et choisir un problème précis, parmi ceux qui émergeaient, auquel serait confrontée la présente recherche de MSc. Conformément à la méthode choisie, la démarche employée est inductive et les objectifs se sont raffinés itérativement avec l'avancement de cette recherche.

La première portion de la recherche a donc consisté à acquérir les connaissances théoriques nécessaires à la compréhension approfondie de la problématique. C'est pourquoi une revue de littérature s'imposait sur les sujets traitant : des couvertures matricielles; des structures multidimensionnelles; des analyses spatiales matricielles; des indexations spatiales et des interrogations des données géospatiales. Le processus de revue de littérature n'est pas un évènement ponctuel et demeure une activité se déroulant tout au long de la recherche. Toutefois, la majorité des lectures furent effectuées au début du processus afin de développer une base de connaissances nécessaire à la poursuite de l'activité de recherche. L'exploration des concepts a permis de formaliser les enjeux de la production **d'analyses spatiales à la volée dans un contexte géodécisionnel. En fonction de l'accroissement de la base de connaissances, ceci a permis de mieux définir le projet à chaque itération et l'amener à la version actuelle.**

Par la suite, une analyse détaillée de cubes spatiaux vectoriels [Bernier, 2008; Bernier et *al.*, 2008; Proulx et *al.*, 2006] et matriciels [Brisebois, 2003; Bernier, 2006; Chaker W. et *al.*, 2009; McHugh, 2008], des données spatiales matricielles (couvertures matricielles) et des analyses spatiales matricielles [Jensen, 2005; De Smith et *al.*, 2007] ont permis d'identifier des facteurs influençant la consommation de ressources et les temps de traitement des analyses. La quatrième étape visait à mettre en relation ces concepts afin d'identifier et de synthétiser les principales causes de la consommation des ressources dans le traitement des analyses (sous-objectif #1) dans le contexte des cubes matriciels. De façon plus détaillée, les sous-étapes consistaient en :

- Choisir le type de dimension spatiale à étudier;
- Choisir le type de cube à étudier;
- Intégrer la notion de couverture matricielle;

- Développer des scénarios pour fins d'analyse;
- Définir les enjeux de l'analyse spatiale matricielle à la volée;
- Développer les scénarios d'intégration de couvertures matricielles et des critères d'élaboration;
- Sélectionner les scénarios les plus favorables;
- Apprécier la capacité d'un scénario à supporter l'analyse spatiale matricielles à la volée;
- Déterminer les scénarios de hiérarchisation des couvertures;
- Identifier les formes de hiérarchies les plus pertinentes;
- Identifier des pistes d'optimisation des analyses spatiales matricielles.

Enfin, la dernière étape consistait à élaborer des recommandations pour favoriser le traitement à la volée des analyses spatiales matricielles (sous-objectif #2) et aussi à étendre le processus de création des cubes spatiaux qui seraient destinés à ce type d'exploitation. De façon plus détaillée, les sous-étapes consistaient en :

- Concevoir des algorithmes spécialisés d'analyse;
- Proposer une stratégie de calculs d'indicateurs statistiques vs calculs des résultats;
- Déterminer les caractéristiques d'un index de cube matriciel;
- Déterminer les conditions optimales pour l'exécution à la volée des analyses spatiales matricielles.

Finalement, les optimisations proposées devraient permettre d'orienter les travaux futurs, faisant suite à cette recherche fondamentale, pour la mise en œuvre de cubes matriciels destinés à l'analyse spatiale matricielle à la volée. Le diagramme d'activité ci-dessous illustre le processus entrepris dans ces travaux théoriques de recherche.

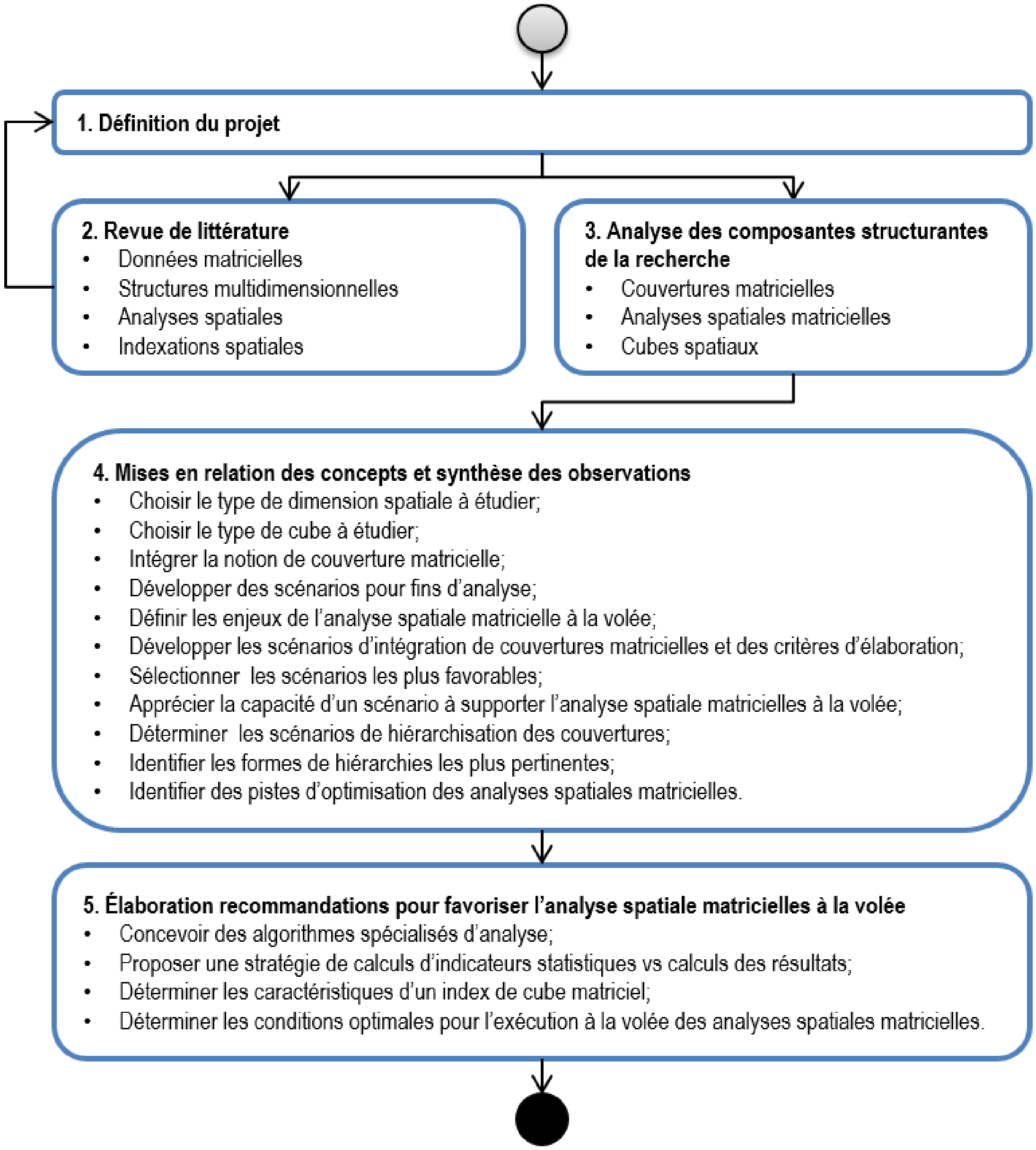


Figure 1.1 — Méthode exploratoire selon une démarche inductive

Enfin, ce projet de recherche présente les enjeux théoriques de la modélisation spatiale matricielle multidimensionnelle et de l'exécution à la volée des analyses spatiales matricielles. La démarche inductive suivie consiste à généraliser des cas particuliers observés et d'en dégager des hypothèses, des modèles et même des lois générales. Les pistes de solution présentées dans le cadre de ces travaux, notamment au chapitre 4 portant sur des scénarios d'optimisation, devront faire l'objet d'expérimentation afin d'être confirmées ou infirmées lors de travaux futurs.

1.5 Conclusion

Le présent chapitre aura permis d'exposer le contexte du projet et d'exprimer la problématique de la recherche. Au moment de l'écriture de ce mémoire, l'analyse spatiale vectorielle à la volée, dans les structures OLAP, s'apparente plus à de l'interrogation à la volée de résultats calculés préalablement qu'à du traitement à la demande des algorithmes d'analyse. Cette situation engendre un manque de flexibilité et de souplesse de la structure pour répondre à des besoins imprévus lors de la conception du cube. Les données matricielles possèdent des caractéristiques singulières qui pourraient permettre l'exécution à la volée des analyses spatiales lorsque certaines conditions sont réunies. Ce travail de recherche tente d'identifier les causes de la consommation de ressources informatiques par les analyses spatiales matricielles et d'identifier des moyens d'optimiser les ressources pour le traitement de ces analyses. Le chapitre suivant démontre les faits saillants de la revue de littérature ayant servi à alimenter le processus de recherche. Le chapitre 3 approfondit les aspects de la théorie de la modélisation matricielle multidimensionnelle et explore les causes de la consommation des ressources des analyses spatiales matricielles. Le chapitre 4 propose des moyens d'optimiser les traitements favorisant l'exécution à la volée en fonction de la demande en ressources pour le traitement des analyses. Finalement, le chapitre 5 présente les discussions et perspectives pour des travaux futurs concernant l'exécution à la volée d'analyses spatiales dans des cubes matriciels.

Chapitre 2. Revue des concepts

L'intégration des données matricielles dans les cubes supportant les analyses spatiales à la volée repose sur trois concepts clés : les données spatiales matricielles; les cubes spatiaux matriciels et les analyses spatiales matricielles. Le présent chapitre expose ces concepts afin de créer la base de connaissances nécessaire pour appuyer la suite de la démarche. En premier lieu, il sera question des concepts entourant les données spatiales matricielles. Ensuite, les méthodes pour intégrer les données matricielles dans les structures multidimensionnelles spatiales seront explorées. Finalement, une revue de l'analyse spatiale matricielle et des caractéristiques essentielles à ce type d'analyse clora ce chapitre.

2.1 Données spatiales matricielles

L'intérêt de cette recherche se porte sur les données matricielles dans un contexte géographique décisionnel. La section suivante s'attarde à définir le jeu de données matricielles spatiales tel que perçu dans le cadre de référence de cette recherche. Dans un premier temps, les caractéristiques essentielles des jeux de données matricielles sont brièvement présentées. Dans un deuxième temps, il est question des couvertures continues matricielles, notion essentielle qui est au cœur de ces travaux.

2.1.1 Structure matricielle

Le domaine d'application SOLAP, pour livrer la performance recherchée, requiert l'utilisation de grilles régulières aux cellules de forme, de taille et d'orientation homogène aussi appelées tessellations régulières. Une tessellation régulière est un découpage continu de l'espace en une collection de polygones réguliers contigus de même forme et de même taille. En présence d'une grille comportant des polygones identiques (côtés et angles intérieurs égaux), les seules primitives géométriques possibles sont : les triangles, les carrés et les hexagones. La tessellation régulière formée de carrés est particulière en ce sens qu'elle est la seule qui peut être décomposée ou agrégée en objets de même forme et de même orientation [Laurini et Thompson, 1992]. C'est également celle qui utilise les algorithmes d'analyse les plus simples. Par ailleurs, les cellules carrées offrent une plus grande flexibilité fonctionnelle dans le sens où l'ensemble des algorithmes d'analyse spatiale en mode matriciel est basé sur ce type de tessellation [Tomlin, 1990; McHugh, 2008]. Les **tessellations régulières carrées** feront donc l'objet de notre attention particulière pour la suite de ce mémoire.

Les données matricielles régulières carrées constituent un jeu de données structuré en lignes et en colonnes à la façon d'une matrice (M) au sens mathématique du terme. L'agencement de ces lignes et de ces colonnes produit une grille. L'espace formé par l'intersection de lignes et de colonnes forme une cellule matricielle. Dans le contexte de la géomatique, la cellule matricielle est associée à une portion de territoire. Cette caractéristique intrinsèque oriente l'ensemble des analyses et des interprétations qui sont faites des données matricielles par la suite. La cellule contient une **valeur numérique** associée à la thématique exprimée par le jeu de données (ex. : humidité et température du sol). L'analyse des valeurs numériques permet d'interpréter la thématique en tenant compte de la position de la cellule ainsi que des valeurs et des positions des cellules voisines. Du moment où la donnée matricielle est considérée géospatiale, cette valeur numérique est associée à une thématique géolocalisée.

En plus de posséder une **position cartésienne** relative au jeu de données (intersection ligne/colonne), la donnée matricielle géospatiale possède une **coordonnée géographique** (x, y, z). La taille d'une cellule en rapport à la réalité décrite permet de déterminer la **résolution spatiale** de la cellule. Dans plusieurs jeux de données, la résolution spatiale précise est une valeur propre à chaque cellule puisque cette valeur peut varier à l'intérieur du jeu de données matricielles. En effet, les données matricielles provenant de la télédétection ou de modèles numériques de terrain présentent parfois des déformations des cellules en fonction des caractéristiques du capteur, des phénomènes physiques ou du relief. Par exemple, les données matricielles satellitaires présentent une déformation de perspective se matérialisant par des cellules plus petites près du nadir (position située directement au-dessous du capteur) et plus grande selon l'éloignement de cette position. La figure ci-dessous décrit cette situation de déformation en fonction de la perspective du nadir du capteur. Toutefois, dans les cubes de données SOLAP, les données matricielles proviennent souvent de la rastérisation des cartes vectorielles et leur résolution est présumée uniforme pour l'ensemble du territoire couvert (malgré les distorsions dues au choix de la projection cartographique, à l'ellipsoïde de référence, au relief et aux capteurs).

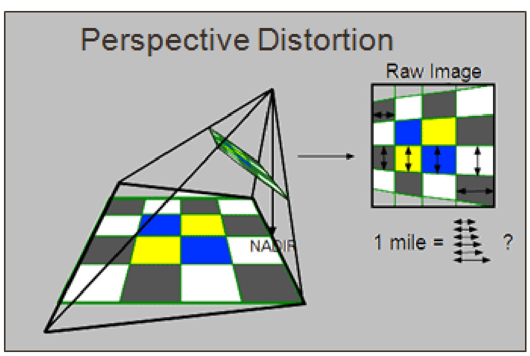


Figure 2.1 — Distorsion de la perspective par éloignement du NADIR[1]

[1] *Source : http://www.cardiofx.com/media/img/perspectiveDistortion.gif*

Les données matricielles possèdent aussi un **domaine de valeurs** numériques. Il sert à définir l'ensemble des valeurs admissibles pour décrire la thématique. La **modalité** sert à exprimer les valeurs thématiques observées dans un jeu de données matricielles. Les cellules possèdent également des **métadonnées** associées notamment à l'acquisition, aux transformations et à la qualité des données.

Enfin, en géomatique il existe deux méthodes pour représenter plusieurs thématiques pour un même territoire. La première consiste à superposer plusieurs grilles matricielles identiques et superposées à l'intérieur d'un même jeu de données matricielles. Ces grilles matricielles sont des **couches matricielles** partageant le même espace géographique, mais exprimant des thématiques différentes. La seconde consiste à associer plusieurs valeurs numériques (sous forme d'un tableau ou d'un vecteur) aux cellules matricielles qui composent le jeu de données. Cette dernière méthode est celle la plus couramment utilisée en géomatique.

2.1.2 Couverture matricielle

Les sciences géomatiques proposent l'utilisation des couvertures pour décrire les phénomènes continus sur le territoire. Un **phénomène continu** est une caractéristique observable sur l'ensemble du territoire. La température, l'humidité et l'utilisation du sol sont des exemples de phénomène continus. Le modèle de couverture proposé par la norme ISO 19123 défini la couverture comme suit : « *abstraction d'un phénomène du monde réel agissant telle une fonction retournant une valeur thématique de son domaine de valeurs pour n'importe quelle coordonnée, à l'intérieur d'un système de référence spatiale, situé dans son domaine spatial, temporel ou spatio-temporel (traduction libre)* [ISO19123, 2005]. » L'interrogation d'une coordonnée de la couverture retourne un objet spatial (géométrie) possédant des attributs. Le domaine spatial est composé d'une collection de géométries. Le type et la configuration de ces géométries identifient le sous-type de la couverture. Le recours à des couvertures discrètes ou continues est lié au phénomène ou à la thématique à représenter à l'aide de la couverture. S'il s'agit d'un phénomène continu sur le territoire, par exemple la température ou l'utilisation du sol, la couverture continue est plus appropriée. Selon la norme, les couvertures continues se présentent en cinq sous-classes distinctes. La figure suivante présente ces différentes sous-classes de couvertures continues.

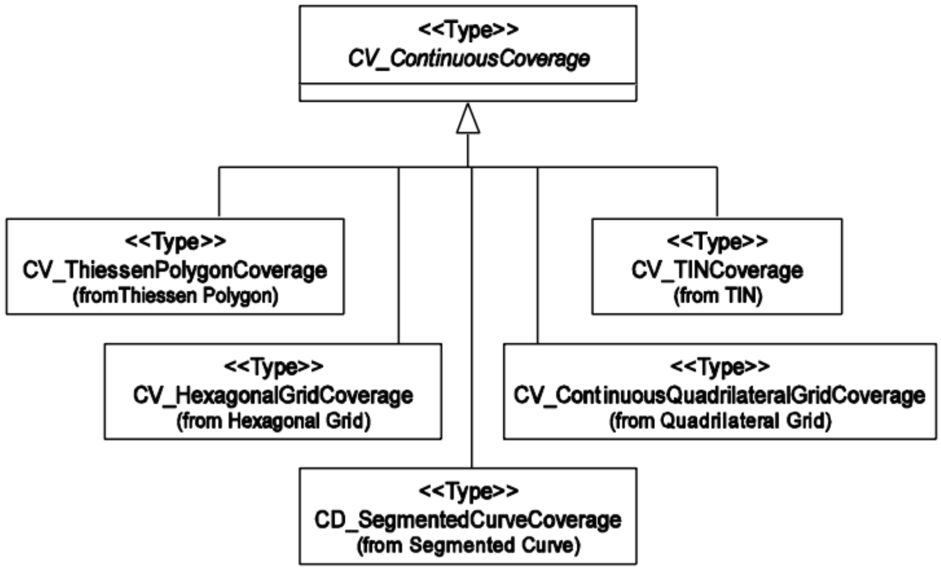


Figure 2.2 — Sous-classes de couvertures continues[2]

De ces différents types de couvertures continues, la **grille quadrilatérale continue** ou « *continuous quadrilateral grid coverage* » est la seule permettant de représenter des tessellations régulières carrées. Ce type de couverture continue se caractérise par un découpage systématique du domaine spatial, temporel ou spatio-temporel. Des courbes quadrillent le domaine dans une orientation horizontale et verticale. La grille formée présente des courbes qui s'intersectent les unes et les autres. L'espace créé entre des courbes verticales et horizontales contigües s'intersectant forme les cellules de la grille. Les cellules démontrent une forme géométrique associée aux parallélogrammes ou aux parallélipipèdes. Toutefois, les grilles quadrilatérales ne forment pas systématiquement des tessellations régulières carrées. Pour former une tessellation régulière carrée, une grille quadrilatérale doit être composée de courbes droites et linéaires. Ces courbes doivent être parallèles et les intersections formées par les courbes doivent être perpendiculaires. De plus, l'espacement entre ces courbes doit être le même afin que la résolution spatiale soit constante sur toute la grille. L'ensemble de ces conditions permet de créer une tessellation régulière carrée à l'aide d'une couverture en grille quadrilatérale.

Ceci dit, la **notion de fonction** associée aux couvertures continues [ISO19123, 2005] est primordiale. Elle implique qu'à l'intérieur du domaine de la fonction, toutes les coordonnées sont associées à un objet spatial. Conséquemment, il ne saurait y avoir d'objet spatial manquant dans la couverture. Ceci dit, il peut y avoir absence de valeur thématique, mais pas absence de géométrie (cellule) dans une couverture. Cette notion de

[2] [ISO 19123, 2005]

continuité est essentielle pour la résolution à la volée des analyses spatiales matricielles et sera explorée plus en détail au chapitre suivant. Dans le cadre de ce mémoire, par souci de concision, de clarté et pour les besoins du SOLAP 2D, le nom **couverture matricielle** sera employé pour désigner expressément ce type de couverture. Les caractéristiques essentielles d'une couverture matricielle sont les suivantes :

1. interrogation d'une coordonnée (ligne, colonne) retourne un objet spatial unique;
2. toutes les cellules possèdent une géométrie;
3. toutes les géométries sont carrées, de même taille et de même orientation;
4. le domaine de valeurs est homogène à l'intérieur d'une couche.

Du point de vue de la performance des requêtes d'analyse spatiale, la **principale caractéristique de l'intérêt de la couverture quadrilatérale continue réside dans le fait qu'elle supporte une énumération séquentielle** des éléments du domaine spatial. De ce fait, l'entreposage et l'accès aux données sont plus efficaces que pour tout autre type de couverture [ISO19123, 2005]. Ce type de couverture est compatible avec la schématisation multidimensionnelle en ce sens qu'une grille est matérialisée pour chaque paire de dimensions géographiques (x, y, z). La coordonnée d'une cellule s'obtient par une séquence de nombres où chaque nombre désigne la position d'un point dans son domaine dimensionnel respectif formant un système de référence. La figure ci-dessous présente le système de coordonnées des couvertures quadrilatérales.

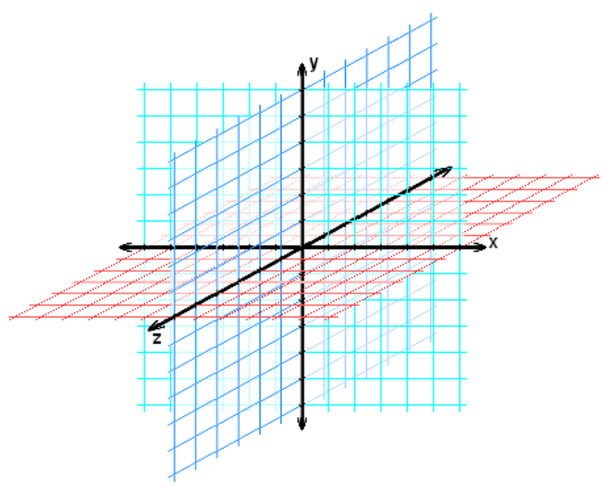


Figure 2.3 — Système de coordonnées cartésien pour les couvertures quadrilatérales

Le fait d'ajouter une ou plusieurs dimensions au jeu de données se fait par l'ajout de variables au système de référence. Ainsi, un jeu de données matricielles spatio-temporel est exploitable à l'aide d'une fonction dont les paramètres sont X, Y, Z et T où T représente le temps. La compatibilité des couvertures matricielles avec la représentation multidimensionnelle est une condition essentielle pour l'utilisation de ce type de structure de données dans les cubes spatiaux.

2.2 Matriciel dans les structures multidimensionnelles décisionnelles

Le contexte géodécisionnel du projet porte à exploiter les structures multidimensionnelles spatiales. Ces structures, référées par le terme cube spatial, sont des cubes OLAP pour lesquels des membres, des dimensions ou les faits (par le biais de mesures spatiales) sont référencés spatialement et peuvent être représentés sur des cartes [Bédard et *al.*, 2008]. Les caractéristiques fondamentales des cubes OLAP spatiaux sont discutés dans de nombreux ouvrages [Bédard et *al.*, 2001, 2004, 2008; Rivest et *al.*, 2003; McHugh, 2008] et ne seront pas présentés plus en détail dans ce document. Il a été démontré que les cubes spatiaux sont des structures de données capables de supporter les données matricielles [McHugh, 2008]. Ces recherches ont permis de distinguer deux types de cubes spatiaux, les cubes d'objets et les cubes matriciels. Cette section présente une revue de l'intégration matricielle dans les cubes spatiaux par l'intermédiaire des dimensions et des faits.

2.2.1 Intégration du matriciel dans les dimensions

Une dimension spatiale est une dimension, au sens prévu par les structures OLAP, dont un ou plusieurs membres peuvent être localisés sur le territoire. *A priori*, la littérature distingue les dimensions géométriques des dimensions non géométriques [Han et *al.*, 1998; Bédard et *al.*, 2001, 2008; Rivest et *al.*, 2003]. Les dimensions non géométriques sont des dimensions spatiales pour lesquelles l'aspect spatial ne provient pas de primitives géométriques, mais plutôt d'une référence nominale. Les dimensions géométriques sont des dimensions pour lesquelles tous les membres, de tous les niveaux possèdent une primitive géométrique. Il peut exister des dimensions mixtes où les membres ne possèdent pas tous des primitives géométriques.

Il existe une distinction concernant les modèles de données spatiales associés aux membres des dimensions pour l'intégration des données matricielles. Les modèles spatiaux vectoriels et matriciels, c'est-à-dire le type de géométrie associée aux membres, permettent de modéliser des dimensions spatiales additionnelles. Il est alors question de dimensions spatiales simples lorsque tous les membres d'une dimension sont du même type soit matriciel, vectoriel ou non géométrique. Les dimensions spatiales mixtes sont celles intégrant à la fois des membres géométriques (vectoriels ou matriciels) et des membres non géométriques. Les dimensions spatiales hybrides sont celles matérialisant des membres géométriques des domaines matriciels et vectoriels. L'ensemble des types dimensionnels spatiaux est illustré ci-dessous.

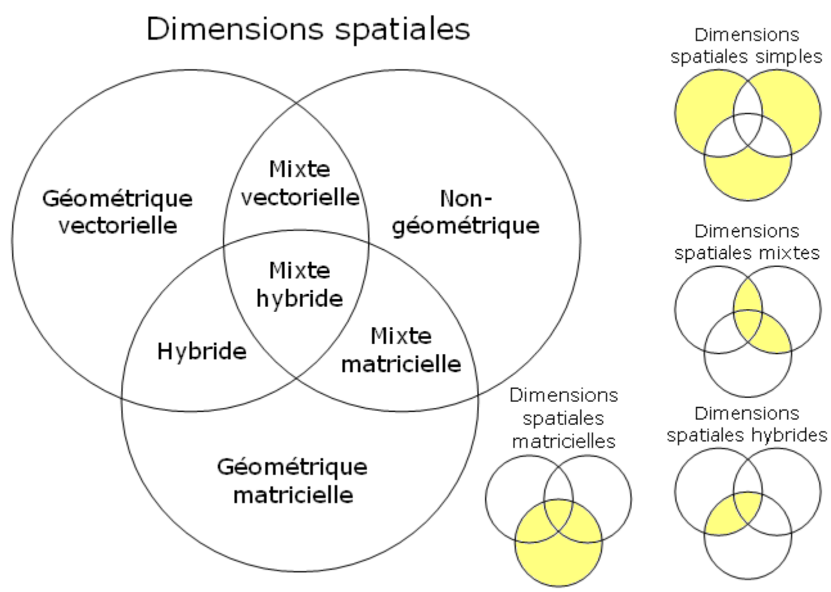


Figure 2.4 — Les types de dimensions spatiales[3]

Spécifiquement, les dimensions spatiales intégrant des données matricielles sont les suivantes [McHugh, 2008] :

a) **Dimension géométrique matricielle** : Dimension spatiale géométrique constituée d'au moins un niveau matriciel et si le nombre de niveaux est supérieur à un, différentes résolutions géométriques sont employées pour former la hiérarchie. La plus grande résolution est utilisée pour le niveau le plus détaillé et les résolutions agrégées sont des multiples entiers de la résolution du niveau détaillé.
b) **Dimension géométrique hybride** : Dimension spatiale géométrique constituée d'au moins un niveau matriciel et d'au moins un niveau vectoriel. Elle ne doit pas contenir de niveau non géométrique. L'ordre des niveaux n'a pas de conséquence. La résolution la plus fine est associée au niveau vectoriel ou matriciel sans distinction ou importance.
c) **Dimension mixte matricielle** : Dimension spatiale constituée d'au moins un niveau matriciel et d'au moins un niveau nominal. Aucun niveau vectoriel n'est admis. L'ordre des niveaux n'a pas d'importance. La résolution la plus fine est associée à un niveau matriciel ou nominal sans distinction ou importance.
d) **Dimension mixte hybride** : Dimension spatiale constituée d'au moins trois niveaux. Elle contient au moins un niveau nominal, au moins un niveau vectoriel et au moins un niveau matriciel. L'ordre des niveaux n'a pas d'importance. La résolution la plus fine est associée à un niveau matriciel, vectoriel ou nominal sans distinction ou importance.

[3] [McHugh, 2008]

Les dimensions géométriques hybrides et les dimensions hybrides mixtes sont particulières en ce sens qu'elles admettent dans certains cas des relations « non strictes ». Une relation non stricte, au sens prévu dans les modèles de bases de données multidimensionnelles, comporte une relation multiple (cardinalité N. N) entre des membres de deux niveaux d'une dimension. Ces cas se produisent lorsque des niveaux vectoriels sont mis en relation avec des niveaux matriciels et que la relation spatiale établit qu'un objet vectoriel ne contient pas entièrement un objet matriciel. La figure suivante illustre une relation non stricte entre un niveau vectoriel et un niveau matriciel.

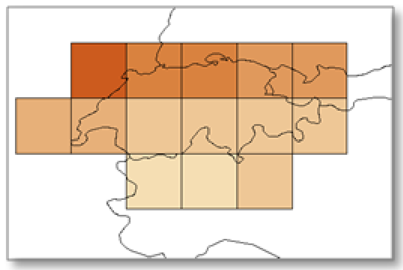


Figure 2.5 — Relation spatiale entre la Suisse et les cellules qui la composent[4]

Or, certaines dimensions hybrides (géométriques ou mixtes) permettent de conserver les relations strictes entre les membres des niveaux vectoriels et matriciels. C'est le cas notamment lorsque les géométries vectorielles sont générées par agrégation de cellules matricielles. La relation est alors stricte (1..N) entre les objets vectoriels et les objets matriciels. La figure suivante illustre une relation stricte entre un niveau vectoriel et un niveau matriciel.

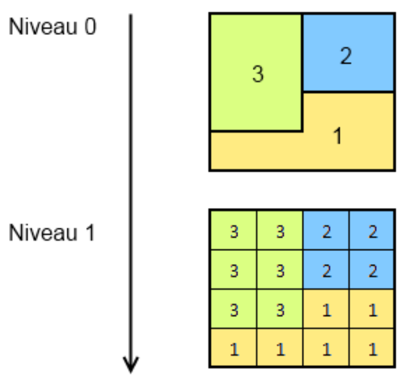


Figure 2.6 — Relation spatiale stricte entre les cellules (matricielles) et les polygones (vectoriels)

[4] [Proulx et coll. 2008]

Observer le degré de recouvrement spatial entre les membres des niveaux lors du traitement des analyses spatiales matricielles augmente les ressources nécessaires au traitement et complexifie les algorithmes.

Afin de ne pas ajouter une charge supplémentaire au traitement des analyses spatiales matricielles, et aussi afin de garder l'ampleur de la présente recherche conforme aux exigences d'un mémoire de M.Sc., **seulement les dimensions spatiales géométriques matricielles** [figure ci-dessous] seront considérées pour la suite de ce mémoire.

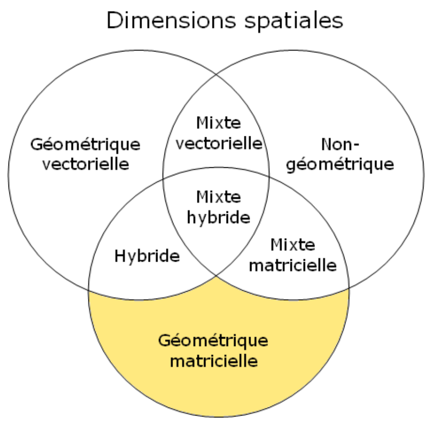


Figure 2.7 - Dimension spatiale géométrique matricielle

La figure présente les différents types de dimensions spatiales [McHugh, 2008] et met en relief la portée des travaux de recherches effectués dans le cadre de ce mémoire.

2.2.2 Intégration du matriciel dans les faits du cube spatial

L'intégration des données matricielles dans les dimensions spatiales, tel que présenté ci-avant, constitue une des approches pour intégrer ce modèle de données spatiales dans les cubes. L'autre approche consiste à intégrer les données matricielles dans les faits. Pour ce faire, une cellule matricielle est associée à chaque fait du cube. Rappelons qu'un fait est l'objet associé au croisement unique d'un membre pour chaque dimension et d'une ou plusieurs mesures dans le cube. Le fait devient alors un fait spatial, puisque la cellule matricielle possède une géométrie par association à sa mesure. La cartographie du fait est alors possible à l'aide de cette géométrie.

Les faits détaillés constituent le niveau de détail le plus fin d'un cube de données. Un fait agrégé est le résultat d'une agrégation statistique effectuée sur des faits détaillés. Les agrégations sont des calculs statistiques tels la somme, la moyenne, le maximum et le minimum effectués sur les données. L'agrégation peut être de nature spatiale, temporelle ou descriptive. L'agrégation descriptive s'effectue sur la valeur numérique ou qualitative associée au fait. Ensuite, l'agrégation spatiale utilise des opérateurs statistiques considérant la localisation, les relations sémantiques et topologiques pour former des agrégats de cellules. Enfin, l'agrégation temporelle utilise des opérateurs logiques pour recréer des groupes temporels, la plupart du temps, selon un calendrier grégorien ou fiscal. Les règles d'agrégation sont établies à partir des hiérarchies des dimensions du cube. Dans le cas où aucune dimension spatiale n'est présente dans le cube, la taille des cellules matricielles sera la même pour l'ensemble des faits matriciels du cube. Il y aura une relation de cardinalité (1..N) entre les membres thématiques agrégés ou détaillés et les faits matriciels géométriques. En présence d'une dimension spatiale dans le cube, l'agrégation spatiale des cellules matricielles dépend de la résolution spatiale associée à la dimension spatiale ou à ses composantes (hiérarchie, niveau, membre). La hiérarchisation matricielle sera abordée plus en détail au chapitre 3.

2.2.3 Cubes matriciels

Les travaux portant sur l'intégration des données matricielles ont permis de distinguer deux types de cubes spatiaux : les cubes d'objets et les cubes matriciels [McHugh, 2008].

> **Cube d'objets** : Cube dont les faits détaillés sont des objets discrets possédant ou non une géométrie. Ceci correspond aux cubes que l'on rencontre habituellement. La géométrie est associée aux faits par la correspondance avec les géométries des membres de dimensions spatiales ou par l'entremise de mesures spatiales.
>
> **Cube matriciel** : Cube dont les faits détaillés sont des cellules d'une grille matricielle. Pour ce type de cube, le fait possède toujours une géométrie intrinsèque. La géométrie est associée aux faits par l'entremise de mesures spatiales ou par la correspondance avec les géométries des membres de dimensions spatiales.

Ces deux modèles de cubes peuvent implémenter l'aspect matriciel. Toutefois, ces cubes se distinguent par le type d'objets spatiaux associé aux faits détaillés et par l'objectif d'analyse poursuivi dans ce type de structure de données [McHugh, 2008]. L'objectif d'analyse par couverture, étudiant la variation spatiale d'un phénomène sans limites particulières, est associé aux cubes matriciels. L'analyse spatiale matricielle poursuit ce même objectif. D'autre part, dans les cubes matriciels « […] les faits ont toujours une géométrie et cette géométrie peut-être propre au fait ou correspondre à la géométrie d'un membre spatial » [McHugh, 2008]. Cette obligation de matérialiser une géométrie pour chaque fait est compatible avec la notion de fonction associée à la couverture matricielle définie précédemment (cf. section 2.1.2). De l'autre côté, le cube d'objet a pour objectif de permettre l'analyse d'objets discrets et ne matérialise pas nécessaire une géométrie matricielle pour chaque fait. **Pour la suite de cette recherche, les cubes matriciels feront l'objet de notre attention particulière pour l'atteinte de nos objectifs.**

2.3 Analyses spatiales matricielles

Le terme analyse spatiale désigne un « raisonnement qui permet de déduire les caractéristiques d'un phénomène en faisant intervenir des données à référence spatiale » [OQLF, 2004]. Il s'agit d'une discipline à part entière, possédant son propre cursus théorique. Elle intègre des méthodes d'évaluation et d'interprétation des relations, de l'organisation, de la répartition et de la position sur le territoire d'évènements ou de phénomènes. L'objectif du processus d'analyse spatiale est de comprendre, quantifier, modéliser et ultimement prédire en exploitant l'espace ainsi que le temps comme axes d'analyse. La figure suivante [De Smith et *al.*, 2007] décrit les étapes du processus générique d'analyse spatiale matriciel.

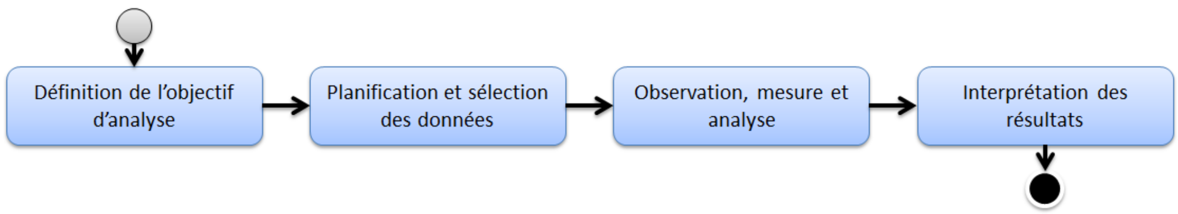


Figure 2.8 — Processus générique d'analyse spatiale

Plus spécifiquement, l'analyse spatiale matricielle désigne l'ensemble des processus d'analyse spatiale exploitant les jeux de données matricielles sous forme d'images, d'orthophotos ou de couvertures matricielles. L'analyse spatiale matricielle n'exclut pas l'intervention de données vectorielles ou nominatives, mais nécessite la présence de données matricielles dans son processus en intrant ou en extrant. La section qui suit présente les opérateurs d'analyse spatiale matricielle et décrit la portée des fonctions d'analyse. Finalement, le balayage des couvertures matricielles pour la résolution des analyses est revu en détail.

2.3.1 Opérateurs d'analyse spatiale matricielle

L'origine des méthodes, des outils et des algorithmes d'analyse spatiale matricielle provient du traitement numérique des images issues des communautés de la photographie et du traitement d'images [Worboys, 1995]. Depuis ces origines, différents algorithmes spécifiques aux données géolocalisées ont été mis au point. D'autre part, le langage Map Algebra, à l'aide de fonctions simples supportant la composition et la combinaison, permet d'interroger et d'analyser les données matricielles par couche [Tomlin, 1990]. Également, il ne faut pas négliger l'apport du domaine fréquentiel et des mathématiques entourant l'acoustique (Fourier, Laplace, etc.) pour la mise au point de filtres spatiaux permettant le rehaussement de l'image qui modifie la valeur des pixels en fonction de la valeur des pixels avoisinants afin de rehausser les zones de haute ou basse fréquence spatiale. D'autres algorithmes spécialisés sont utilisés en analyse spatiale matricielle. Ils permettent notamment de mieux comprendre la distribution des phénomènes, l'autocorrélation spatiale (ex. : I de Moran et C de Geary) et de produire des interpolations (ex. : poids inverse à la distance et Krigeage). Cependant, ces algorithmes, tous aussi spécialisés soient-ils, reposent sur des fonctions qui elles sont composées d'opérateurs élémentaires d'analyse spatiale appliqués aux données matricielles.

Dans l'espace 2D, l'analyse spatiale s'obtient à partir d'opérateurs thématiques, métriques et topologiques. Ces opérateurs simples peuvent être combinés pour créer des opérateurs complexes notamment les opérateurs de proximité, de superposition, d'intersection, de « *clustering* », etc.

Opérateurs thématiques

Les opérateurs thématiques sont indépendants des composantes spatiale et temporelle. Ils visent à sélectionner des objets selon leurs attributs. Ils appliquent des fonctions logiques, arithmétiques, mathématiques ou statistiques pour opérer une sélection et procéder à des calculs. L'intervention de ce genre d'opérateur signifie généralement un balayage de l'ensemble des données une cellule à la fois.

Opérateurs métriques

Les opérateurs métriques utilisent la composante géométrique de la cellule. Ils visent à mesurer la taille, la position, l'orientation ou la forme des objets spatiaux. Des exemples d'application sont le calcul de la distance entre deux cellules ou la longueur d'une frontière entre deux conglomérats de cellules. Un moins grand nombre de cellules sont utilisées pour ce type d'opérateur puisque l'opération s'effectue seulement sur un sous-ensemble des données. « Métrique » signifie « qui se mesure » et conséquemment les paramètres variables ainsi que le résultat nécessitent une unité de mesure. Typiquement, un changement de système de référence spatiale occasionne une modification au résultat. Autrement dit, le résultat d'opérations métriques est variable en fonction du système de référence utilisé.

Opérateurs topologiques

La topologie traite de ce qui est invariant lorsqu'on déforme ou change le système de référence spatiale, par exemple l'adjacence, l'inclusion, l'intersection et l'égalité. Aucune unité de mesure n'est requise pour exprimer le résultat. Les opérateurs topologiques utilisent théoriquement la relation spatiale unissant les objets (quoiqu'en pratique, les SIG utilisent les informations métriques pour construire la topologie avant d'effectuer les opérations topologiques). Il est donc possible d'obtenir des informations sur l'intersection, la contiguïté, la disjonction, l'inclusion ou l'égalité à partir d'objets spatiaux et d'opérateurs topologiques. Les données matricielles simplifient grandement les problématiques de performance liées à ce genre d'opérateurs, car la topologie de ce type de donnée est explicitement exprimée en termes de lignes et de colonnes. Ainsi, la cellule à la ligne 4 colonne 4 a comme voisins immédiats les cellules de la ligne 4, colonnes 3 et 5 puis les cellules des lignes 3 et 5 de la colonne 4. Aucun opérateur métrique n'est nécessaire pour effectuer cette analyse spatiale et la réponse va demeurer invariable même si on déforme ou change le système de référence spatiale.

Les algorithmes d'analyse spatiale utilisent rarement les opérateurs décrits précédemment de manière isolée. Ils sont plutôt le résultat de combinaisons ou de compositions d'opérateurs et utilisent parfois les résultats d'autres analyses sous forme de chaînes de traitement. Ces chaînes de traitement sont parfois itératives et incrémentielles exploitant les résultats d'autres analyses thématiques, temporelles ou spatiales (matricielles ou vectoriels).

2.3.2 Portée des fonctions d'analyse spatiale matricielle

Un algorithme d'analyse spatiale est composé de fonctions élémentaires. Chacune de ces fonctions interroge, interprète, transforme et traite un ensemble de cellules matricielles. Une fonction possède une portée matricielle propre, c'est-à-dire qu'elle exploite un nombre fini de cellules matricielles pour produire un résultat. Un bon nombre de ces fonctions sont récursives et exploitent l'ensemble ou un sous-ensemble de cellules de la couverture matricielle. Afin de caractériser la portée d'un algorithme d'analyse spatiale matricielle, il est nécessaire d'analyser la portée des différentes fonctions qui les composent.

La classification des fonctions élémentaires pour le traitement des données matricielles [Tomlin, 1990] permet de synthétiser la notion de portée. La classification catégorise les fonctions selon la portée matricielle exploitée :

Fonction locale

Le résultat de la fonction est obtenu par le traitement d'une seule cellule à partir d'une ou plusieurs couches matricielles superposées. L'addition de cellules matricielles pour former une couche composite est un exemple de fonction locale.

Fonction focale

Le résultat de la fonction est obtenu par le traitement d'une cellule et des cellules voisines à partir d'une ou plusieurs couches matricielles superposées. Le voisinage est restreint aux cellules contigües à la cellule d'origine ou étendue selon un paramètre de distance. Les filtres par convolution constituent un exemple de fonction focale.

Fonction zonale

Le résultat de la fonction est obtenu par le traitement des cellules composant une zone donnée à partir d'une ou plusieurs couches matricielles superposées. La zone, de taille et de forme variée, est définie à l'aide d'un objet vectoriel ou d'une couche matricielle. Le calcul de température moyenne pour une région est un exemple de fonction zonale.

En complément de cette classification, une fonction globale, est une fonction zonale dont la portée est celle de la couverture matricielle fournie en intrant. C'est donc nécessairement l'ensemble des cellules qui sont traitées pour produire un résultat dans une fonction globale. La figure suivante [Câmara et *al.*, 2005] illustre des types de fonction provenant de Map Algebra (fonction locale, focale et zonale).

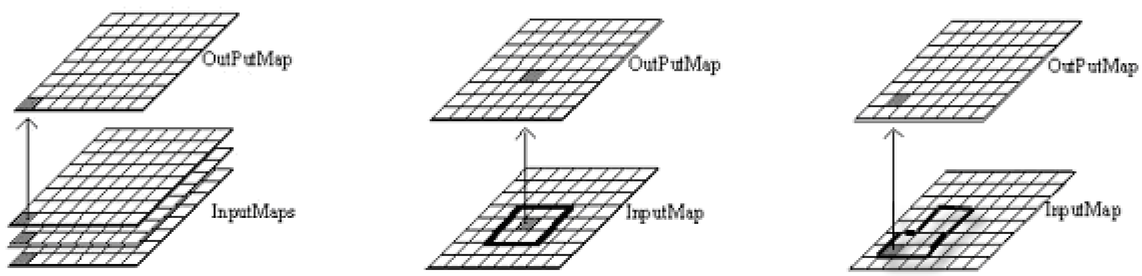


Figure 2.9 — Types de fonctions de Map Algebra

La mise en œuvre de la récursivité dans les algorithmes d'analyse spatiale fait en sorte qu'une fonction peut être répétée de façon itérative et incrémentale sur l'ensemble d'une couche ou d'une couverture matricielle. Par exemple, un filtre par convolution est appliqué localement, sur une zone ou sur l'ensemble des cellules du jeu de données matricielles. De ce fait, il est nécessaire de distinguer la portée d'une fonction et la portée d'un algorithme d'analyse spatiale matricielle. Considérant qu'une analyse spatiale est le résultat de la composition d'algorithmes, qui sont eux-mêmes des compositions de fonctions élémentaires, la portée d'une analyse spatiale correspond à l'union des ensembles formés par les portées des algorithmes et des fonctions qui la compose. Cette portée a une incidence sur le temps nécessaire au traitement de l'analyse spatiale matricielle et sera abordée plus en détail à chapitre 4.

2.3.3 Balayage des couvertures matricielles

Dans le domaine vectoriel, le balayage des objets géographiques se résume généralement à parcourir de façon séquentielle ou à l'aide des relations topologiques les objets spatiaux en interrogeant leurs caractéristiques et attributs afin de les analyser. Or, le domaine matriciel offre lui aussi une grande diversité de méthodes de balayage qui ne sauraient être comparées d'un domaine à l'autre. Le balayage, dans le contexte des données matricielles, consiste à parcourir de façon ordonnée et cohérente l'ensemble des cellules d'une couverture en observant les caractéristiques de chaque cellule. Ces caractéristiques, tel qu'énoncé précédemment, sont les attributs thématiques, le positionnement relatif dans le jeu de données, la géométrie de la cellule et ses métadonnées. En fonction des algorithmes d'analyse spatiale, le balayage considère les cellules matricielles comme des entités ponctuelles indépendantes ou dans d'autres cas comme des objets interactifs. Les interactions entre les cellules sont de nature spatiale ou thématique. La forme d'interaction la plus commune est de nature topologique et s'intéresse au voisinage de la cellule. Au sens large, ce sont les algorithmes d'analyse spatiale qui dictent la façon dont les données doivent être balayées pour obtenir le résultat anticipé.

Dans le domaine de l'imagerie numérique, il n'est pas rare d'utiliser un balayage linéaire ou en colonne. C'est notamment le cas pour les algorithmes élémentaires de chargement d'image, de compression (RLE) et les formats d'échange (BMP, PNG, GIF). Les cellules sont lues une à la fois et une séquence cohérente est formée pour en permettre le traitement. En présence de plusieurs couches matricielles superposées, comme c'est le cas pour les canaux RVB (Rouge, Vert, Bleu) des images en couleurs, la façon d'entreposer les couvertures joue un rôle très important pour le balayage des cellules matricielles et l'extraction des valeurs thématiques. En effet, les formats de données numériques matricielles communs sont organisés en intervalle par pixel (BIP « *band interleaved by pixel* »), par ligne (BIL « *band interleaved by line* ») ou en format séquentiel (BSQ « *band sequential format* »). La figure ci-dessous présente ces différents formats.

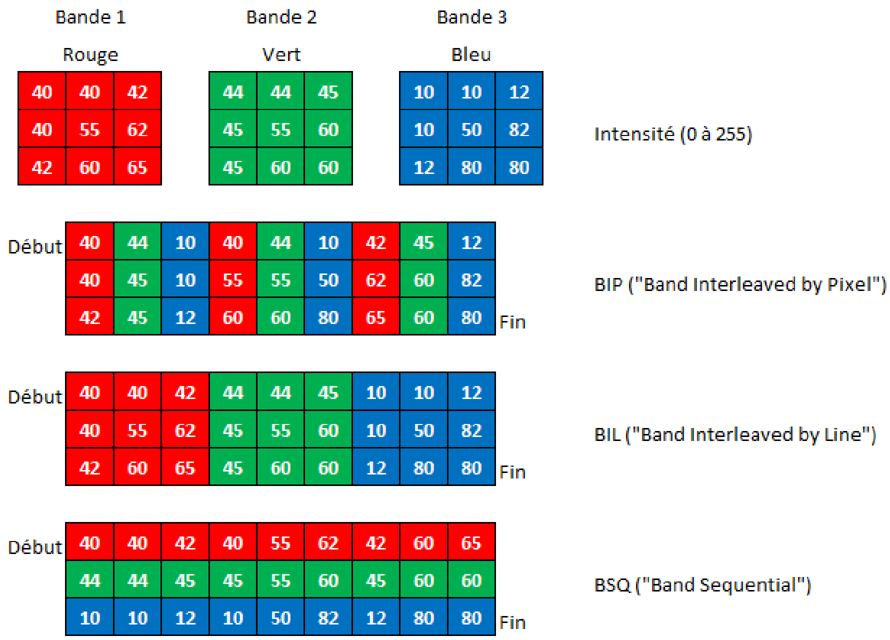


Figure 2.10 — Codage des bandes matricielles[5]

Toutes les analyses spatiales matricielles parcourent un sous-ensemble de cellules pixel par pixel. Cependant, l'assemblage sous forme de groupe de cellules est nécessaire pour produire les résultats de certains types d'analyse spatiale matricielle. C'est le cas notamment pour les analyses d'interpolation, de variogramme et de krigeage qui utilisent une ligne entière ou une colonne d'un jeu de données matricielles pour produire un résultat. Dans les cas de balayage linéaire, chaque pixel est lu, un à un, de façon séquentielle avant de passer à la ligne suivante. La figure suivante présente quelques méthodes de balayage des jeux de données matricielles.

[5] [Jensen, 2005]

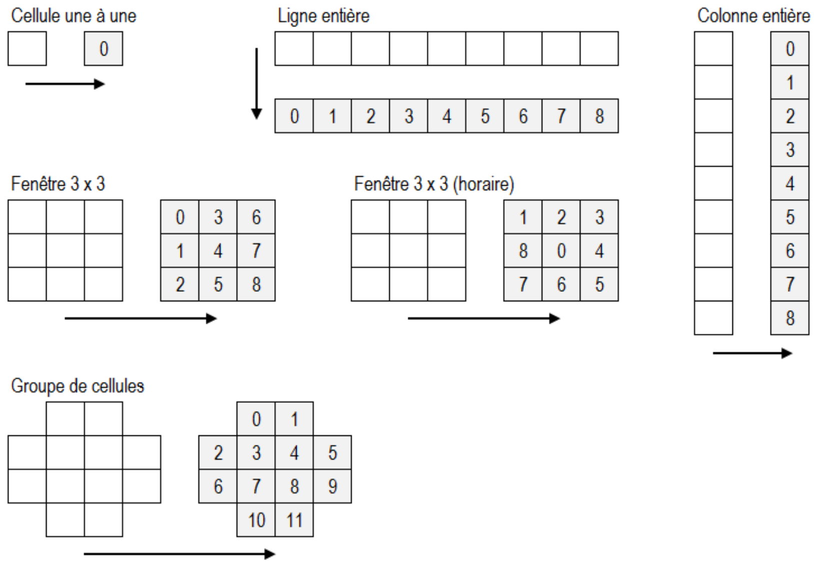


Figure 2.11 — Méthodes de balayage des données matricielles

Le balayage linéaire ou en colonne possède un inconvénient majeur pour les couvertures matricielles notamment lorsqu'il est question de mesurer l'autocorrélation spatiale. Ces types de balayage entrainent une distance logique informatique potentiellement très élevée entre deux cellules qui sont voisines d'un point de vue géospatial. La figure suivante illustre le balayage linéaire de gauche à droite ou en lacet.

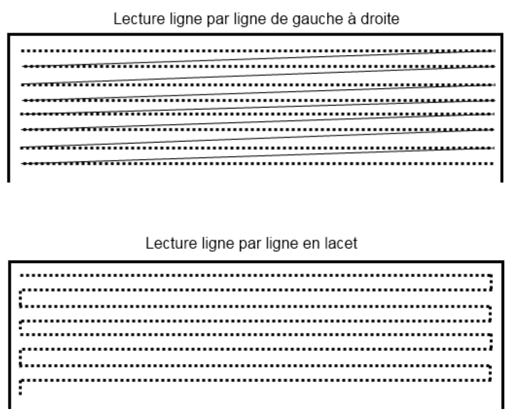
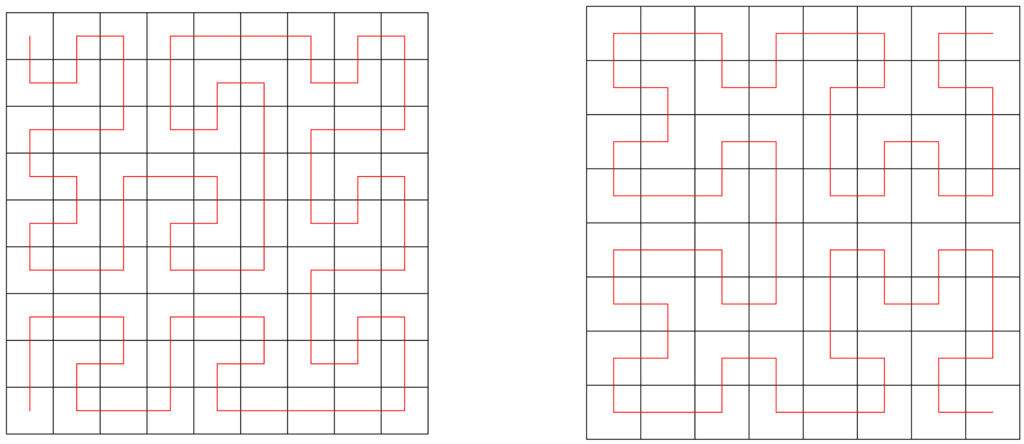


Figure 2.12 — Exemples de balayage linéaire

En réponse à cette problématique, les courbes de remplissage ou « *space-filling curves* » ou courbes de Peano sont utilisées pour parcourir les données matricielles. « Ces courbes s'avèrent être d'une grande utilité pour balayer complètement les espaces et servir de base à l'indexation spatiale » [Laurini Thompson, 1992]. Les courbes de Peano ou de Hilbert peuvent être utilisées pour les applications multidimensionnelles et sont compatibles avec les hypercubes [Laurini Milleret-Raffort, 1993]. La figure suivante illustre un exemple des courbes de remplissage de Peano et de Hilbert.



Peano – Second order space filling curve

Hilbert – Third order space filling curve

Figure 2.13 — Exemples de courbes de remplissage (Peano et Hilbert)

Pour leur part, les filtres par convolution et les analyses basées sur l'inverse de la distance utilisent des fenêtres de pixels à la manière des fonctions focales [Tomlin, 1990]. Ces fenêtres formées d'un nombre impair de cellules par côté, c'est-à-dire 3 x 3, 5 x 5 ou 7 x 7 cellules sont très communes pour les filtres d'atténuation, de rehaussement et de lissage dans le traitement des analyses d'image de télédétection. Ces procédés peuvent également être appliqués à des données interprétées pour rehausser certaines caractéristiques du jeu de données thématiques. À l'intérieur d'une même fenêtre de 3 x 3, il peut exister une séquence de lecture différente qui aura pour effet de varier les résultats finaux. Le filtre Laplacien utilise une fenêtre de 3 x 3 et ne possède aucune restriction pour l'ordre de lecture des cellules de la fenêtre. Alors que la séquence de lecture pour l'extraction de texture « *texture unit* » [Wang et He, 1990] doit être fait en sens horaire en débutant par le pixel supérieur à gauche dans une fenêtre de 3 x 3 tel qu'illustré par la figure ci-dessous.

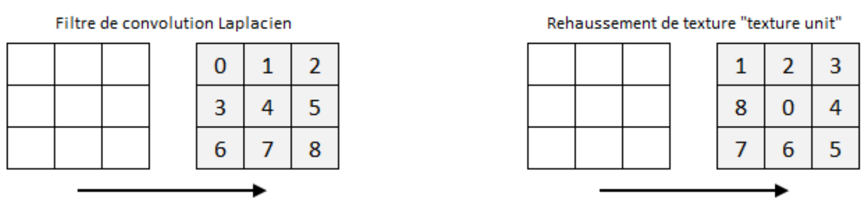


Figure 2.14 — Filtres exploitant des fenêtres de 3 par 3

Les algorithmes d'analyse spatiale matricielle exploitant des groupes de pixels formant des zones, parcourent généralement les cellules, une à une, une ligne à la fois de façon itérative partant du plus petit index vers le plus grand jusqu'à ce que toute la zone soit couverte.

Enfin, la notion de couverture matricielle, telle que définie (*cf.* section 2.1.2), est importante pour les analyses spatiales matricielles. En effet, les fonctions et algorithmes d'analyses exploitent les valeurs thématiques véhiculées par les cellules matricielles. Lors du balayage du jeu de données matricielles, la valeur thématique peut être nulle ou manquante [De Smith et *al*, 2007], mais la cellule elle-même est toujours définie. C'est la notion de fonction, associée à la couverture matricielle, qui prend toute son importance lorsqu'il est question de produire des analyses spatiales matricielles. La notion de fonction matricielle prévoit qu'en tout temps il existe une cellule matricielle associée à une coordonnée formée par le croisement de ligne et de colonne. Cette notion sera mise en perspective au chapitre 3 (*cf.* section 3.2.1) lorsqu'il sera question d'intégration des couvertures matricielles dans les structures de données OLAP spatiales.

2.4 Conclusion

Ce chapitre présentait les bases de cette recherche en abordant les notions essentielles liées aux données matricielles, l'intégration de ces structures dans les cubes OLAP spatiaux et explore les caractéristiques des analyses spatiales matricielles. Il faut rappeler que les données matricielles sont exploitées pour différents usages plus ou moins éloignés du domaine géodécisionnel. Une cellule matricielle, lorsqu'elle est utilisée dans un contexte géographique, porte beaucoup plus d'information que sa contrepartie mathématique. Une couverture matricielle constitue une spécialisation des jeux de données matricielles utilisées en imagerie et en télédétection. Les couvertures matricielles, telles que définies dans ce chapitre (*cf.* section 2.1), possèdent une richesse singulière arborant une multitude d'opportunités d'exploitation.

L'intégration des données matricielles dans les cubes spatiaux et de la façon de la réaliser est au cœur de ce chapitre. L'intégration des données matricielles géolocalisées, dans les structures OLAP, se fait par l'association de cellules matricielles aux faits du cube. Cette association est possible par la matérialisation de dimensions spatiales comportant des cellules matricielles ou à l'aide de mesures géométriques matricielles. L'observation de certaines caractéristiques, notamment le niveau de détail le plus fin du cube, permet de caractériser le type de cube matriciel dont il est question.

Finalement, les analyses spatiales matricielles furent explorées. L'objectif de ce type d'analyse est d'améliorer la compréhension de l'organisation, de la répartition, de la position et les relations entre les phénomènes sur le territoire. Afin de quantifier ou qualifier les évènements et les phénomènes, les algorithmes d'analyse spatiale matricielle ont recours à des opérateurs d'analyse et à des fonctions spatiales. C'est la combinaison de ces fonctions et de ces opérateurs qui permet d'extraire l'information pertinente des jeux de données matricielles. Le balayage permet de décrire pour chaque fonction la séquence d'interrogation des cellules matricielles afin d'extraire le résultat. Le chapitre suivant propose d'approfondir les aspects d'intégration des données matricielles dans les cubes spatiaux. La notion de couverture matricielle, telle que présentée dans le chapitre précédent, est au cœur des propos du chapitre à venir.

Chapitre 3. Intégration du matriciel dans les cubes matriciels

3.1 Introduction

Le chapitre précédent a permis d'identifier les notions essentielles à la conduite de cette recherche. Les notions de couvertures matricielles, d'analyse spatiale matricielle et de cubes matriciels furent notamment explorées. Les travaux antérieurs portant sur l'intégration des données matricielles dans les cubes spatiaux ont démontré qu'il existe différentes façons d'intégrer ce type de données dans les structures OLAP spatiales. Ils se sont toutefois attardés aux données matricielles au sens large sans accorder une importance particulière à la notion de couverture matricielle. Or, cette notion est pourtant essentielle aux analyses spatiales matricielles si nous voulons les exécuter à la volée (*cf.* section 2.1, 2.2 et 2.3). Dans une telle optique, il est donc nécessaire de revisiter le cadre théorique de l'intégration des données matricielles dans les structures OLAP spatiales afin d'y intégrer formellement la notion de couverture matricielle telle que décrite au chapitre précédent.

Le présent chapitre vise à étendre les concepts d'intégration des données matricielles dans les cubes spatiaux pour y inclure les considérations particulières aux couvertures matricielles en vue d'effectuer des analyses spatiales à la volée. En premier lieu, deux problématiques liées à l'analyse spatiale matricielle à la volée dans les cubes matriciels seront exposées. Ensuite, il sera question des façons d'intégrer les couvertures matricielles dans les cubes matriciels. Nous allons analyser différents scénarios pour ensuite les comparer quant à leur potentiel pour l'analyse spatiale à la volée. Ainsi, nous serons en mesure de mieux répondre à des questions telles que « est-ce que tous ces scénarios possibles permettent de mettre en œuvre des couvertures continues nécessaires à l'analyse spatiale matricielle? Est-ce que le volume de données matricielles produites serait réaliste? » Également, il sera question de la façon de hiérarchiser l'information spatiale et thématique des couvertures matricielles pour l'intégration de ces structures dans les dimensions. Finalement, une étude du modèle physique des hiérarchies mènera à étendre le cadre décisionnel pour inclure des considérations particulières à la hiérarchisation des couvertures matricielles.

3.2 Enjeux spécifiques de l'analyse spatiale matricielle à la volée dans les cubes matriciels

L'exécution à la volée des analyses spatiales matricielles dans les cubes matriciels est le point central de cette recherche. La réduction de la taille des jeux de données, en l'occurrence des cubes matriciels, constitue une approche généralement privilégiée en optimisation afin d'atteindre les performances d'exécution requises. Par exemple, un cube trop volumineux peut être partitionné en enlevant des dimensions, des niveaux hiérarchiques ou des membres. Conceptuellement, ces stratégies de partitionnement consistent à :

- ✓ Remplacer un cube spatio-temporel par deux cubes, un spatial uniquement où on enlève l'axe temporel et un temporel uniquement où on enlève l'axe spatial;
- ✓ Découper une hiérarchie temporelle en regroupant dix (10) années de façon à créer plusieurs cubes : un cube par année et un autre regroupant toutes les années;
- ✓ Produire deux (2) cubes portant sur cinq (5) années au lieu d'un seul cube incluant dix (10) années.

Toutefois, cette approche limite la capacité d'analyse du cube en réduisant le nombre de faits, de membres ou d'axes d'analyse.

Nous croyons qu'il existe d'autres options complémentaires afin de tendre vers le traitement à la volée des analyses spatiales matricielles. L'optimisation du traitement des analyses spatiales matricielles dans les cubes matriciels sera abordée dans le chapitre 4 de ce document. Toutefois, avant d'optimiser le traitement des analyses spatiales matricielles dans les cubes matriciels, certaines problématiques fondamentales de ces structures, par rapport aux analyses matricielles à la volée, doivent être soulevées.

3.2.1 Discontinuité du domaine spatial des cubes matriciels

Les algorithmes d'analyse spatiale matricielle sont conçus pour des couvertures matricielles. Or, pour ces algorithmes, la continuité est une caractéristique *sine qua non* d'une couverture matricielle. Cette caractéristique doit être observable sur l'ensemble de la couverture analysée. Par contre, un cube matriciel, même s'il est conçu pour faire des analyses continues (*cf.* section 2.2), ne possède pas toujours une continuité sur l'ensemble de son domaine spatial (étendue spatiale) lors de son utilisation. En effet, les fonctionnalités SOLAP permettent de sélectionner ou filtrer les éléments spatiaux et ainsi engendrer une discontinuité sur le domaine spatial de la couverture matricielle résultante. La discontinuité, dans ce type de structure, résulte d'opérations qui filtrent les géométries des cellules au même titre que les valeurs thématiques provenant de la mesure. La figure suivante illustre un exemple de filtre sur les cellules matricielles produisant une discontinuité de la couverture matricielle.

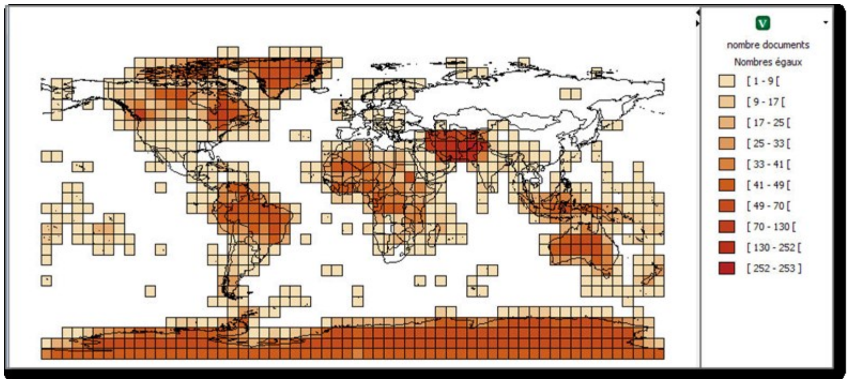


Figure 3.1 — Filtre sur la couverture mondiale pour les documents de date de validité > 2005[6]

Cette discontinuité possible du domaine spatial des cubes matriciels n'est pas favorable à une exécution à la volée des analyses spatiales matricielles. Dans le contexte de cette recherche[7] se posent alors deux scénarios possibles consistant à intervenir en amont, c.-à.-d. avant que la discontinuité soit observable, ou en aval, à savoir lorsque la discontinuité de la couverture est observée.

En intervenant en aval de la discontinuité du domaine spatial, il est possible de recourir à une *transposition* des résultats du paradigme multidimensionnel vers le paradigme de couverture matricielle. Cette opération consiste à extraire les valeurs thématiques pertinentes du cube et à les insérer dans une couverture matricielle[8]. En se référant à la figure précédente, le processus de *transposition* initialise une couverture matricielle avec des valeurs thématiques non-significatives (valeur nulle ou égale à zéro). Ensuite, les valeurs thématiques du sous-ensemble discontinu de cellules (nombre de documents) sont copiées dans cette nouvelle couverture continue. La figure suivante illustre cette étape de transposition d'un modèle multidimensionnel à un modèle de couverture matricielle.

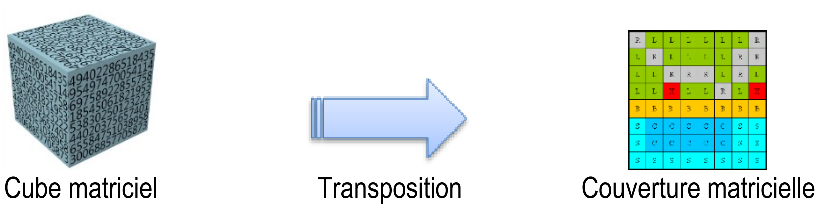


Figure 3.2 — Transposition du modèle multidimensionnel au modèle de couverture matricielle

---

[6] [Proulx et *al.,* 2008]

[7] Portée de la recherche qui ne vise pas à traiter des fonctionnalités SOLAP

[8] Cette couverture matricielle doit posséder des caractéristiques équivalentes (résolution spatiale, domaine de valeurs) et une étendue spatiale égale au rectangle minimal englobant « *MBR* » exprimé par la requête dans le cube matriciel.

L'algorithme d'analyse spatiale matricielle peut ensuite être appliqué sur cette couverture matricielle de référence. Cette étape de transposition s'ajoute au traitement de l'analyse spatiale matricielle, ce qui n'est pas favorable à son exécution à la volée, car il s'agit d'une étape additionnelle.

L'alternative, consistant à intervenir en amont de la discontinuité du domaine spatial, a pour but d'associer nativement une ou plusieurs couvertures matricielles aux composantes du cube matriciel (ex. : membre, niveau, hiérarchie, dimension, mesure). Cette approche vise à spécialiser le cube matriciel afin de tendre vers l'analyse spatiale matricielle à la volée. Ainsi, toutes les requêtes sur le cube matriciel spécialisé retourneraient le résultat sous la forme d'une couverture matricielle et non plus comme une collection de cellules matricielles indépendantes. Cette approche constitue une alternative à ce qui a été proposé jusqu'à maintenant pour l'intégration du matriciel dans les cubes. Elle sera développée plus en détail dans la section 3.3 de ce chapitre.

3.2.2 Disparité des caractéristiques des couvertures matricielles pour analyse spatiale

Les algorithmes d'analyse spatiale matricielle sont conçus pour s'exécuter soit sur une seule couverture matricielle (ex. : Filtre Laplacien) ou soit sur plusieurs à la fois (ex. : Map Algebra). Les analyses spatiales matricielles s'effectuant sur une seule couverture matricielle, *a priori*, ne posent pas de problème particulier de performance pour l'exécution à la volée. Toutefois, des complications se posent lorsque plusieurs couvertures matricielles sont utilisées comme intrant pour le traitement d'une analyse spatiale matricielle. D'entrée de jeu, le nombre de couvertures matricielles impliquées dans le traitement de l'analyse spatiale matricielle aura un impact sur le temps d'exécution. Cependant, même si le nombre de couvertures demeure restreint, la disparité des caractéristiques entre ces couvertures matricielles peut entrainer des délais de traitement.

Il n'est pas rare, en analyse spatiale matricielle, de traiter des couvertures possédant une résolution spatiale ou un domaine de valeurs différent. En de telles situations, il est nécessaire de *fusionner* les couvertures matricielles préalablement au traitement de l'analyse spatiale matricielle. La fusion [Jensen, 2005] permet d'établir un cadre de référence commun pour les couvertures impliquées dans l'analyse. Les valeurs thématiques sont insérées dans la couverture de référence afin de rendre possible l'analyse spatiale matricielle. La figure suivante illustre un exemple de fusion de couvertures matricielles où une seule couverture est altérée pour établir le référentiel commun.

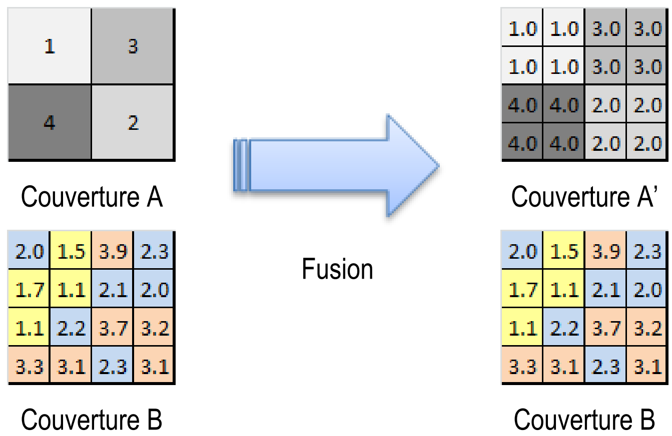


Figure 3.3 – Fusion de couvertures matricielles en prévision d'une analyse spatiale matricielle

La fusion des couvertures, lorsqu'elle est nécessaire, n'est pas favorable à l'exécution à la volée, car elle ajoute une étape supplémentaire au traitement. Le scénario qui consiste à intervenir en amont de la problématique de disparité vise à limiter, lors du design du cube, les disparités potentielles entre les couvertures formant les cubes matriciels. De cette façon, les analyses spatiales matricielles nécessitant une fusion de couvertures sont réduites au minimum et dans certains cas non requises. La section suivante propose des scénarios d'intégration des couvertures dans les hiérarchies des dimensions spatiales matricielles afin d'identifier les combinaisons les plus favorables à l'exécution à la volée d'analyses spatiales matricielles.

3.3 Intégration des couvertures matricielles dans les cubes matriciels

Le précédent chapitre présente les modèles de cubes spatiaux matriciels. Les deux formalismes de cubes proposés sont : les cubes d'objets et les cubes matriciels. D'entrée de jeu, ces deux modèles de cubes peuvent implémenter l'aspect matriciel. Ils se distinguent par le type d'objets spatiaux associé aux faits ainsi que par l'objectif d'analyse de couvertures continues ou d'objets discrets [McHugh, 2008]. Par définition, le cube matriciel est le seul type de cube matérialisant obligatoirement une couverture continue de type matriciel. Il intègre des cellules matricielles dans ses faits détaillés et possède comme objectif l'analyse par couverture. Toutefois, dans la première partie de ce chapitre, nous ne désirons pas nous limiter à cette définition du cube matriciel pour mettre en œuvre l'analyse spatiale matricielle à la volée. Il sera donc question de cubes intégrant les couvertures matricielles au sens large. Ces **cubes matriciels spécialisés** ont comme objectif de produire des analyses par couverture, mais n'intègrent pas obligatoirement des cellules matricielles dans les faits détaillés.

Les dimensions spatiales décrites dans ce chapitre réfèrent uniquement aux dimensions spatiales géométriques matricielles (voir figure ci-dessous). Les autres types de dimension spatiale matricielle intégrant des données vectorielles (dimension hybride) ou des données nominales (dimension mixte) ou les deux (dimension mixte hybride) ne seront pas abordés. Ce choix s'explique par la nécessité de traiter seulement des dimensions spatiales permettant de matérialiser des couvertures matricielles en tout temps et dans toutes les hiérarchies. La figure présente les différents types de dimensions spatiales [McHugh, 2008] et met en relief la portée des travaux de recherches effectués dans le cadre de ce mémoire.

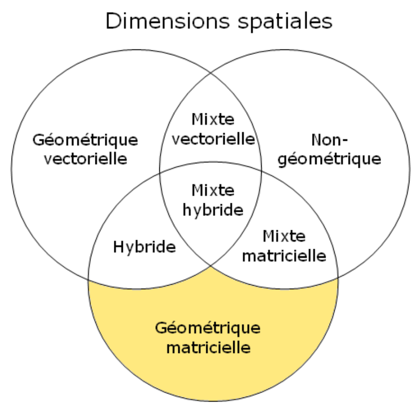


Figure 3.4 — Dimension spatiale géométrique matricielle

Également, la définition de l'intégration du matriciel dans les dimensions géométriques matricielles (*cf.* section 2.2.1) prévoit que les cellules matricielles sont associées aux membres de la dimension. L'agrégation des cellules matricielles est de nature géométrique et fait en sorte que la résolution spatiale évolue au rythme des niveaux de la hiérarchie. La figure suivante [McHugh, 2008] présente un exemple de dimension géométrique matricielle.

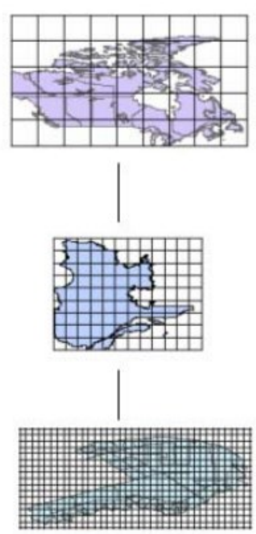


Figure 3.5 — Dimension géométrique matricielle

Dans le paradigme vectoriel, l'agrégation des niveaux d'une hiérarchie vectorielle peut être de nature **géométrique**, **thématique**, **temporelle** ou être constituée d'une combinaison de ces aspects. À titre d'exemple, l'agrégation des membres d'une hiérarchie spatiale vectorielle découle très souvent de la combinaison des aspects spatiaux et thématiques (ex. : pays >> province >> municipalité). Or, la définition découlant des travaux précédents portant sur l'intégration du matriciel dans les dimensions spatiales matricielles repose uniquement sur l'agrégation géométrique entre les niveaux sans égard à l'aspect sémantique (thématique ou temporel). La présente section vise à explorer différents scénarios d'intégration des couvertures matricielles dans les composantes d'une dimension spatiale (dimension, hiérarchie, niveau et membre). Il est sous-entendu que l'aspect géométrique ne constitue pas la seule façon de hiérarchiser ou agréger les membres matriciels d'un niveau à l'autre de la dimension. La section suivante (*cf.* section 3.4) décrit plus en détail les stratégies de hiérarchisation des couvertures matricielles. Pour le moment, le simple fait d'envisager l'agrégation des membres matriciels selon d'autres aspects que la géométrie est suffisant pour explorer les scénarios d'intégration des couvertures matricielles dans les cubes matriciels.

L'élaboration de scénarios d'intégration des couvertures matricielles dans les cubes pour faire de l'analyse spatiale matricielle exige d'aborder simultanément deux paradigmes : le fait détaillé (*niveau de granularité le plus fin du cube matriciel*) du cube et la couverture matricielle. De la perspective du fait détaillé, il faut tout d'abord déterminer si ce fait est associé à **une cellule de la couverture matricielle** ou à une **couverture matricielle entière**. La figure suivante illustre cette dualité dans l'intégration des couvertures matricielles dans les cubes matriciels où le fait détaillé est associé à une cellule de la couverture matricielle ou à une couverture matricielle.

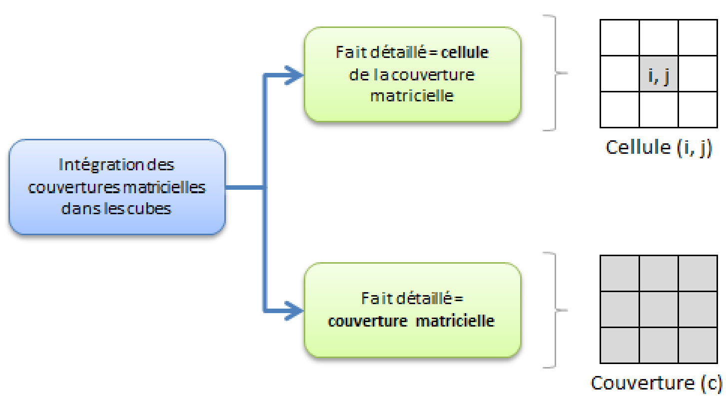


Figure 3.6 — Intégration conceptuelle de couvertures matricielles dans les cubes matriciels

Les travaux précédents [McHugh, 2008] ont porté uniquement sur les faits détaillés associés à des cellules matricielles. À ce stade-ci, il est envisageable qu'une cellule matricielle ou une couverture matricielle soit partagée avec plusieurs faits. Cette notion de partage est pertinente notamment pour réduire la taille des cubes et éviter la redondance des géométries, ce qui est en accord avec les objectifs d'optimisation décrits à la section suivante [*cf.* Section 4 - Optimisation des analyses spatiales matricielles). Ainsi, de la perspective de la couverture matricielle, il faut déterminer le nombre de faits qui sont associés à une couverture matricielle ou inversement, le nombre de cellules matricielles associées à un fait. Considérant qu'un cube peut matérialiser plusieurs couvertures matricielles, se posent alors plusieurs cas possibles. Ces scénarios sont résumés schématiquement à la figure 22. Le premier consiste à associer une seule couverture matricielle à tous les faits du cube (faible nombre de géométries). Ensuite, une même couverture peut être partagée entre plusieurs faits sans toutefois être associée à tous les faits. Finalement, chaque fait détaillé ou agrégé peut posséder sa propre couverture matricielle (très grand nombre de géométries). La figure suivante illustre les perspectives du fait détaillé et de la couverture matricielle ainsi que les différentes combinaisons « fait/couverture » abordées dans le cadre de cette recherche. Tous ces scénarios sont présentés individuellement dans les sections 3.3.1 et 3.3.2 du présent chapitre.

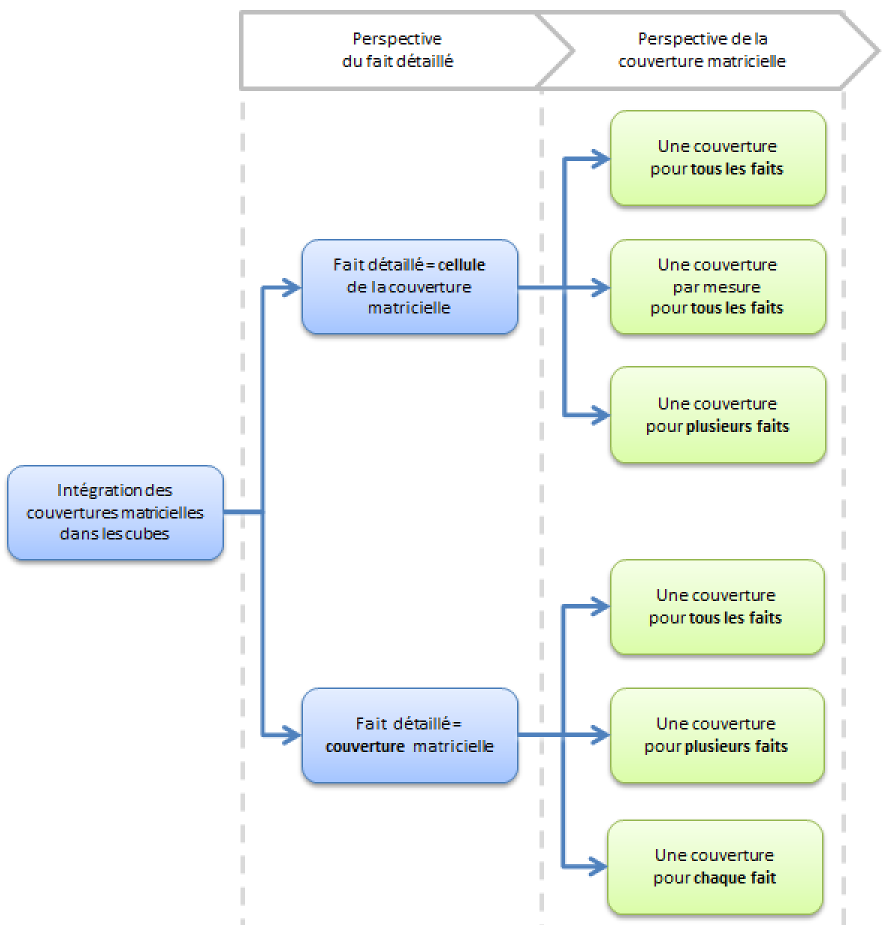


Figure 3.7 – Intégration conceptuelle des couvertures matricielles (perspectives du fait détaillé et de la couverture matricielle)

Enfin, il existe des règles strictes pour définir ce qu'est une couverture matricielle au sens prévu dans ces travaux (*cf.* chapitre 2.1.2). Une couverture matricielle est une grille spatiale matricielle possédant les caractéristiques suivantes :

1. toutes les cellules possèdent une géométrie;
2. toutes les géométries sont carrées, de même taille et de même orientation;
3. une coordonnée (ligne, colonne) retourne un objet spatial unique;
4. le domaine de valeurs des cellules matricielles, formant la couverture, est homogène.

De ces caractéristiques découlent certains constats eu égard aux caractéristiques descriptives des couvertures matricielles. Notamment, la **résolution spatiale** est homogène sur l'ensemble de la couverture et des couches qui la compose. La **variable thématique** est une propriété associée à la couverture en présence d'une seule couche matricielle. Si la couverture matricielle possède plusieurs couches, chacune des couches exprime sa propre variable thématique et son propre domaine de valeurs. La **valeur thématique** décrit la valeur numérique associée à chaque cellule de la couche matricielle. L'ensemble des valeurs thématiques observées sur une couche matricielle correspond au **domaine de valeurs** de la couche ou de la couverture. Enfin, la caractérisation de l'intégration des couvertures matricielles dans les cubes repose sur l'observation de ces caractéristiques. Par exemple, un cube matriciel qui intègre une seule couverture matricielle possède une résolution spatiale constante pour l'ensemble du cube. Or, un cube qui possède plusieurs couvertures aura une résolution spatiale variable en fonction de ces couvertures.

C'est la combinaison des perspectives « fait/couverture » à l'évolution des caractéristiques des couvertures qui permet de décrire des scénarios possibles d'intégration des couvertures dans les cubes. Ces scénarios d'intégration sont élaborés de façon à intégrer une ou plusieurs couvertures aux composantes essentielles du cube : un membre, un niveau, une hiérarchie, une dimension, un fait, une mesure ou le cube lui-même. Ces scénarios sont théoriques et leur valeur par rapport à l'analyse spatiale matricielle à la volée est discutée à la fin de cette section. Ces scénarios sont au nombre de sept (7) et se déclinent selon les modalités présentées dans le schéma ci-dessous.

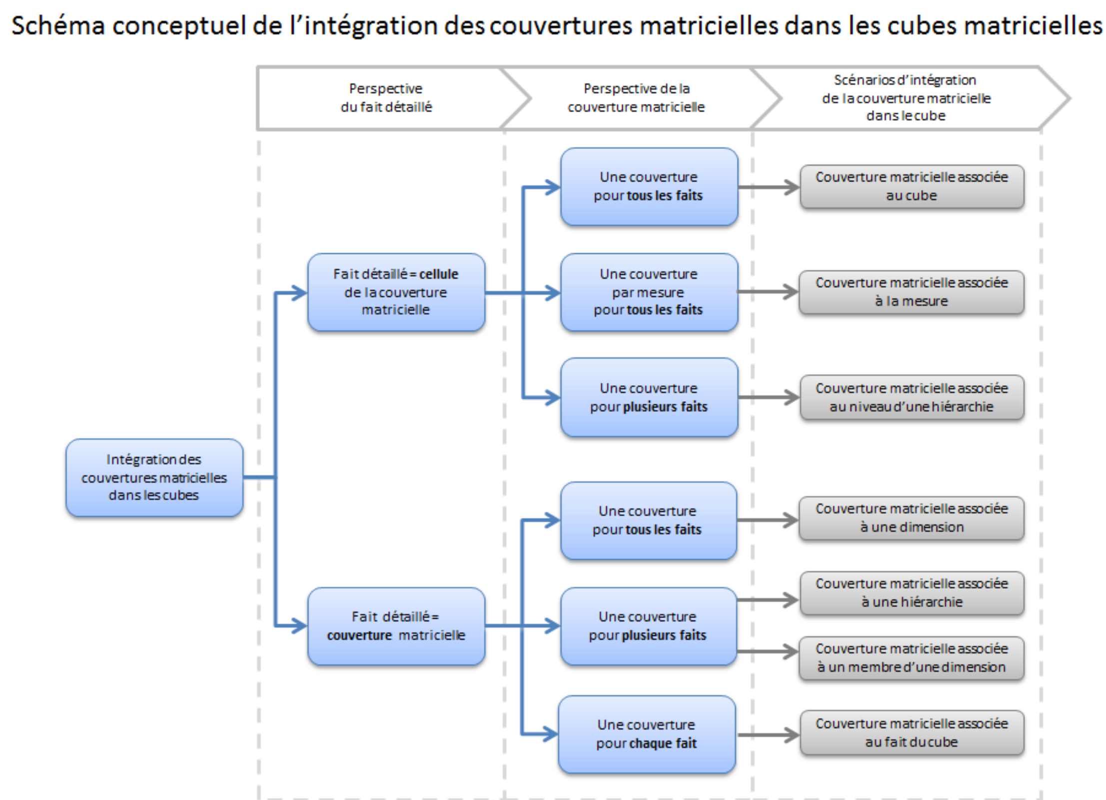


Figure 3.8 — Scénarios d'intégration des couvertures matricielles dans les cubes matriciels

La section qui suit vise à explorer ces scénarios d'intégration des couvertures matricielles dans les cubes spatiaux. Ces scénarios seront décrits conceptuellement et caractérisés selon leur propension à favoriser ou non l'analyse spatiale matricielle à la volée. À cet effet, les enjeux décrits précédemment c.-à-d. la discontinuité du domaine spatial et la disparité des caractéristiques des couvertures matricielles serviront de critères discriminants. En premier lieu, il sera question des scénarios associant un fait à une cellule matricielle. Ensuite, les autres scénarios associant un fait à une couverture matricielle seront décrits. Finalement, ces scénarios seront analysés afin de décrire les conditions sous lesquelles ils sont favorables ou non à l'exécution à la volée des analyses spatiales matricielles. Il est sous-entendu que ces scénarios ne sont pas tous favorables à l'analyse spatiale matricielle à la volée. Un tableau récapitulatif et comparatif des scénarios est présenté à la fin de cette section (*cf.* Tableau 3.1 — Résumé des caractéristiques des scénarios d'intégration des couvertures matricielles en fonction des perspectives du cube ou de la couverture). Enfin, seuls les scénarios les plus favorables à l'analyse spatiale matricielle à la volée seront décrits plus en détail.

3.3.1 Fait détaillé associé à une cellule de couverture matricielle

L'intégration des couvertures matricielles dans les cubes matriciels, où le fait détaillé est associé à une cellule matricielle, signifie que nous sommes en présence de cubes matriciels authentiques (*cf.* chapitre 2.2.3). Les scénarios d'intégration sont au nombre de trois (3) et se distinguent par l'association de la couverture matricielle à :

a) tous les faits;
b) tous les faits d'une même mesure;
c) plusieurs faits.

Les scénarios d'intégration des couvertures matricielles dans les cubes matriciels sont présentés dans la figure ci-dessous qui constitue un extrait de la figure précédente.

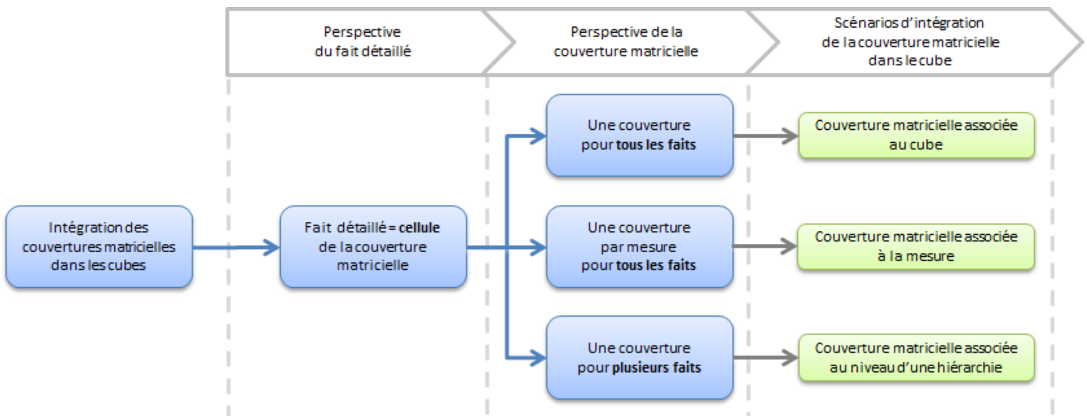


Figure 3.9 – Extrait des scénarios d'intégration des couvertures matricielles dans les cubes matriciels

Le premier scénario signifie qu'il n'y a qu'une seule couverture matricielle pour l'ensemble du cube. Pour les deux autres scénarios, il y a potentiellement plusieurs couvertures matricielles s'il y a plusieurs mesures dans le cube ou plusieurs niveaux dans la dimension matricielle.

Couverture matricielle associée au cube matriciel

L'association d'une couverture matricielle à un cube matriciel signifie que celui-ci possède une seule couverture matricielle. La résolution spatiale et l'étendue spatiale sont des propriétés du cube, car elles sont constantes pour l'ensemble du cube matriciel. Incidemment, l'agrégation des faits détaillés, en fonction des axes d'analyse, est de nature temporelle ou thématique, car la résolution spatiale est constante. Le fait issu du croisement des axes d'analyse est associé à une seule cellule de la couverture matricielle. Par contre, une cellule de la couverture peut être associée à plusieurs faits.

Les valeurs numériques des mesures correspondent aux valeurs thématiques des cellules de la couverture. La mise en œuvre de plusieurs mesures dans le cube est l'équivalent de matérialiser plusieurs couches dans une couverture matricielle. À titre d'exemple, la figure suivante illustre des couches d'une couverture matricielle exprimant les variations numériques des mesures par intensités de rouge (R), de vert (V) ou de bleu (B) pour les mêmes coordonnées matricielles.

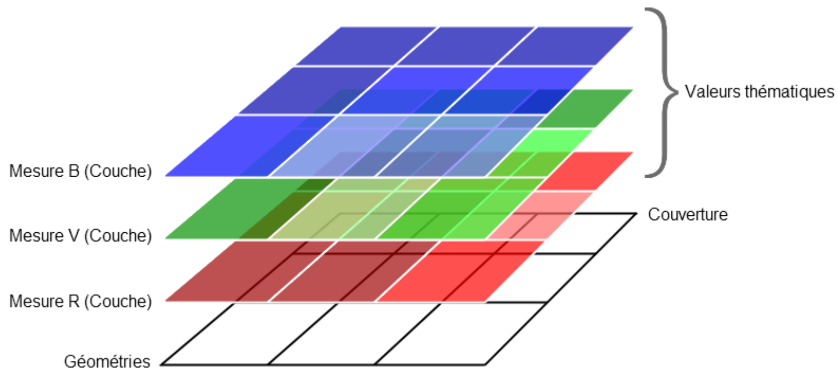


Figure 3.10 — Couches d'une couverture matricielle associée aux mesures d'un cube matriciel

Le fait est associé à la couverture matricielle à l'aide d'un identifiant de cellule (*i, j*). Toutes les mesures d'un fait font référence à la même cellule de la couverture et partagent le même identifiant. La figure suivante illustre conceptuellement le partage entre plusieurs mesures d'un fait d'une référence vers une seule cellule de la couverture matricielle.

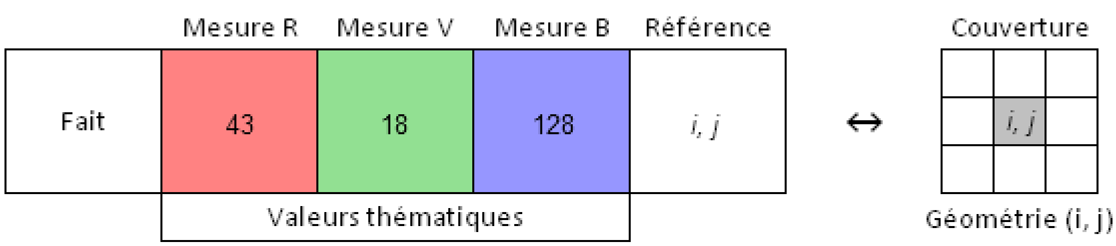


Figure 3.11 — Partage d'une même cellule matricielle entre plusieurs mesures

Dans le monde OLAP, les mesures possèdent souvent des domaines de valeur thématique différents (ex. : nombre de personne et température moyenne). En associant les mesures à des couches de la couverture matricielle, ceci permet de faire varier le domaine de valeurs thématiques d'une couche à l'autre. Toutefois, cette variation du domaine de valeurs entre les mesures n'est pas favorable à l'exécution à la volée des analyses spatiales matricielles. En effet, elle implique de faire la fusion des couches avant de procéder à une analyse spatiale matricielle. En contrepartie, l'homogénéisation de la variable exprimant les valeurs numériques des mesures permet d'éviter le traitement de fusion. À cet effet, d'autres recommandations favorisant l'analyse spatiale matricielle à la volée sont regroupées à la fin de cette section. Enfin, les cubes matriciels formés par ce type d'intégration constituent une des méthodes les plus simples pour intégrer des couvertures matricielles dans les cubes OLAP spatiaux. La prochaine méthode explorée augmente le niveau de flexibilité et de complexité du cube matriciel.

Couverture matricielle associée à une mesure du cube matriciel

L'association d'une couverture matricielle à une mesure du cube matriciel plutôt qu'au cube lui-même ouvre la porte à une intégration plus flexible. Elle est plus flexible, car la résolution spatiale, l'étendue spatiale et le domaine de valeurs thématique constituent des caractéristiques de la mesure plutôt que du cube en entier. Ainsi, ces caractéristiques sont constantes pour une mesure, mais variables à l'intérieur du cube.

Le fait issu du croisement des axes d'analyse est associé, à l'aide d'un identifiant de couverture et de cellule (*c, i, j*), à la couverture matricielle pertinente. Les valeurs numériques des mesures correspondent aux valeurs thématiques des cellules de la couverture. La mise en œuvre de plusieurs mesures dans le cube matérialise plusieurs couvertures matricielles. Enfin, une mesure contient une ou plusieurs couches en fonction des membres des dimensions du cube matriciel. Cependant, les couches d'une même mesure possèdent toutes le même domaine de valeurs thématiques, car il s'agit d'une propriété de la mesure du cube.

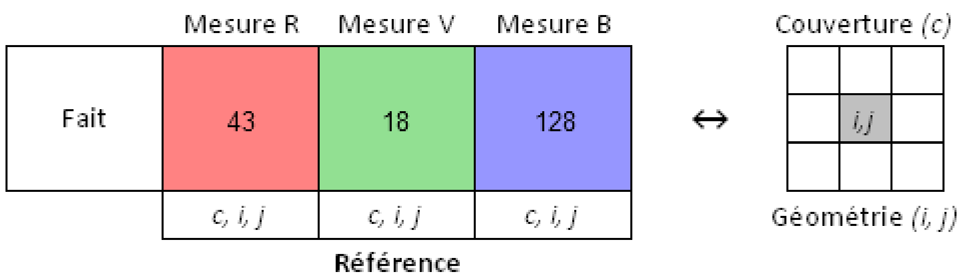


Figure 3.12 — Référence à une cellule d'une couverture matricielle pour chaque mesure

La couverture matricielle d'une mesure peut potentiellement présenter des caractéristiques très différentes de la couverture d'une autre mesure du cube. La figure suivante illustre une dramatisation de la variation des caractéristiques des couvertures matricielles entre les mesures d'un même cube.

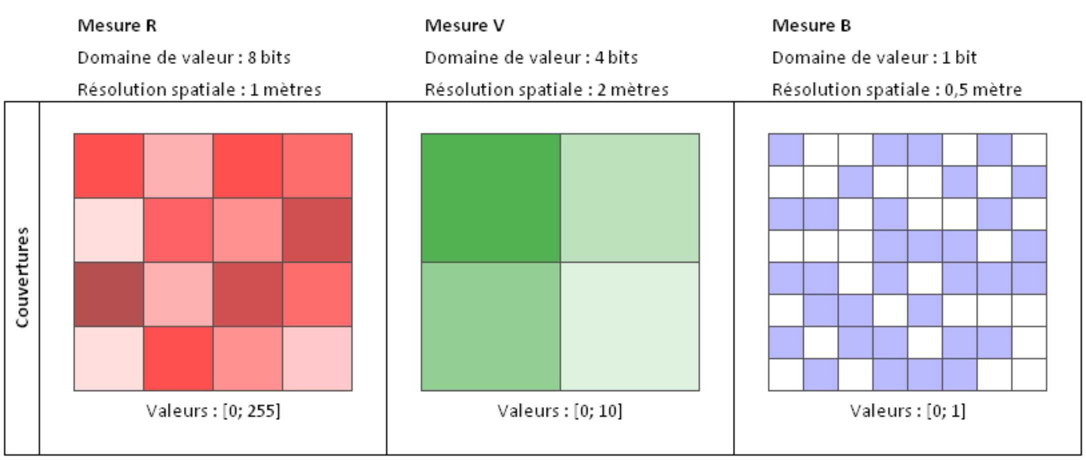


Figure 3.13 — Variation des caractéristiques de couverture entre les mesures du cube

Les analyses spatiales matricielles impliquant la comparaison de mesures différentes peuvent alors devenir complexes et nécessiter une fusion des couvertures impliquées dans l'analyse. Toutefois, en effectuant une analyse spatiale matricielle sur les différentes couches d'une même mesure il n'est pas nécessaire de recourir à une fusion de couverture. Par exemple, l'addition des précipitations liquides mensuelles localement pour l'année 2010 constitue une analyse spatiale matricielle d'une seule couverture et d'une seule mesure, mais provenant de faits différents. Dans ce type d'intégration, l'analyse spatiale matricielle « intra mesure » est favorable à l'exécution à la volée, alors que l'analyse « inter mesure » est moins favorable.

Couverture matricielle associée au niveau d'une hiérarchie

L'intégration de la couverture matricielle dans le niveau consiste à associer chaque cellule formant la couverture matricielle à un membre du niveau. Dans les dimensions spatiales matricielles, chaque niveau est composé d'une couverture matricielle différente. Les couvertures formant la hiérarchie permettent de représenter différentes échelles du territoire. L'association d'une couverture matricielle à un niveau d'une hiérarchie d'une dimension ouvre la porte à une intégration plus flexible, mais plus complexe. Cette méthode est plus flexible, car la résolution spatiale et l'étendue spatiale constituent des caractéristiques propres à chaque niveau. Elle est plus complexe, car la présence de plusieurs niveaux entraine le besoin de hiérarchiser les couvertures matricielles entre elles afin d'établir la subordination entre les membres des niveaux. La hiérarchisation des couvertures matricielles est un processus relativement complexe qui fera l'objet d'une étude plus approfondie à la section suivante (*cf.* section 3.4). La figure suivante illustre conceptuellement ce type d'intégration de couverture matricielle dans les dimensions spatiales matricielles.

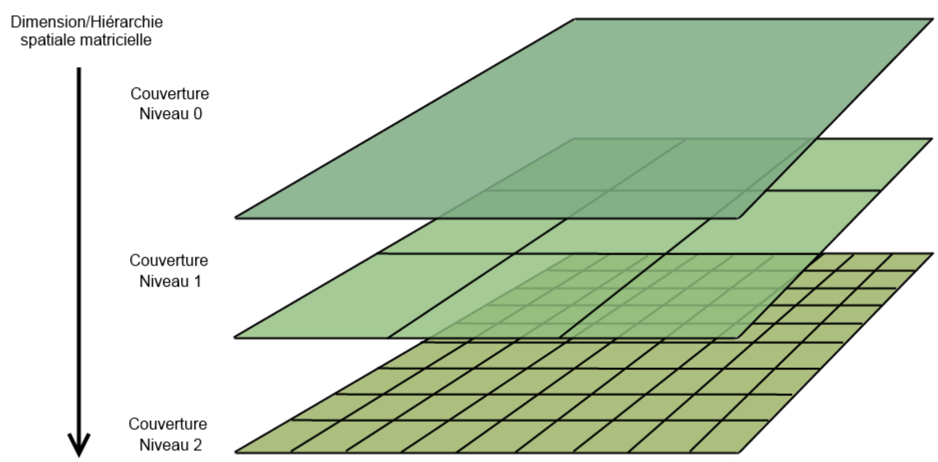


Figure 3.14 — Exemple de couvertures associées aux niveaux d'une hiérarchie

Le fait issu du croisement des axes d'analyse est associé, à l'aide d'un identifiant de membre à une cellule de la couverture du niveau auquel il appartient. Les valeurs numériques des mesures correspondent aux valeurs thématiques des cellules de la couverture. La mise en œuvre de plusieurs mesures dans le cube matérialise plusieurs couches de couvertures matricielles.

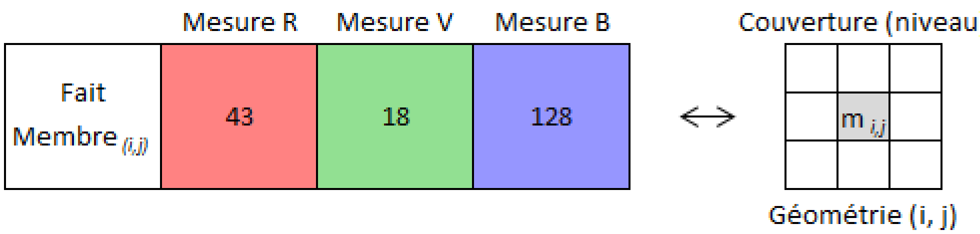


Figure 3.15 — Référence à une cellule d'une couverture matricielle d'un niveau

Les analyses spatiales matricielles impliquant un seul niveau et une seule mesure sont très favorables à l'exécution à la volée. En effet, dans ce contexte la résolution spatiale, l'étendue spatiale et le domaine de valeurs sont constants. La fusion des couvertures n'est pas nécessaire dans ce contexte, car la seule variation est imputable à la variation thématique. Toutefois, la situation se complique lorsque plusieurs niveaux ou plusieurs mesures sont nécessaires à l'analyse spatiale matricielle.

3.3.2 Fait détaillé associé à une couverture matricielle

La première partie de cette section présentait des scénarios théoriques où le fait détaillé (*niveau de granularité le plus fin du cube matriciel*) du cube est associé à une cellule matricielle provenant d'une couverture matricielle. Pour les scénarios qui suivent, le fait détaillé ne réfère plus (*directement ou par l'intermédiaire d'une dimension matricielle)* à une seule cellule, mais plutôt à un ensemble de cellules formant la couverture matricielle. Ainsi, la mesure du cube ne contient pas de valeur thématique, mais seulement une référence (*pointeur*) vers un tableau de valeurs thématiques et une couverture matricielle (*objet spatial*). La figure suivante illustre conceptuellement la référence entre les mesures d'un fait et les valeurs thématiques et la couverture matricielle.

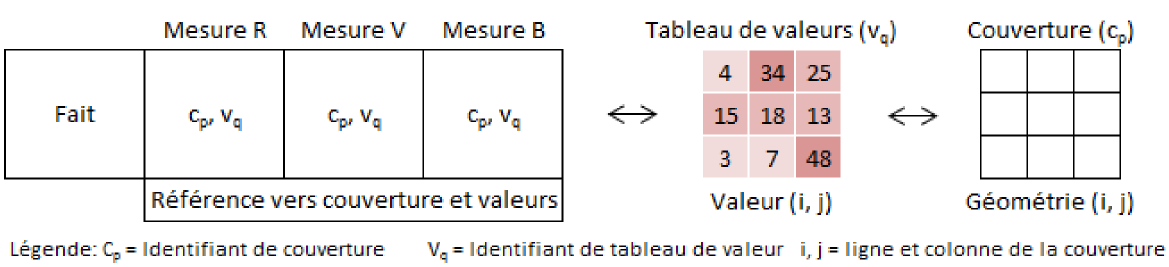


Figure 3.16 – Exemple d'intégration d'une couverture matricielle dans un fait détaillé d'un cube spatial

L'intégration des couvertures matricielles dans les cubes spatiaux où le fait détaillé est associé à une couverture matricielle signifie que nous sommes en présence de cubes d'objets (*cf.* chapitre 2.2.3). Toutefois, un cas particulier se pose lorsque tous les faits et toutes les mesures partagent la même couverture matricielle. En effet, il ne s'agit plus d'objets spatiaux indépendants du point de vue géospatial, mais bien d'une couverture continue pour tous les faits du cube. Dans une telle situation, il n'est plus nécessaire de procéder à la fusion des couvertures matricielles, car les caractéristiques de la couverture sont homogènes pour tous les faits. Il s'agit donc d'une situation favorable à l'analyse spatiale matricielle à la volée. C'est ce qui explique l'intérêt que peut présenter l'intégration des couvertures matricielles dans les faits détaillées.

Les scénarios d'intégration sont au nombre de quatre (4) et se distinguent par le fait que la couverture matricielle peut être la même pour :

a) tous les faits;
b) pour plusieurs faits;
c) pour chaque fait.

Il faut rappeler ici qu'il s'agit de scénarios théoriques d'intégration et qu'ils ne sont pas tous favorables à l'analyse spatiale matricielle. Les scénarios d'intégration des couvertures matricielles dans les cubes d'objets sont présentés dans la figure ci-dessous.

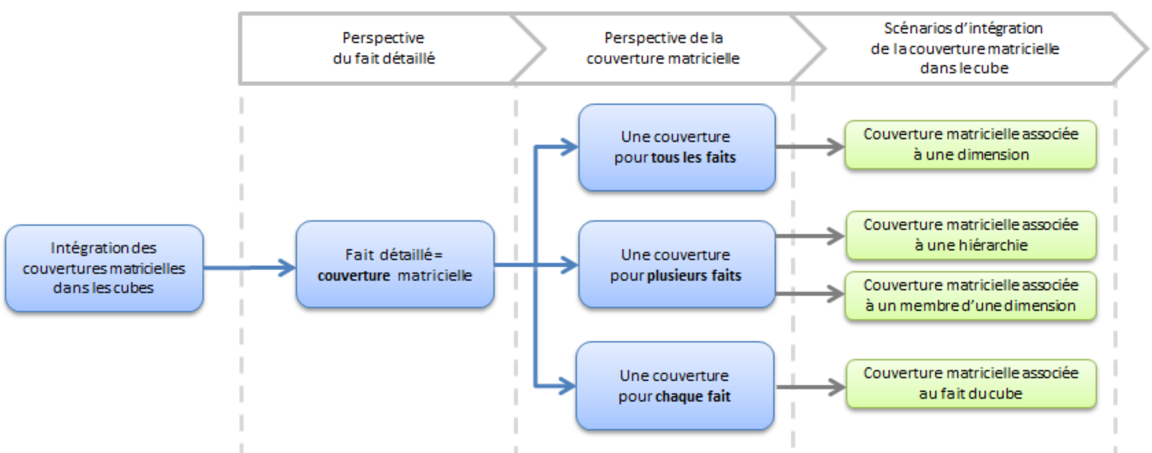


Figure 3.17 – Extrait des scénarios d'intégration des couvertures matricielles dans les cubes matriciels

Le premier scénario signifie qu'il n'y a qu'une seule couverture matricielle pour l'ensemble de la dimension et incidemment pour le cube en présence d'une seule dimension spatiale matricielle. Dans les deux scénarios suivants, il y a potentiellement plusieurs couvertures matricielles pour un même cube. Finalement, le dernier scénario prévoit qu'il y a autant de couvertures que de faits dans le cube.

Couverture matricielle associée à une dimension

L'association d'une couverture matricielle à une dimension signifie que nous sommes en présence d'une dimension spatiale matricielle. Tous les membres de la dimension font référence à une seule couverture matricielle. Les membres de la dimension, en combinaison avec les mesures du cube, forment les différentes couches de la couverture matricielle. La résolution spatiale et l'étendue spatiale sont des propriétés de la dimension spatiale. Le fait du cube réfère à un tableau de valeurs (*couche de la couverture*) et tous les faits partagent la même couverture matricielle. La figure suivante illustre la référence vers un tableau de valeurs et une couverture unique pour chaque fait et chaque mesure.

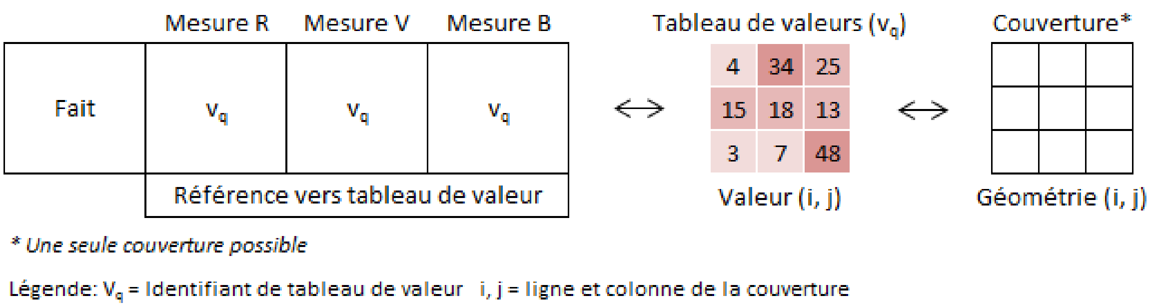


Figure 3.18 — Référence à un tableau de valeur et à une couverture unique

L'intégration de couvertures matricielles dans les dimensions entraine la création d'une dimension spatiale matricielle géométrique (*cf.* chapitre 2.2.3). En présence de plusieurs niveaux, la hiérarchisation des membres n'est pas effectuée sur une base géométrique, mais sur la base de l'information thématique (sémantique) et/ou temporelle. La hiérarchisation des couvertures matricielles sera décrite plus en détail à la section suivante (*cf.* section 3.4).

Afin d'illustrer le concept de dimension spatiale thématique, le cube spatial de foresterie [Proulx et *al.*, 2006] a été modifié de façon à intégrer, à même la dimension « Essence », la couverture matricielle. La figure suivante présente le modèle multidimensionnel de cube d'objet vectoriel original.

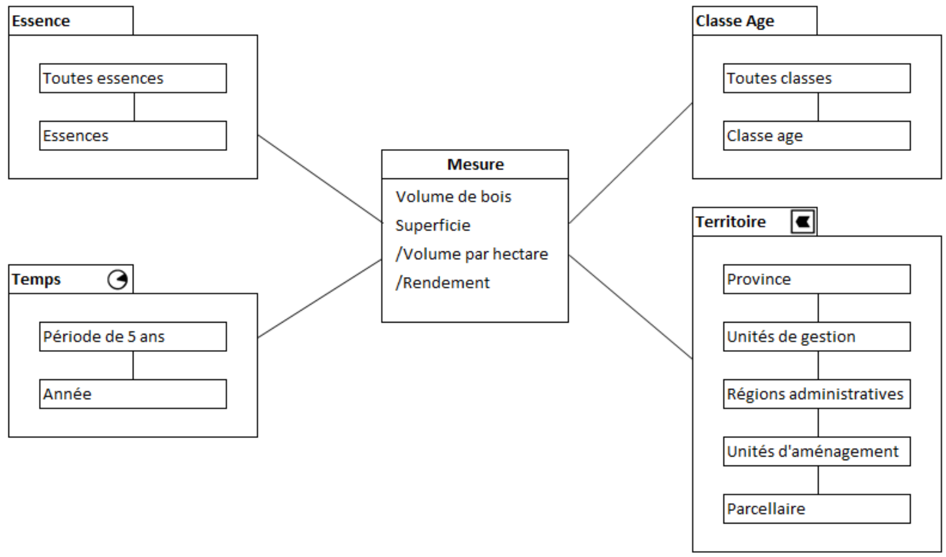


Figure 3.19 — Modèle multidimensionnel du cube foresterie[9]

La dimension spatiale territoire est de type vectoriel. Afin de transformer ce modèle en cube matriciel, la dimension spatiale territoire a été retirée. La couverture matricielle a été intégrée à la dimension « Essence » afin d'en faire une dimension spatiale et thématique. La figure suivante présente le modèle multidimensionnel adapté du cube matriciel.

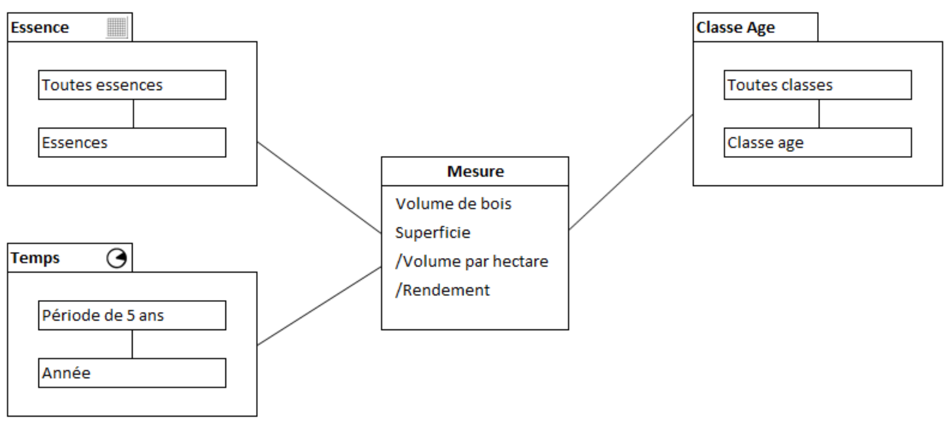


Figure 3.20 — Modèle multidimensionnel d'un cube matriciel en foresterie

La figure ci-dessous présente des exemples d'exploitation du cube matriciel en foresterie. Peu importe le membre sélectionné dans la dimension « Essence », le résultat est présenté sous forme de couverture matricielle, mais une ou plusieurs couches différentes.

---

[9] [Proulx et *al.*, 2006]

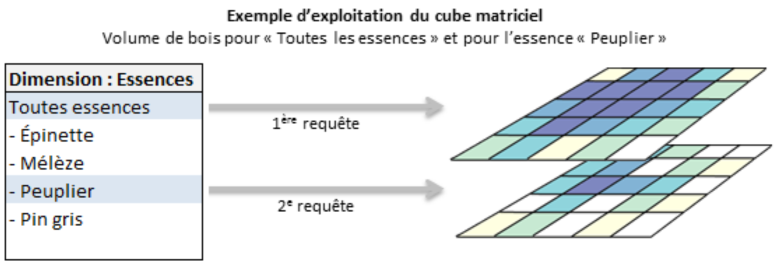


Figure 3.21 — Exploitation d'une dimension spatiale thématique matricielle

L'agrégation des valeurs par cellule est de nature thématique et spatiale, mais pas géométrique. Tel que la figure suivante l'illustre, le résultat de la sélection de tous les « membres enfants » d'un « membre parent » équivaut à la sélection de ce « membre parent ».

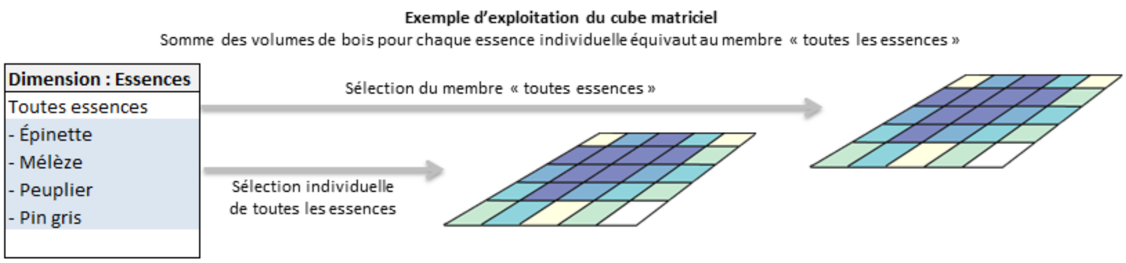


Figure 3.22 – Comparaison de sélection de tous les membres enfants et du membre parent

L'analyse spatiale matricielle à la volée est très favorable dans ce contexte, car tous les faits partagent la même couverture et incidemment les mêmes caractéristiques de couvertures. La figure suivante illustre un exemple d'analyse spatiale matricielle qui peut être réalisé à la volée. L'addition locale (cellule par cellule) des volumes de bois constitue une opération triviale et par conséquent performante selon les critères établis précédemment.

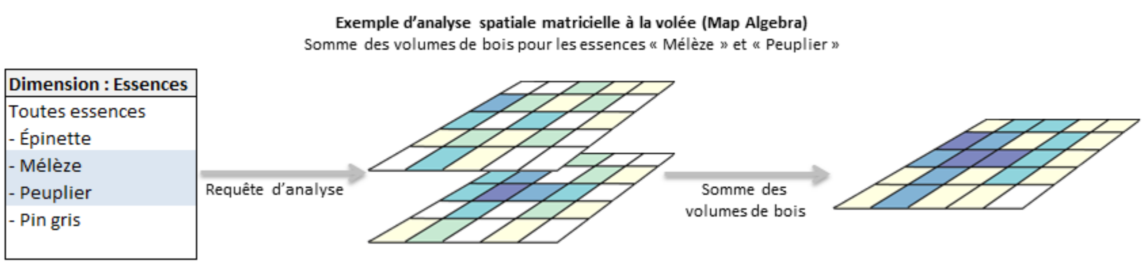


Figure 3.23 — Analyse spatiale matricielle à la volée

L'analyse spatiale matricielle à la volée est moins favorable lorsqu'elle implique plusieurs mesures différentes (variation du domaine de valeurs), mais demeure tout de même favorable. L'intégration de plusieurs couvertures matricielles, dans une même dimension, nécessite d'introduire la notion de hiérarchie dans les dimensions spatiales matricielle.

Couverture matricielle associée à une hiérarchie

L'association d'une couverture matricielle à une hiérarchie signifie que tous les membres de la hiérarchie font référence à une seule couverture matricielle. Il s'agit d'un scénario d'intégration plus flexible, car il est possible d'avoir plusieurs hiérarchies matricielles et donc plusieurs couvertures matricielles à l'intérieur d'une même dimension. Le nombre de niveaux d'une hiérarchie à l'autre peut différer et il n'y a pas toujours de correspondance entre les niveaux de ces hiérarchies. La résolution spatiale et l'étendue spatiale sont des propriétés de la hiérarchie spatiale. Le fait du cube réfère à un tableau de valeurs et plusieurs faits partagent la même couverture matricielle. Enfin, un fait est associé à une seule couverture. La figure suivante illustre la référence vers un tableau de valeur et une couverture pour chaque fait et chaque mesure.

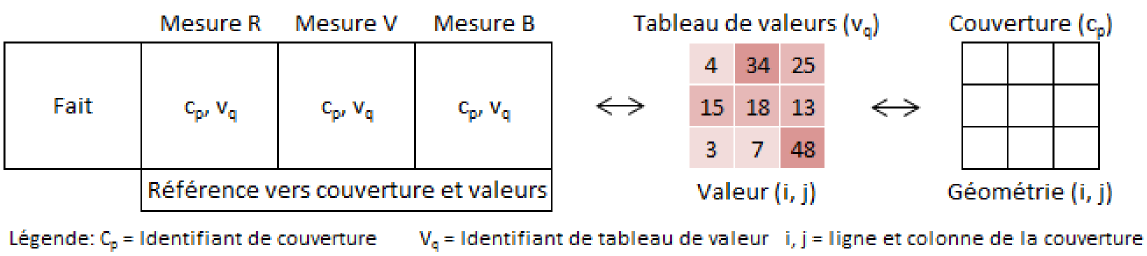


Figure 3.24 — Référence à un tableau de valeurs et à une couverture

L'avantage principal de ce scénario par rapport au précédent consiste à matérialiser plusieurs couvertures différentes à l'intérieur d'une même dimension. La figure suivante présente un exemple de dimension spatiale matricielle thématique intégrant des couvertures différentes dans deux hiérarchies. La première hiérarchie possède une couverture avec des cellules de cinq (5) mètres de résolution spatiale et une couverture avec 50 mètres de résolution spatiale. La résolution la plus fine permet de produire des analyses locales, par exemple à l'échelle d'une plantation, alors que la résolution plus grossière est destinée aux analyses régionales ou écosystémiques.

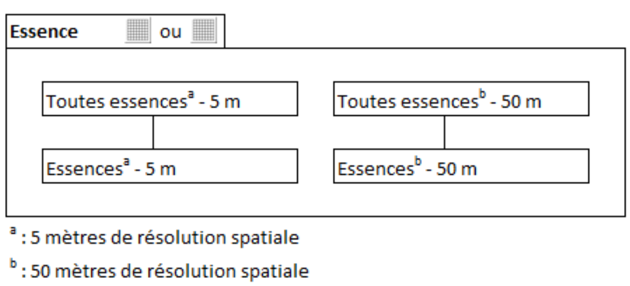


Figure 3.25 — Dimension spatiale matricielle thématique matérialisant deux couvertures matricielles dans deux hiérarchies différentes

Ce scénario est très favorable à l'analyse spatiale matricielle à la volée si toutes les couvertures utilisées dans l'analyse proviennent de la même hiérarchie. Dans ce type d'intégration, l'analyse spatiale matricielle « intra hiérarchie » est favorable à l'exécution à la volée. Pour sa part, l'analyse « inter hiérarchie » n'est pas favorable à l'exécution à la volée notamment lorsque la résolution spatiale est différente d'une hiérarchie à l'autre.

Couverture matricielle associée à un membre d'une dimension

La mise en œuvre de ce scénario d'intégration des couvertures matricielles implique que chaque membre de la dimension spatiale matricielle est associé à une couverture matricielle. La résolution spatiale et l'étendue spatiale sont des caractéristiques propres à chaque membre. Le fait du cube réfère à un tableau de valeurs et plusieurs faits partagent la même couverture matricielle. Il s'agit d'un scénario d'intégration très flexible, mais qui n'est pas favorable à l'analyse spatiale matricielle à la volée. En effet, les couvertures peuvent posséder des caractéristiques très distinctes (disparité des caractéristiques des couvertures) et même être associées à des territoires différents (discontinuité du domaine spatial). Un exemple d'application consisterait à créer un catalogue de couvertures matricielles, structuré en fonction du capteur, de l'année, des corrections et des rehaussements apportés aux valeurs thématiques. Selon les enjeux évoqués précédemment (*cf.* 3.2 Enjeux spécifiques de l'analyse spatiale dans les cubes matriciels), l'exécution à la volée est envisageable seulement pour les faits provenant du même membre spatial, c.-à-d. de la même couverture matricielle. L'analyse spatiale matricielle « inter mesure » ou « inter membre spatial » nécessite une fusion des couvertures matricielles et n'est donc pas favorable à l'analyse spatiale matricielle à la volée.

Couverture matricielle associée au fait du cube

L'association d'une couverture matricielle à un fait consiste à obtenir une couverture matricielle pour chaque fait du cube. Le recours à une dimension spatiale matricielle n'est pas nécessaire pour procéder à cette intégration. La navigation dans les axes d'analyse équivaut à sélectionner une couverture matricielle et la sélection de mesure permet d'identifier les couches d'intérêt. D'un point de vue utilitaire, tout comme le scénario précédent, cette approche pourrait servir à consulter une banque d'image structurée sous forme d'axes d'analyse (années, capteur, corrections, rehaussement, etc.).

Dans le contexte de cette recherche, ce scénario d'intégration n'est pas envisageable pour l'analyse spatiale matricielle à la volée. En effet, la possibilité d'associer à chaque fait sa propre couverture entraine l'indépendance des faits entre eux. Cette indépendance est caractéristique des objets spatiaux discrets ce qui constitue l'apanage des cubes d'objets. L'intégration des couvertures matricielles dans les faits détaillés ne sera pas approfondie davantage dans le cadre de cette recherche, mais pourrait comporter un intérêt pour des travaux futurs sur l'analyse spatiale matricielle dans les structures OLAP spatiales.

3.3.3 Constats et recommandations pour l'intégration des couvertures matricielles dans les cubes matriciels pour l'analyse spatiale matricielle à la volée

L'objectif était de présenter les scénarios possibles d'intégrations des couvertures matricielles dans les cubes matriciels et d'en observer les spécificités dans une perspective d'analyse spatiale à la volée. *A priori*, ce ne sont pas tous les scénarios d'intégration qui permettent de mettre en œuvre des couvertures continues pour l'ensemble du cube. L'intégration des couvertures matricielles par le cube, la mesure, le niveau, la dimension et la hiérarchie constituent les cinq (5) scénarios où la continuité est assurée. Cette continuité est toutefois conditionnelle à la mise en œuvre de dimensions spatiales géométriques matricielles. La figure suivante illustre les différents scénarios matérialisant des couvertures continues dans le cube.

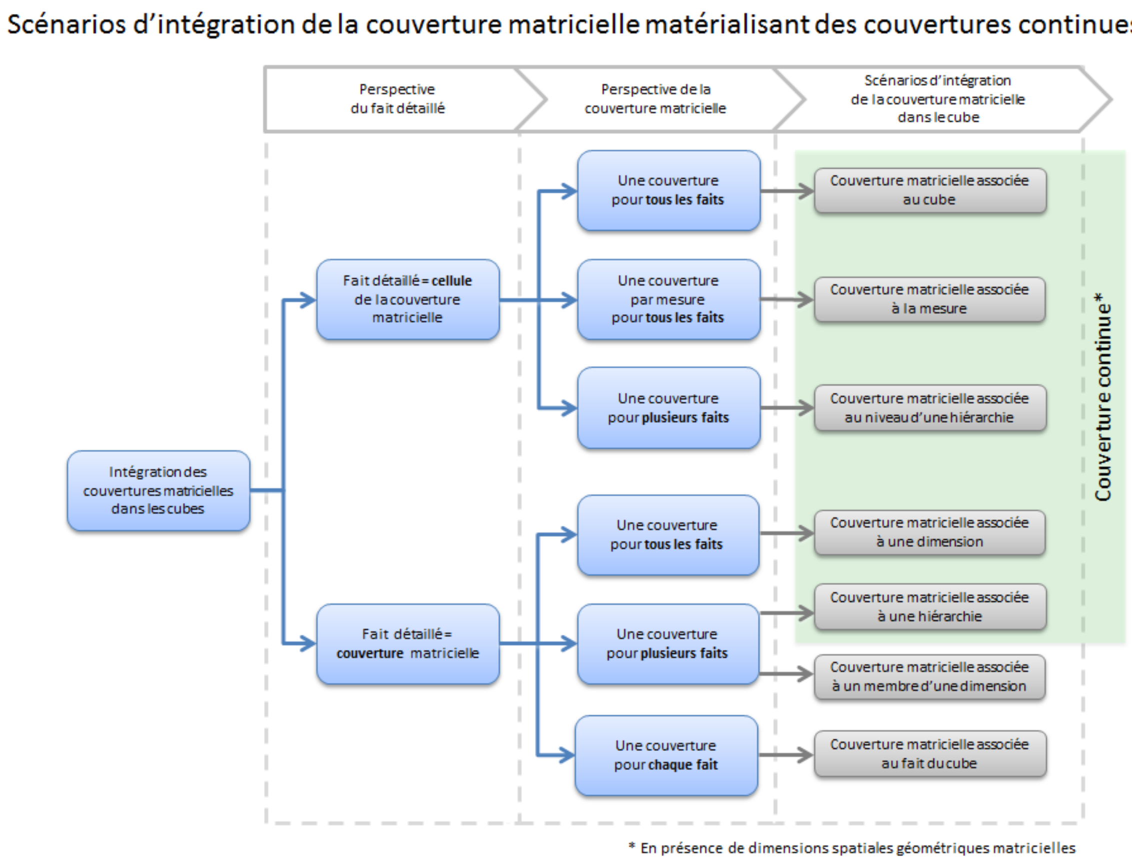


Figure 3.26 — Scénarios d'intégration de la couverture matricielle matérialisant des couvertures continues

Ce sont aussi les seuls scénarios matérialisant des cubes matriciels. Toutefois, il est nécessaire d'élargir la définition de cube matriciel afin d'inclure également les cas où le fait détaillé intègre une couverture matricielle et non seulement une cellule de la couverture.

> **Cube matriciel** : *Cube spatial intégrant une ou plusieurs couvertures matricielles. Les faits détaillés sont associés aux cellules d'une couverture ou à une couverture matricielle. Pour ce type de cube, les faits sont toujours géométriques et cette géométrie est propre au fait ou provient d'une association à un membre spatial d'une dimension. Enfin, un cube matriciel a pour objectif de répondre à un besoin d'analyse par couverture continue, ce qui le distingue fondamentalement du cube d'objet.*

Aussi, il faut souligner que les cubes matriciels destinés à l'analyse spatiale matricielle peuvent comporter des dimensions spatiales matricielles mixtes ou hybrides. Cette recherche ne visait pas à identifier toutes les combinaisons possibles pour ces types de dimension spatiale, mais bien à explorer des scénarios favorables à l'analyse spatiale matricielle à la volée à partir des situations les plus simples (ce qui est normal pour le stade actuel des connaissances et l'ampleur d'une maîtrise).

Les scénarios d'intégration présentés précédemment sont nombreux et possèdent des caractéristiques distinctes. Le Tableau 3.1 — Résumé des caractéristiques des scénarios d'intégration des couvertures matricielles en fonction des perspectives du cube ou de la couverture ci-dessous présente les principales caractéristiques des scénarios d'intégration des couvertures matricielles. Il résume également l'appréciation de la capacité d'un scénario d'intégration des couvertures matricielles à supporter l'analyse spatiale matricielle à la volée. Pour chaque scénario, les conditions d'exploitation les plus favorables à l'analyse spatiale matricielle à la volée sont présentées. Ces conditions d'exploitation sont « **très favorables** » lorsqu'il n'est pas requis de fusionner des couvertures. Les conditions sont « **favorables** » si la fusion requise vise seulement à uniformiser le domaine de valeurs thématiques exprimé par les couvertures matricielles. Enfin, les conditions d'exploitation sont « **défavorables** » si les couvertures nécessitent une fusion pour uniformiser la résolution spatiale. À titre d'exemple, l'intégration des couvertures par le cube est très favorable à l'analyse spatiale matricielle lorsqu'une seule mesure du cube est impliquée dans l'analyse. Lorsque plusieurs mesures sont impliquées, le domaine de valeurs thématiques risque de varier d'une mesure à l'autre, ce qui nécessiterait une fusion des domaines de valeurs avant de procéder à l'analyse spatiale matricielle.

En complément des stratégies de réduction de la taille des jeux de données énoncés précédemment, l'**uniformisation du domaine de valeurs** doit être privilégiée pour favoriser l'exécution à la volée des analyses spatiales matricielles. Cette stratégie consiste à utiliser le même type de variable (ex. : entier ou réel) pour exprimer toutes les mesures du cube matriciel destiné à l'analyse spatiale matricielle à la volée. La nécessité de recourir à la fusion de couverture est ainsi grandement réduite. Seulement les cas où la résolution spatiale ou l'étendue spatiale sont différentes d'une couverture à l'autre nécessiteront le recours à la fusion de couvertures préalablement à l'exécution de l'analyse spatiale matricielle. Par exemple, cette stratégie d’uniformisation appliquée au scénario d’intégration des couvertures matricielles par le cube permet de s’assurer que le cube matriciel est toujours très favorable à l’analyse spatiale matricielle à la volée. En procédant ainsi, l’exploitation du cube matriciel, dans un contexte d’analyse spatiale matricielle à la volée, est beaucoup plus flexible pour la plupart des scénarios (voir Tableau 3.2 — Conditions d'exploitation pour l'analyse spatiale matricielle à la volée des scénarios d'intégration après une uniformisation des domaines de valeurs des mesures ci-dessous).

Tableau 3.1 — Résumé des caractéristiques des scénarios d'intégration des couvertures matricielles en fonction des perspectives du cube ou de la couverture

Perspectives		Intégration des couvertures matricielles						
		par le cube	par la mesure	par le niveau	par la dimension	par la hiérarchie	par le membre	par le fait
Cube	Cube	Une seule couverture	Une ou plusieurs couvertures	Une ou plusieurs couvertures	Une seule couverture	Une ou plusieurs couvertures	Une ou plusieurs couvertures	Une ou plusieurs couvertures
	Mesure	Une seule couverture	Une seule couverture	Une ou plusieurs couvertures	Une seule couverture	Une ou plusieurs couvertures	Une ou plusieurs couvertures	Une ou plusieurs couvertures
	Fait détaillé	Une seule cellule	Une seule cellule	Une seule cellule	Une couverture	Une couverture	Une couverture	Une couverture
Couverture	Étendue spatiale	Constante pour le cube	Variable entre les mesures	Variable entre les niveaux	Constante pour le cube	Variable entre les hiérarchies	Variable entre les membres	Variable entre les faits
	Résolution spatiale	Constante pour le cube	Variable entre les mesures	Variable entre les niveaux	Constante pour le cube	Variable entre les hiérarchies	Variable entre les membres	Variable entre les faits
	Domaine de valeurs	Variable entre les mesures	Variable entre les mesures	Variable entre les mesures	Variable entre les mesures	Variable entre les mesures	Variable entre les mesures	Variable entre les mesures
	Valeur thématique	Propre à chaque fait	Propre à chaque fait	Propre à chaque fait	Propre à chaque cellule du fait	Propre à chaque cellule du fait	Propre à chaque cellule du fait	Propre à chaque cellule du fait
Analyse spatiale matricielle à la volée	Très favorable pour	Tous les membres et une seule mesure	Tous les membres et une seule mesure	Un seul niveau et une seule mesure	Tous les membres et une seule mesure	Tous les membres de la même hiérarchie et une seule mesure	Un seul membre et une seule mesure	Un seul fait et une seule mesure
	Favorable pour	Tous les membres et plusieurs mesures	S.O.	Un seul niveau et plusieurs mesures	Tous les membres et plusieurs mesures	Tous les membres de la même hiérarchie et plusieurs mesures	Un seul membre et plusieurs mesures	Un seul fait et plusieurs mesures
	Défavorable pour	S.O.	Tous les membres et plusieurs mesures	Plusieurs niveaux et une seule ou plusieurs mesures	S.O.	Plusieurs membres de plusieurs hiérarchies et une ou plusieurs mesures	Plusieurs membres et une ou plusieurs mesures	Plusieurs faits et une ou plusieurs mesures

Tableau 3.2 — Conditions d'exploitation pour l'analyse spatiale matricielle à la volée des scénarios d'intégration après une uniformisation des domaines de valeurs des mesures

Perspectives		Intégration des couvertures matricielles						
		par le cube	par la mesure	par le niveau	par la dimension	par la hiérarchie	par le membre	par le fait
Analyse spatiale matricielle à la volée	Très favorable pour	Tous les membres et plusieurs mesures	Tous les membres et une seule mesure	Un seul niveau et plusieurs mesures	Tous les membres et plusieurs mesures	Tous les membres de la même hiérarchie et plusieurs mesures	Un seul membre et plusieurs mesures	Un seul fait et plusieurs mesures
	Défavorable pour	S.O.	Tous les membres et plusieurs mesures	Plusieurs niveaux et une seule ou plusieurs mesures	S.O.	Plusieurs membres de plusieurs hiérarchies et une ou plusieurs mesures	Plusieurs membres et une ou plusieurs mesures	Plusieurs faits et une ou plusieurs mesures

Enfin, le principal constat de cette analyse des scénarios d'intégration possibles est que les scénarios présentés ne sont pas tous favorables à l'analyse spatiale matricielle à la volée. Une des caractéristiques qui distingue les scénarios décrits précédemment réside dans l'**intention de hiérarchiser l'information géométrique**. L'intégration des couvertures matricielles par le cube, par la mesure, par la dimension, par la hiérarchie, par le membre ou par le fait ne permet pas de hiérarchiser géométriquement l'information. Les cellules matricielles doivent avoir la même taille, la même forme et la même orientation lorsqu'elles sont intégrées directement dans les faits. Pour ces scénarios, la hiérarchisation géométrique (variation de la résolution spatiale) n'est pas admissible. Or, l'intégration des couvertures matricielles par les dimensions, les hiérarchies et les niveaux ouvrent des opportunités pour hiérarchiser spatialement les couvertures matricielles qui seront approfondies à la section suivante (*cf.* section 3.4) sans faire varier la résolution spatiale.

L'autre aspect considérable pour l'intégration des couvertures matricielles réside dans la capacité ou la **volonté de partager les couvertures entre les éléments fondamentaux du cube**. Par le mot partage ici est sous-entendue l'association des faits et des couvertures ainsi que le nombre de faits associés à une même couverture. L'intention de partager une couverture et ses cellules peut notamment provenir d'un souci d'espace mémoire ou de performance. En effet, une seule couverture matricielle pour l'ensemble du cube occupe moins d'espace que plusieurs couvertures matricielles pour un même cube. Cependant, l'intention de partager une couverture entre les éléments du cube peut également provenir d'un intérêt de simplification de l'analyse spatiale matricielle. L'intégration d'une couverture matricielle par le cube constitue un modèle simple et est favorable à l'exécution à la volée des analyses spatiales matricielles. La combinaison des membres et des mesures du cube crée différentes couches (variables thématiques) d'une seule et même couverture. Ainsi, l'analyse spatiale matricielle multivariée est facilitée par ce référentiel spatial unique. L'intégration des couvertures par les dimensions ou par les hiérarchies produira un effet relativement semblable. À l'opposé, une intégration des couvertures par les niveaux entraine la mise en place de plusieurs couvertures aux attributs distincts (étendue spatiale, résolution spatiale et domaine de valeurs), ce qui complexifie l'analyse spatiale matricielle multivariée. Cependant, l'intégration des couvertures matricielles dans les niveaux de la dimension constitue une approche plus flexible justement parce que ces caractéristiques sont variables. Il y a donc un **effort de planification à faire lors du design du cube pour établir le scénario d'intégration le plus pertinent** selon la flexibilité recherchée, les gains d'espace mémoire ou la capacité de produire des analyses spatiales matricielles à la volée.

3.4 Hiérarchisation des couvertures matricielles dans les dimensions spatiales matricielles

Dans le monde OLAP, la dimension joue un rôle de premier plan pour l'interrogation et la découverte des connaissances. La hiérarchie qui la compose permet de structurer les membres d'une dimension sous forme d'arbre présentant des niveaux de granularité de l'information. Ces niveaux contribuent notamment à atteindre la performance d'interrogation recherchée. Dans un cadre plus large, la hiérarchie est une structure de l'information servant à établir une relation de subordination entre les objets d'un ensemble. Pour qu'une hiérarchie soit spatiale, elle doit considérer la localisation des objets mis en relation. La subordination spatiale est établie en considérant la sémantique des attributs descriptifs et spatiaux tels que la représentation géométrique, la localisation, le niveau de détail ou les relations topologiques entre les objets. C'est le cas notamment lorsque vient le temps de hiérarchiser des zones de code postal, des provinces ou des états à l'intérieur de leur pays respectif.

Les hiérarchies spatiales matricielles, dans le contexte de cette recherche, impliquent la présence de plusieurs couvertures matricielles. La mise en œuvre de ces hiérarchies entraine la superposition spatiale des géométries des cellules des couvertures. Une couverture est associée à chaque niveau de la hiérarchie. Les membres de la dimension pour leur part associés aux cellules des couvertures. La figure suivante illustre un exemple de hiérarchie de couvertures matricielles.

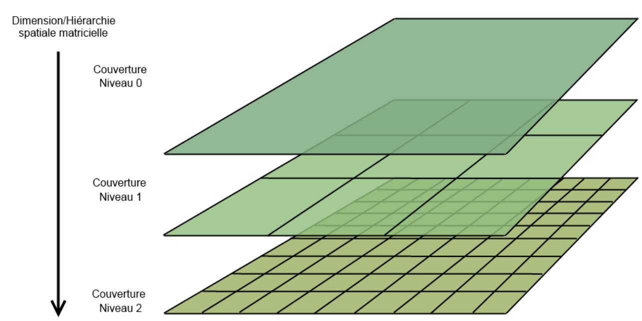


Figure 3.27 — Exemple de couvertures associées aux niveaux d'une hiérarchie

La nature de la relation entre les niveaux hiérarchiques dépend des attributs utilisés pour établir la relation et la sémantique associée à ces attributs. La hiérarchisation matricielle est le processus par lequel des attributs sont sélectionnés pour établir des règles et créer les relations de subordination spatiale entre les cellules. Les hiérarchies spatiales matricielles, telles les pyramides de tuiles [Goffe, 2011] des services géographiques, exploitent la taille et la position des tuiles pour établir la hiérarchie spatiale. C'est le cas également des dimensions spatiales matricielles présentant plusieurs niveaux telles que décrites précédemment (*cf.* chapitre 2.2.1). Ces hiérarchies matricielles exploitent la résolution spatiale et la position des cellules pour établir les règles de hiérarchisation. Or, nous croyons qu'il est possible de créer des hiérarchies spatiales de couvertures matricielles en considérant aussi la variable thématique de la couverture et non seulement sa résolution spatiale. Les hiérarchies ainsi formées créent des dimensions matricielles sans faire varier la résolution spatiale d'un niveau à l'autre.

Cette section décrit les caractéristiques des hiérarchies de couvertures matricielles dans un contexte SOLAP et propose des façons d'intégrer les couvertures matricielles dans ces hiérarchies. Également, elle vise à déterminer quels sont les scénarios de hiérarchisation des couvertures matricielles les plus favorables à l'exécution à la volée des analyses spatiales matricielles.

3.4.1 Hiérarchisation géométrique et thématique des couvertures matricielles

A priori, une hiérarchie spatiale de couvertures matricielles possède des caractéristiques inhérentes à ces couvertures, aux couches et aux cellules matricielles qui les composent (ex. : étendue spatiale, résolution, orientation, domaine de valeurs). L'observation de ces caractéristiques vise à mieux comprendre les relations entre les niveaux de la hiérarchie. Une hiérarchie de couvertures matricielles est formée de plusieurs couvertures[10] matricielles où chaque niveau de la hiérarchie est associé à sa propre couverture. Les géométries des cellules matricielles formant une couverture sont associées aux membres formant le niveau de la hiérarchie. Les caractéristiques d'une hiérarchie de couvertures matricielles sont les suivantes :

- la relation entre les cellules de niveaux hiérarchiquement adjacents doit être exprimée par la relation 1..*N* où *N* est une valeur entière;
- la valeur *N* doit être constante entre deux niveaux hiérarchiquement adjacents pour l'ensemble des cellules matricielles;

Il découle certains constats de l'observation de ces caractéristiques essentielles. Le premier est que **l'étendue spatiale des couvertures matricielles doit être identique à tous les niveaux**. Un manquement à cette règle entrainerait une incapacité de mettre en œuvre la relation 1..*N* où *N* est constant sur l'ensemble de la couverture. D'autre part, comme il s'agit de couvertures matricielles, il est **exclu que des cellules matricielles, qui sont en fait les membres de la dimension, ne soient pas matérialisées**. La mise en œuvre d'une hiérarchie de couvertures matricielles n'exclut pas que la valeur de *N* soit variable à l'intérieur de la hiérarchie pour des couvertures qui ne sont pas associées à des niveaux adjacents.

---

[10] Le terme couverture est utilisé pour simplifier l'explication. Pour les cas où la résolution spatiale et le domaine de valeurs sont contant pour l'ensemble des couvertures, il pourrait être question de couches d'une même couverture plutôt que de couvertures différentes.

La variation de la résolution spatiale fut jusqu'à maintenant la seule variable considérée dans le processus permettant de créer des dimensions spatiales matricielles géométriques. Il faut rappeler qu'une dimension spatiale matricielle géométrique (*cf.* Section 2.2.1) est constituée seulement de données matricielles. Elle ne comporte pas de données vectorielles ou de données spatiales nominales comme c'est le cas pour les dimensions spatiales non-géométriques, géométriques vectorielles, hybrides, mixtes matricielles ou mixtes-hybrides. Cette section explore l'utilisation de la variable thématique associée à la couverture matricielle pour établir la hiérarchie en complément de la résolution spatiale. En mettant en interaction ces deux variables, il est possible d'élaborer des combinaisons de règles formant des scénarios de hiérarchisation distincts. La figure suivante illustre les combinaisons des caractéristiques des couvertures matricielles qui sont discutées dans cette section pour former des dimensions matricielles géométriques.

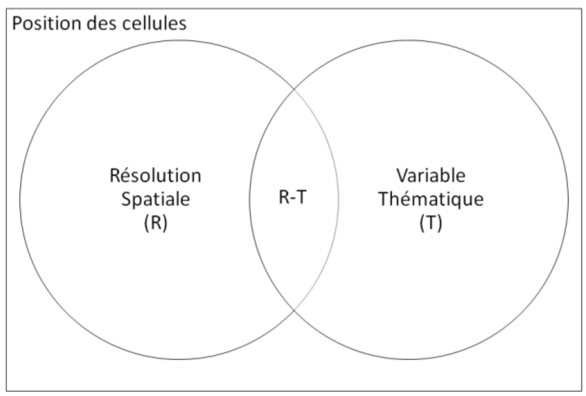


Figure 3.28 — Variables participant à la hiérarchisation des couvertures matricielles

Considérant que tous ces scénarios visent à produire des hiérarchies spatiales, la position des cellules formant les couvertures sont toujours impliquées dans le processus. Il est possible de dénombrer trois (3) scénarios pour mettre en œuvre des hiérarchies de couvertures matricielles : 1) géométrique (résolution spatiale); 2) thématique ainsi que 3) géométrique et thématique. Les principales caractéristiques de ces scénarios sont présentées ci-dessous.

Hiérarchisation spatiale matricielle géométrique (R)

Cette hiérarchisation consiste à mettre en relation les couvertures matricielles à l'aide de la position des cellules et de la résolution spatiale des couvertures. Les couvertures matricielles sont chacune associées à un niveau de la hiérarchie. Les membres du niveau sont liés aux cellules matricielles formant la couverture. La résolution spatiale des couvertures évolue au rythme du gradient de profondeur de la hiérarchie.

La hiérarchie est créée par un procédé géométrique d'agrégation (détaillé vers général) ou de division (général vers détaillé). La position et la résolution sont deux variables indépendantes. La variable thématique est une variable dépendante du processus de hiérarchisation, car elle est affectée par le procédé de hiérarchisation (agrégation ou division). L'agrégation nécessite de synthétiser ou « *summarize* » les valeurs thématiques des cellules fines, à l'aide d'opérateurs statistiques considérant la localisation, les relations sémantiques et topologiques, afin de conserver une forme de représentativité pour les valeurs agrégées (moyenne, somme, etc.). Or, le procédé par division entraine une propagation des valeurs thématiques observées dans la cellule plus grossière vers les cellules fines nouvellement formées. La figure suivante illustre les procédés d'agrégation et de division.

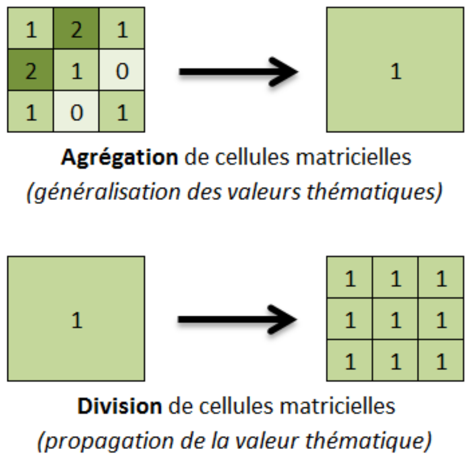
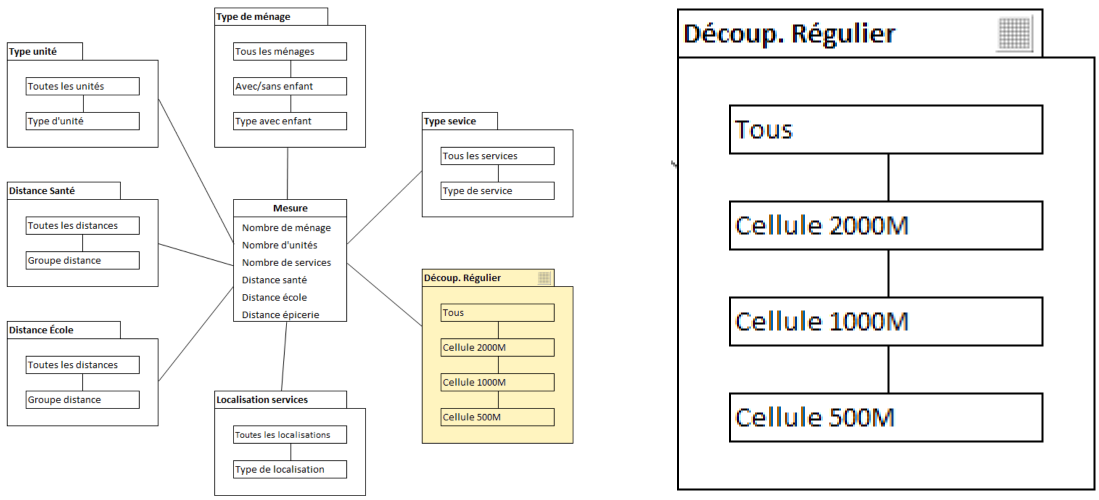


Figure 3.29 — Procédés permettant de créer des nouvelles couvertures matricielles

Le processus de division est notamment utilisé lors de traitements visant à fusionner ou à rehausser des images [Jensen, 2005]. Dans le contexte d'application SOLAP, le procédé d'agrégation est sans équivoque le plus répandu. Dans la dimension matricielle géométrique, dont la hiérarchie est produite en considérant uniquement la position des cellules et la résolution des couvertures, les niveaux sont identifiés par leur résolution spatiale spécifique comme c'est le cas dans le cube matriciel Muscamags [McHugh, 2008] illustré ci-dessous.



A. Cube matriciel Muscamags **B. Hiérarchie spatiale matricielle géométrique**

Figure 3.30 — Exemple de dimension spatiale géométrique obtenu par variation de la résolution spatiale

Dans ce type de hiérarchie matricielle, la résolution spatiale est constante à l'intérieur d'un niveau hiérarchique, mais varie au rythme du gradient de profondeur des niveaux de la hiérarchie. La variation de la résolution spatiale est exprimée par un entier (**δ**) décrivant la relation entre les cellules d'un niveau observé (**ρ**) et d'un niveau adjacent précédent (**ρ-1**) ou suivant (**ρ+1**) dans la hiérarchie. Cette valeur **δ** est variable entre les niveaux de la hiérarchie, mais peut également être constante pour tous les niveaux. La matérialisation de la hiérarchie spatiale entraine la mise en relation et la superposition de plusieurs couvertures matricielles, possédant des résolutions spatiales différentes, telles qu'illustrées par la figure suivante. La résolution spatiale y est présentée tel un attribut variable entre les niveaux, mais constant pour le niveau observé. La valeur **δ** décrivant la relation spatiale entre les niveaux est constant pour la figure (a) où $\delta_{(0;\,1)} = 9 = 3^2$ et $\delta_{(1;\,2)} = 9 = 3^2$ alors qu'elle est variable pour la figure (b) où $\delta_{(0;\,1)} = 9 = 3^2$ et $\delta_{(1;\,2)} = 4 = 2^2$. Le chapitre suivant (*cf.* section 4.2.1) discute de l'impact de la variation de δ sur l'analyse spatiale matricielle à la volée.

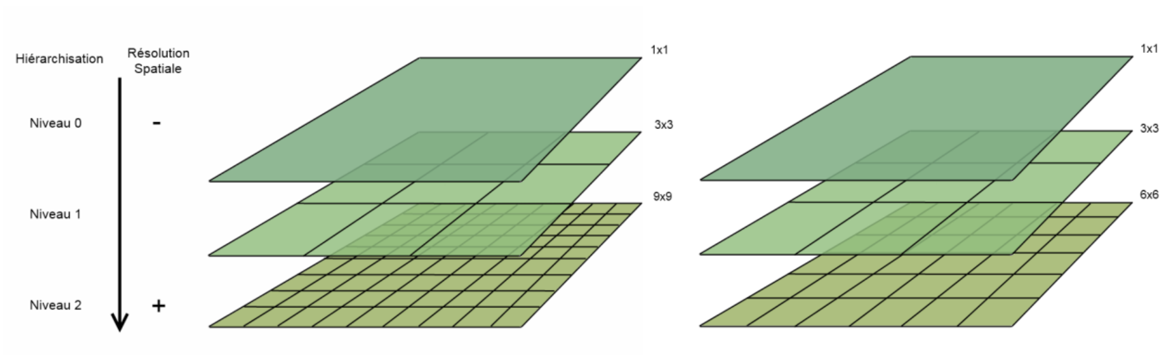


Figure 3.31 — Exemples de hiérarchies de couvertures matricielles géométriques

Ce type de hiérarchie matricielle est très répandu en géomatique. Notamment, c'est une des méthodes utilisées pour établir des niveaux de granularités dans les pyramides matricielles ou les pyramides de tuiles [WMTS, 2012] des services cartographiques. Aussi, il s'agit de la méthode décrite précédemment (*cf*. section 2.2.1) pour intégrer le matriciel dans les dimensions spatiales matricielles.

Pour l'analyse spatiale matricielle à la volée, ce type de hiérarchie possède comme principal intérêt d'utiliser moins de ressources informatiques pour le traitement des analyses spatiales matricielles dans les niveaux agrégés par comparaison aux niveaux détaillés. En effet, un seul objet spatial d'un niveau agrégé, en l'occurrence une cellule matricielle, occupe le même espace géographique que plusieurs cellules des niveaux plus détaillés. Toutefois, dans une telle situation, la performance accrue est atteinte au détriment de la précision thématique. En effet, la cellule agrégée contient une valeur thématique synthèse qui n'est pas nécessairement appropriée au niveau de détail souhaité pour l'analyse spatiale matricielle.

Hiérarchisation spatiale matricielle thématique (T)

Cette hiérarchisation consiste à mettre en relation les couvertures matricielles à l'aide de la position des cellules et de la variable thématique des couvertures. La hiérarchie est créée par la superposition de couvertures. La subordination provient de la variable thématique associée à la couverture. Les couvertures matricielles sont chacune associées à un niveau de la hiérarchie. Les membres du niveau ne sont pas associées aux cellules matricielles formant la couverture, mais plutôt à la variable thématique et au domaine de valeurs présent sur la couverture. La figure suivante illustre l'évolution de la variable thématique en fonction du gradient de profondeur de la hiérarchie.

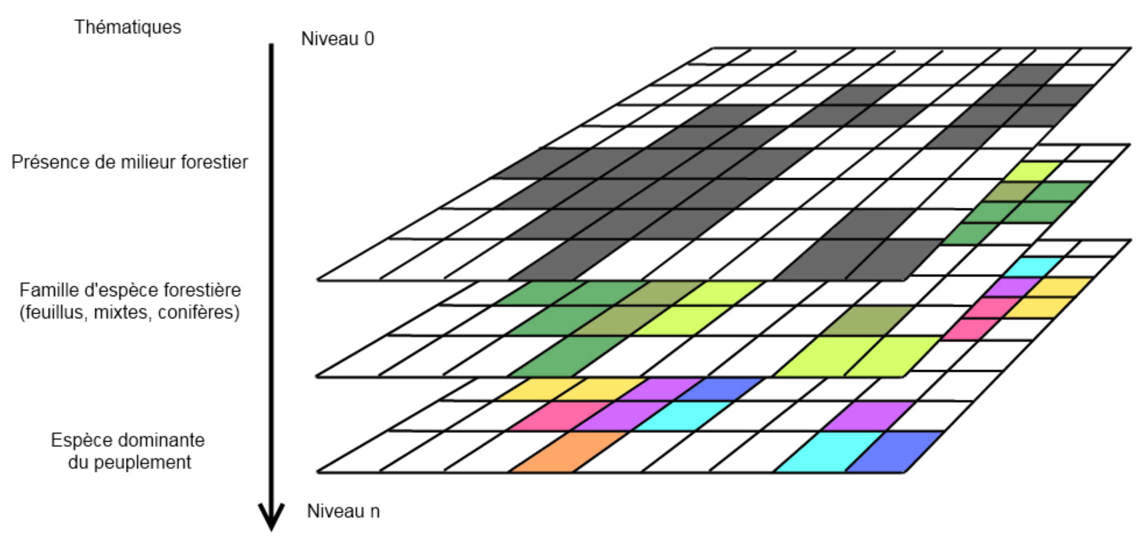


Figure 3.32 — Exemple de hiérarchie spatiale matricielle thématique

Cette hiérarchie matricielle constitue une version simple se distinguant de la superposition spatiale par l'évolution de la variable thématique qui confère une relation de subordination aux cellules entre elles. Ce type de hiérarchie matricielle correspond à une couverture associée à une dimension où chaque couche représentant une variable thématique est associée à un niveau de la hiérarchie (*cf.* section 3.3.2). Pour ce type de hiérarchie, la résolution spatiale est constante pour l'ensemble de la hiérarchie. Le cube spatial de foresterie [Proulx et *al.*, 2006] qui a été modifiée précédemment constitue un exemple de hiérarchie spatiale matricielle thématique. La couverture matricielle a été intégrée à la dimension « Essence » afin d'en faire une dimension spatiale thématique. La figure suivante présente le modèle multidimensionnel adapté du cube matriciel.

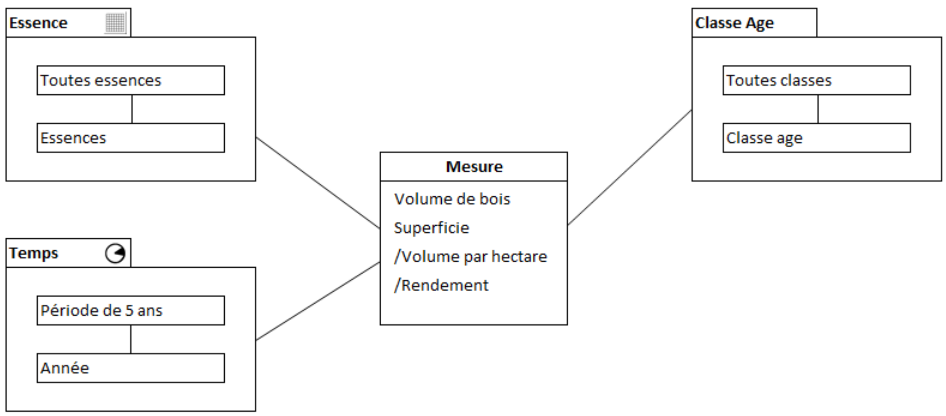


Figure 3.33 — Modèle multidimensionnel d'un cube matriciel en foresterie

Dans ce type de hiérarchie, la continuité des couvertures peut s'avérer problématique. En effet, il peut se former des îlots ou ensembles de cellules qui ne seront pas hiérarchisés, car ils sont hors de la portée de la thématique d'intérêt (ex. : milieu forestier vs autres milieux). La continuité doit être assurée sur tout le domaine d'intérêt afin de former une couverture. Toutefois, les zones en périphéries peuvent être exclues du domaine spatial d'intérêt. Une façon de remédier à cette problématique est de limiter l'étendue spatiale de la couverture pour assurer la continuité sur l'ensemble de la couverture. Ainsi, il sera question de définir l'étendue spatiale de la couverture à l'aide d'une enveloppe convexe ou « *convex hull* » plutôt qu'un rectangle minimal englobant ou « *minimal bounding rectangle* » (voir figure ci-dessous).

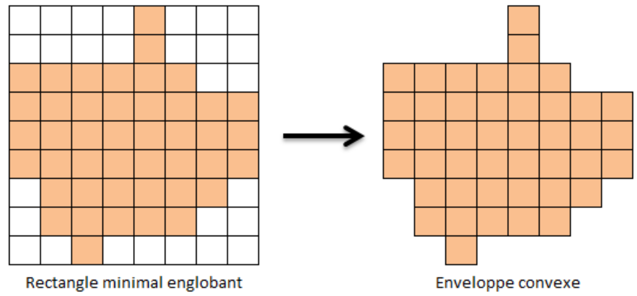


Figure 3.34 — Étendue spatiale de la couverture limitée à la zone d'intérêt

L'étendue spatiale de la couverture définie par l'enveloppe convexe assure la continuité sur l'ensemble de la zone d'intérêt pour la hiérarchisation. Ainsi, seules les cellules contribuant à la hiérarchisation sont exploitées. Toutefois, les caractéristiques d'une hiérarchie de couvertures doivent être respectées (**δ** $\in$ [1, N] où N est contant entre deux niveaux hiérarchiquement adjacents). Les cas de discontinuité à l'intérieur de la zone d'intérêt sont plus complexes et ne sont pas favorables à une exécution à la volée des analyses spatiales matricielles. Ils doivent être traités en amont (*cf.* section 3.2.1 Discontinuité du domaine spatial des cubes matriciels) de l'analyse spatiale matricielle et de la création du cube matriciel.

Enfin, la hiérarchie spatiale matricielle thématique est très favorable à l'analyse spatiale matricielle à la volée, car tous les faits partagent la même couverture et incidemment les mêmes caractéristiques de couvertures (*cf.* section 3.3.2 — Couverture matricielle associée à une dimension). Les ressources informatiques nécessaires au traitement des niveaux plus généraux sont les mêmes que pour traiter les niveaux plus détaillés. Il n'y aura théoriquement pas de dégradation de performance en produisant des analyses sur les niveaux détaillés. De plus, le recours à une enveloppe convexe, pour délimiter la zone d'intérêt, peut s'avérer bénéfique pour le traitement des analyses. Toutefois, les couvertures matricielles qui ne sont pas carrées présentent un niveau de complexité additionnel qui ne sera pas abordé plus en détail dans le cadre de ces travaux de recherche.

Hiérarchisation spatiale matricielle géométrique et thématique (R-T)

Cette hiérarchisation consiste à mettre en relation simultanément la position des cellules, la résolution spatiale et la variable thématique des couvertures. La hiérarchie spatiale matricielle est créée par un procédé d'agrégation spatiale et thématique. Les couvertures matricielles sont chacune associées à un niveau de la hiérarchie. Les membres du niveau sont liés aux cellules matricielles formant la couverture. La résolution spatiale et la variable thématique évoluent au rythme du gradient de profondeur de la hiérarchie. La figure suivante illustre un exemple de hiérarchie spatiale matricielle géométrique et thématique. Le niveau le plus général indique la présence ou l'absence de milieu forestier et le niveau plus détaillé la famille d'espèce dominante à des résolutions spatiales différentes.

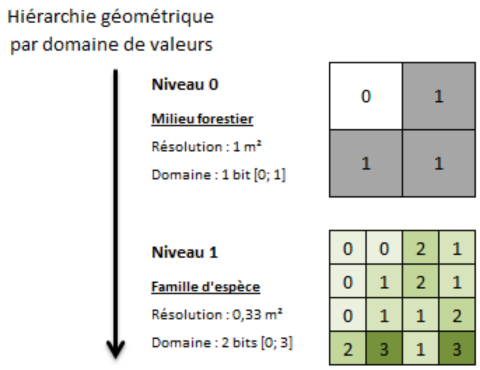


Figure 3.35 — Exemple de hiérarchie géométrique et thématique

Dans ce type de hiérarchie, la résolution spatiale est constante à l'intérieur d'un même niveau de la hiérarchie. Pour illustrer la distinction entre cette méthode de hiérarchisation et celles présentées précédemment, le cube matriciel de foresterie [Proulx et *al.*, 2006] a été adapté. La dimension « Essence » comporte maintenant deux couvertures distinctes possédant des résolutions spatiales différentes. La figure suivante présente le modèle multidimensionnel adapté du cube matriciel.

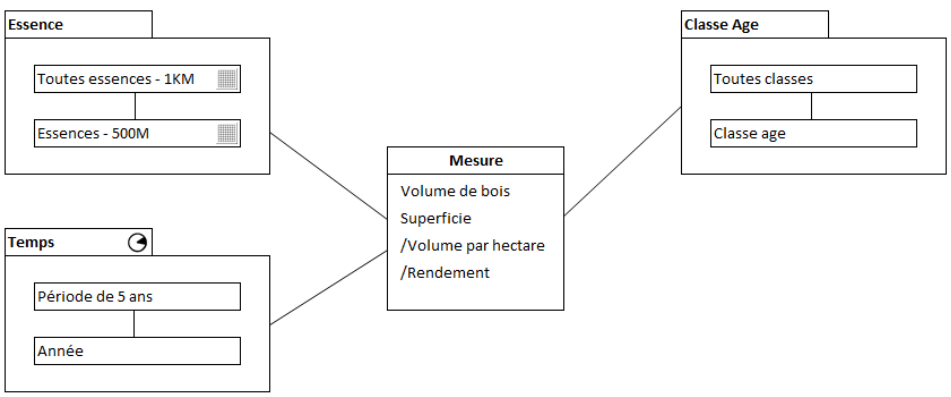


Figure 3.36 – Modèle multidimensionnel intégrant une hiérarchie géométrique et thématique

L'exemple précédent illustre un cas où les deux variables indépendantes (résolution spatiale et valeurs thématiques) convergent, c.-à-d. que leur niveau de détail s'accroît au rythme du gradient de profondeur de la hiérarchie (général vers détaillé). Toutefois, il est théoriquement possible que les deux variables indépendantes évoluent de façon divergente et d'obtenir aussi une hiérarchie de couverture matricielle. Ce cas est possible si la résolution spatiale est plus fine près des feuilles et plus grossière près de la racine (général vers détaillé), mais que la variable thématique évoluait à l'inverse (détaillé vers général). La figure suivante illustre un exemple de divergence entre les variables indépendantes de résolutions spatiale et variable thématique des couvertures.

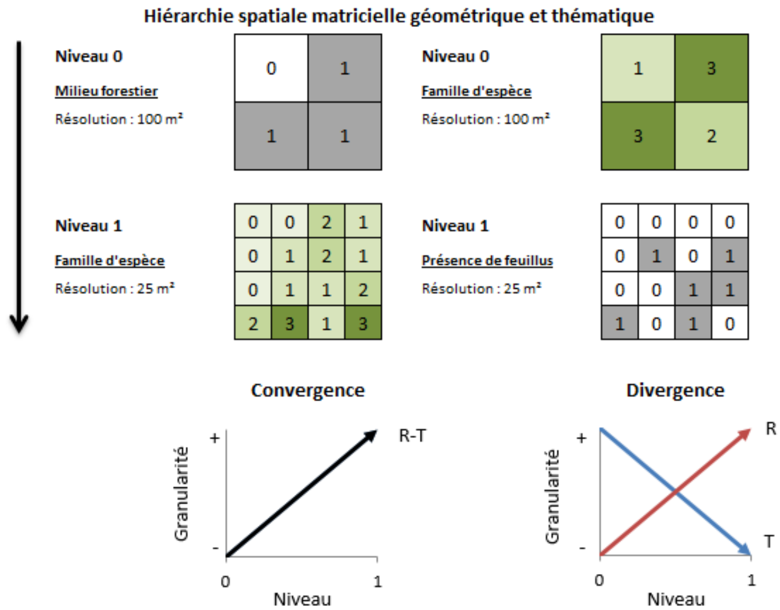


Figure 3.37 — Convergence et divergence des variables indépendantes

Un autre cas se pose où la résolution spatiale est plus détaillée près de la racine et plus agrégée près des feuilles (détaillé vers général). Toutefois, il ne s'agirait plus d'une hiérarchie de couvertures applicable au SOLAP, car l'arborescence ainsi formée serait inversée (plus grand nombre de membre près de la racine) compromettant ainsi les performances d'interrogation recherchées en OLAP.

Enfin, ce type de hiérarchie possède comme principal intérêt pour l'analyse spatiale matricielle à la volée d'utiliser moins de ressources informatiques pour le traitement des analyses spatiales matricielles pour les niveaux agrégés que pour les niveaux détaillés. Elle ne possède toutefois pas l'avantage de posséder une seule couverture matricielle comme c'est le cas pour les hiérarchies spatiales matricielles thématiques.

3.4.2 Constats et recommandation sur la hiérarchisation des couvertures matricielles dans les cubes matriciels pour l'analyse spatiale à la volée

La hiérarchisation des couvertures matricielles est régie par des procédés de superposition, d'agrégation ou de division des cellules. Les règles permettant de créer ces hiérarchies reposent sur l'interaction de la position des cellules, la résolution spatiale (taille des cellules) et la variable thématique propre à la couverture (*cf.* Figure 3.28 Variables participant à la hiérarchisation). Dans le cadre de cette recherche, une réflexion sur l'apport théorique du domaine de valeurs des couvertures matricielles dans le processus de hiérarchisation a été réalisée. L'intérêt d'utiliser le domaine de valeurs pour la hiérarchisation provenait du désir d'optimisation des ressources informatiques pour le traitement des analyses spatiales. En théorie, une couverture matricielle comportant un niveau de détail plus faible (ex. : écoumène/non-écoumène) est moins exigeante à traiter qu'une couverture plus détaillée (ex. : température de surface). Cette approche est notamment viable pour créer des hiérarchies de couvertures matricielles dans un cadre géomatique de traitement d'image. La figure suivante illustre un exemple de hiérarchie matricielle par domaine de valeurs.

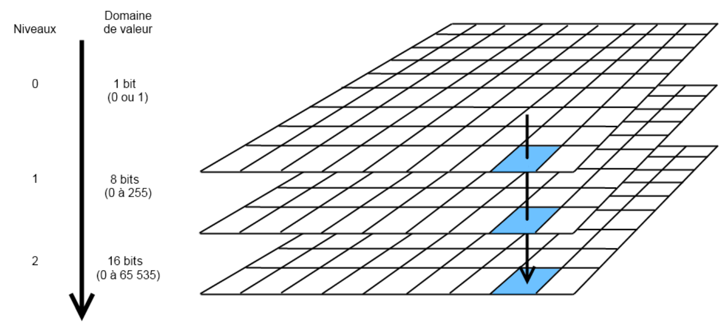


Figure 3.38 — Exemple de hiérarchie matricielle par domaine de valeur

Cependant, dans le contexte SOLAP, le domaine de valeurs est associé aux mesures du cube (ex. : volume de bois en m³) ou à la variation thématique des membres d'une dimension (ex. : toutes les essences, famille d'essence, essence individuelle). Pour la hiérarchisation des couvertures matricielles, nous avons choisi d'inclure le domaine de valeurs à la variable thématique de la couverture matricielle. De cette façon, le domaine de valeurs n'affecte pas la précision des mesures du cube matriciel. Enfin, son effet bénéfique potentiel sur le temps de traitement est intégré à la hiérarchisation thématique par la matérialisation d'un nombre plus grand ou plus restreint de membres selon le niveau.

La hiérarchisation spatiale des couvertures matricielles apporte une nouvelle perspective pour la définition de ce qu'est une dimension géométrique matricielle. Il n'est pas nécessaire de faire évoluer la résolution spatiale au gré des niveaux de la hiérarchie pour obtenir une dimension spatiale matricielle. Il est donc essentiel de proposer une nouvelle définition de cette dimension. Cette définition doit notamment être compatible avec la distinction de la hiérarchie matricielle géométrique et de la hiérarchie matricielle thématique.

> **Dimension spatiale matricielle** : Une dimension spatiale matricielle est constituée d'au moins un niveau matriciel. Si le nombre de niveaux est supérieur à un, des caractéristiques de la couverture matricielle (ex. : résolution spatiale et variable thématique) sont mises en relation avec la position des cellules pour former la hiérarchie. L'agrégation des cellules est de nature thématique ou géométrique en fonction des combinaisons de caractéristiques mises en œuvre pour établir les niveaux de la hiérarchie.

Cette définition plus générale prend en considération les observations décrites ci-dessus et est compatible avec la notion de couverture matricielle essentielle à l'analyse spatiale matricielle.

En ce qui concerne l'exécution à la volée des analyses spatiales matricielles, il est théoriquement possible qu'il y ait une dégradation de la performance pour les hiérarchies de couvertures géométriques (thématiques ou non). En effet, le nombre de cellules nécessaires pour couvrir le même territoire aux niveaux agrégés est inférieur au nombre de cellules au niveau détaillé. L'exécution est théoriquement plus rapide lorsque le niveau de détail est moindre. Pour leur part, les hiérarchies matricielles thématiques devraient en théorie offrir une performance plus constante, car le nombre de cellules demeure constant à chaque niveau. Afin d'obtenir de meilleurs temps de traitement avec les hiérarchies matricielles thématiques, il est nécessaire d'utiliser les méthodes d'optimisation usuelles consistant à réduire l'étendue spatiale, le nombre de membre ou de fait dans le cube matriciel.

Toutefois, le fait de pouvoir interroger une étendue spatiale similaire à l'aide de cellules matricielles agrégées constitue un élément favorable à l'exécution d'analyses spatiales matricielles à la volée. Cette analyse spatiale matricielle devrait notamment être capable d'interroger progressivement les valeurs thématiques à des niveaux agrégés et ensuite les plus détaillées afin de limiter l'étendue spatiale d'intérêt à chaque itération. C'est d'ailleurs cette idée qui est à l'origine du concept des analyses spatiales matricielles multi-niveaux (*cf.* section 4.2) présentés au chapitre suivant.

3.5 Modélisation physique des hiérarchies de couvertures matricielles

Les cubes matriciels reposent sur l'intégration des couvertures matricielles dans les faits directement ou à l'aide d'une dimension matricielle. Lorsque l'intégration se fait à l'aide d'une dimension il est possible d'avoir un seul ou plusieurs niveaux. Une dimension matérialisant un seul niveau est dite plate et ne matérialise pas de hiérarchie. Les dimensions plates sont des dimensions dont tous les membres se situent au même niveau que la racine. Or, en présence d'une hiérarchie dans le modèle physique de la dimension, la hiérarchie peut prendre plusieurs formes. Les formes des hiérarchies sont les suivantes : balancée, non-balancée et non-stricte. L'étude de la forme des hiérarchies revêt son importance pour l'intégration des couvertures matricielles. *A priori*, ce ne sont pas toutes les formes de hiérarchie qui sont aptes à matérialiser des hiérarchies de couvertures matricielles. Cette section vise à identifier quelles sont les formes de hiérarchies qui permettent de matérialiser une hiérarchie de couvertures matricielles.

L'étude de la forme des hiérarchies dépend essentiellement de trois critères : la présence ou l'absence de hiérarchie, la cardinalité entre les membres de niveaux différents; et la symétrie de l'arborescence de la hiérarchie. La figure suivante illustre le processus de classification des formes des différentes hiérarchies dimensionnelles spatiales.

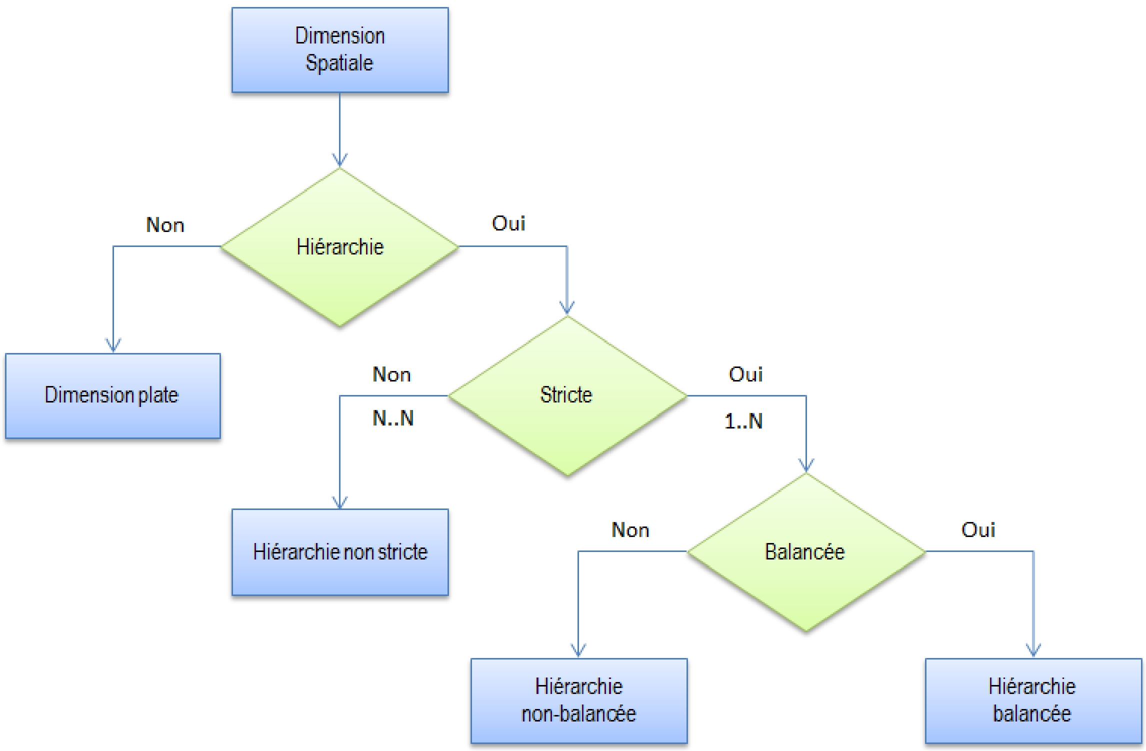


Figure 3.39 – Clé d'identification de la forme des hiérarchies

Les hiérarchies de couvertures matricielles doivent minimalement être composées de deux couches thématiques pour une même couverture ou de deux couvertures superposées. La relation δ {1;N} | N ε ℕ* exprime la cardinalité de la relation entre les membres de couvertures ou de couches logiquement adjacentes. La valeur de N est constante pour deux couvertures ou couches logiquement adjacentes (ρ et $\rho^{\pm 1}$), mais peut varier à l'intérieur de la hiérarchie pour des couches non-adjacentes (ex. : ρ et $\rho^{\pm 2}$). De ces règles, il est possible de déduire que l'étendue spatiale des couvertures matricielles hiérarchisées doit être identique et le recouvrement spatial doit être complet (aucune cellule manquante) pour les couvertures entre elles. Rappelons également qu'une couverture matricielle doit présenter une grille homogène, formée de cellules carrées et posséder une géométrie pour chaque coordonnée du jeu de données. À la fin de cette section se trouve un tableau portant sur les modèles physiques des hiérarchies dimensionnelles qui récapitule les caractéristiques de chaque forme de hiérarchie décrite ci-dessous.

Hiérarchie non-stricte

La hiérarchie non-stricte est une hiérarchie où la cardinalité entre les niveaux ρ et $\rho^{\pm 1}$ est exprimée par la relation N..N. Ce cas se produit notamment lorsqu'il est nécessaire de mettre en œuvre une correspondance entre des jeux de données vectorielles et matricielles. L'appariement n'est généralement pas parfait entre les deux jeux de données. La représentation de l'arborescence sous forme de graphe (figure ci-dessous) de cette hiérarchie met en évidence son incapacité à matérialiser une hiérarchie de couverture. La présence d'un lien double représente la liaison non-stricte entre les éléments du niveau 1 et du niveau 2.

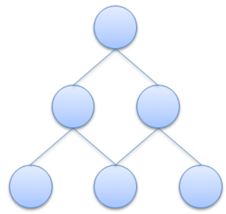


Figure 3.40 — Hiérarchie non-stricte illustrée sous la forme d'un graph

Cette forme de hiérarchie, même lorsqu'elle est symétrique, ne peut matérialiser une hiérarchie de couvertures matricielle à cause du non-respect de la constance de N pour deux niveaux contigus **δ** {1;N} où N est variable entre les membres.

Hiérarchie non-balancée

La hiérarchie non-balancée constitue une arborescence déséquilibrée. Représenté sous la forme d'un graphe, celui-ci est asymétrique. Les relations entre les niveaux ρ et $\rho^{\pm 1}$ sont strictes et exprimées par 1..N. Toutefois, N n'est pas constant pour tous les membres d'un même niveau dans cette forme de hiérarchie.

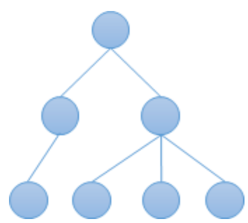


Figure 3.41 — Hiérarchie non-balancée illustrée sous la forme d'un graph

Ce cas se produit notamment lorsque l'étendue spatiale ou le recouvrement des couvertures matricielles n'est pas identique. Il n'est donc pas possible de matérialiser une hiérarchie de couvertures matricielles à l'aide de cette forme de hiérarchie.

Hiérarchie balancée

La hiérarchie balancée est caractérisée par une arborescence équilibrée. Elle présente, lorsqu'illustrée sous la forme d'un graphe, une symétrie évidente. Les relations entre les niveaux ρ et $\rho^{\pm 1}$ sont strictes et exprimées sous la forme 1..N. Toutefois, bien que la valeur de N tende à être constante, il existe des cas où N est variable pour des membres d'un même niveau. La figure ci-dessous illustre un graph représentant un cas particulier de hiérarchie balancée où N est variable à l'intérieur d'un niveau. Les membres au centre du graphe possèdent un nombre d'enfants différent par rapport aux membres situés aux extrémités sur le niveau médian.

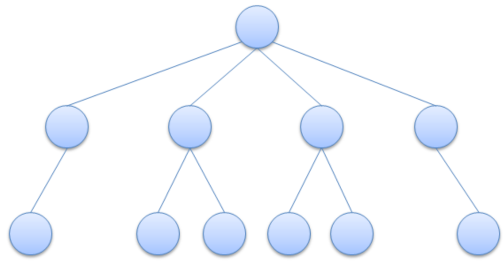


Figure 3.42 — Hiérarchie balancée illustrée sous la forme d'un graph

La hiérarchie balancée, parce qu'elle admet une variabilité de sa valeur N au sein d'un même niveau, ne peut pas matérialiser une hiérarchie de couvertures matricielles dans tous les cas possibles. La définition de deux nouvelles formes de hiérarchies s'avère nécessaire afin de décrire la relation **δ** entre les niveaux d'une hiérarchie de couvertures matricielle. Se pose alors deux possibilités. La première consiste à avoir une relation **δ** {1;N} où N est constant entre deux niveaux contigus. La seconde constitue une spécialisation de la première visant à obtenir une relation **δ** {1;N} où N est de valeur constante pour l'ensemble de la hiérarchie.

Hiérarchie aux niveaux homogènes

Forme de hiérarchie constituée d'une arborescence équilibrée. Illustrée sous forme d'un graphe, cette hiérarchie est symétrique. Les relations de cardinalité entre les niveaux contigus ρ et $\rho^{\pm 1}$ sont strictes et exprimées sous la forme δ {1;N}. La valeur de N est une constante pour tous les membres d'un même niveau. Cette valeur est toutefois variable d'un niveau à l'autre. La figure suivante illustre un exemple d'arborescence décrivant la hiérarchie homogène.

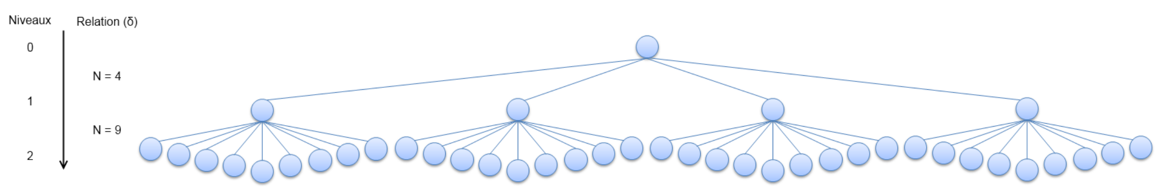


Figure 3.43 — Hiérarchie homogène illustrée sous la forme d'un graph

Hiérarchie uniforme

La hiérarchie uniforme se distingue de la hiérarchie aux niveaux homogènes par le fait qu'elle présente en plus la même relation de cardinalité pour tous les niveaux qui la compose. Tous les couples de niveaux ρ et $\rho^{\pm 1}$ possèdent une cardinalité **δ** {1;N} où la valeur de N est la même pour tous les niveaux. La figure suivante illustre un exemple d'arborescence décrivant la hiérarchie uniforme.

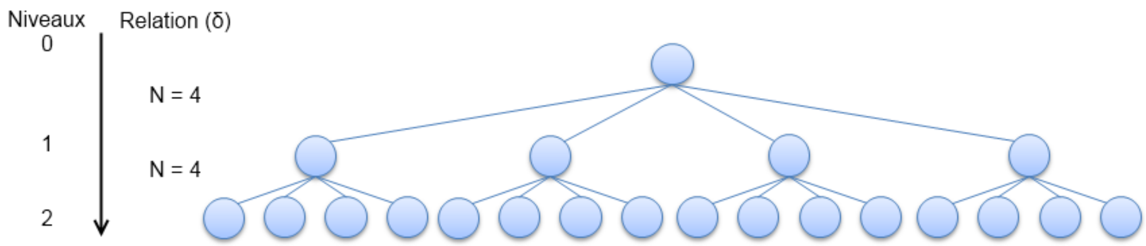


Figure 3.44 — Hiérarchie uniforme illustrée sous la forme d'un graph

Ce cas représente une hiérarchie où la résolution spatiale évolue à l'intérieur de la hiérarchie selon le gradient de profondeur des niveaux. Dans le cas d'une hiérarchie n'impliquant pas de changement de résolution spatiale, l'arborescence décrivant les niveaux est illustrée simplement par une relation **δ** {1;1}. Le tableau suivant décrit les formes des hiérarchies et en résume les principales caractéristiques.

Tableau 3.3 — Modèles physiques des hiérarchies dimensionnelles

Hiérarchie	**Symétrie**	**Cardinalité**	**Caractéristique de la relation δ**
Non-stricte	Non	N.. N	N est variable pour tous les membres
Non-balancée	Non	1..N	N est variable pour un niveau
Balancée	Oui	1..N	N est variable pour un niveau
Homogène	Oui	1..N	N est constant pour un niveau, variable entre les niveaux
Uniforme	Oui	1..N	N est constant pour l'ensemble de la hiérarchie

L'ajout de ces nouvelles hiérarchies modifie le schéma décisionnel d'identification de la forme des hiérarchies. La figure suivante présente une extension de la clé d'identification. L'ajout de deux points de contrôle est nécessaire pour évaluer s'il s'agit d'une hiérarchie homogène ou uniforme dans le cas où la hiérarchie est balancée.

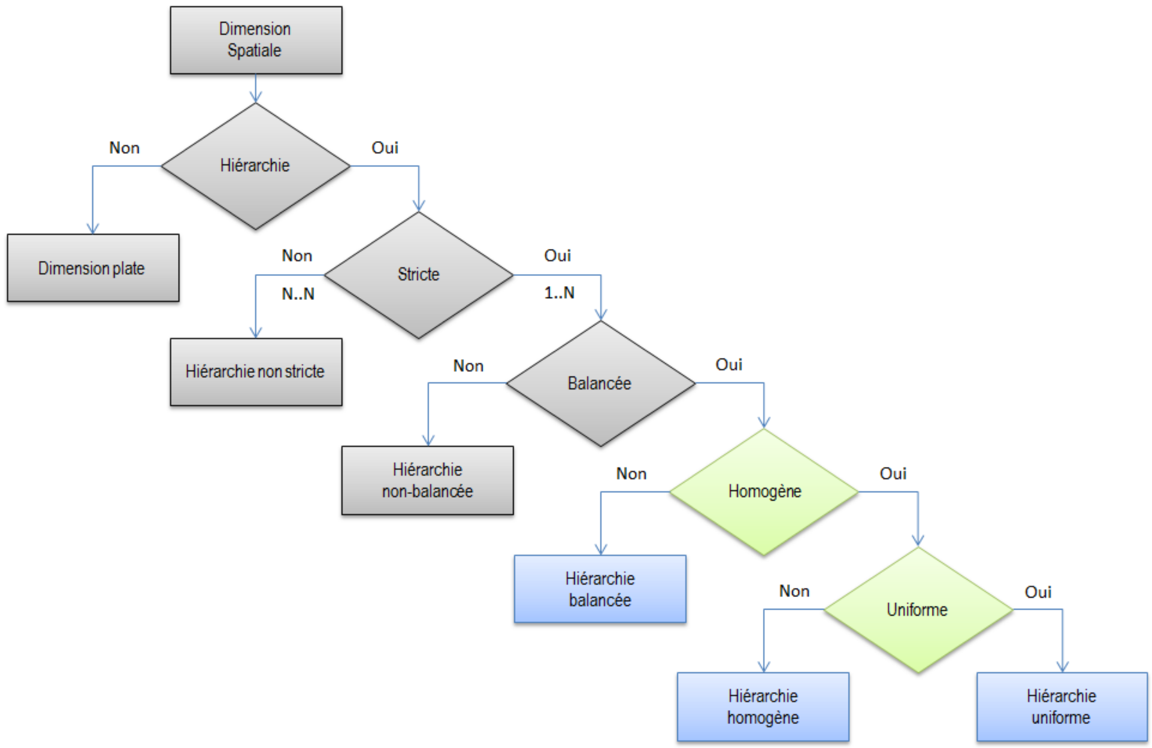


Figure 3.45 – Extension de la clé d'identification de la forme des hiérarchies

Le point de contrôle sur l'homogénéité nécessite le dénombrement des membres de niveaux adjacents ρ et $\rho^{\pm 1}$ afin de déterminer la nature de la relation **δ** {1;N} entre les niveaux. Le point de contrôle portant sur l'uniformité valide le nombre de membres parents-enfants pour tous les membres de tous les niveaux de la hiérarchie. Enfin, seules les hiérarchies homogènes et uniformes permettent de matérialiser des hiérarchies de couvertures matricielles. Ce sont les seules formes de hiérarchies qui assurent la constance de la valeur N entre deux niveaux adjacents. Le chapitre suivant explore l'effet de la forme de la hiérarchie sur le temps de traitement des analyses spatiales matricielles.

3.6 Conclusion

Ce chapitre a permis d'approfondir le cadre théorique entourant la modélisation matricielle multidimensionnelle. L'intégration de couvertures matricielles, aux fins d'analyse spatiale, repose sur les caractéristiques du cube matriciel et des couvertures elles-mêmes. Cette analyse a permis de dégager deux grandes tendances liées à l'intention de modélisation favorisant une analyse thématique, par couche d'une seule couverture, ou une analyse géométrique mettant en relation des couvertures entre elles. Ensuite, la hiérarchisation matricielle a été revue afin de considérer d'autres attributs de hiérarchisation spatiale en complément de la résolution spatiale. Cette analyse a permis de déterminer qu'il existe cinq combinaisons d'attributs permettant de créer des hiérarchies spatiales matricielles compatibles avec la notion de couverture matricielle. Les hiérarchies ainsi formées ont été caractérisées selon qu'il s'agit d'une hiérarchie spatiale matricielle thématique ou géométrique. Enfin, l'introduction de la notion de couverture dans les hiérarchies spatiales matricielles fait en sorte que la clé d'identification du modèle physique devenait incomplète. Une nouvelle version de cette clé d'identification, adaptée à la hiérarchisation spatiale matricielle, fut proposée. Le prochain chapitre s'intéresse particulièrement à l'analyse spatiale matricielle à la volée dans les dimensions matricielles géométriques. L'intérêt sera notamment porté sur les perspectives offertes par la hiérarchisation spatiale matricielle des couvertures et les considérations entourant ce concept aux fins d'analyse.

Chapitre 4. Optimisation des analyses spatiales matricielles à l'aide d'une dimension matricielle géométrique

4.1 Introduction

Le chapitre précédent faisait état de sources potentielles de la consommation excessive de ressources informatiques pour le traitement des analyses spatiales matricielles dans un contexte SOLAP qui sont liées à la structure de données matricielles. Dans ce chapitre, il sera question de la consommation des ressources liées aux algorithmes d'analyse spatiale matricielle, et de l'optimisation de ces algorithmes dans le cadre plus restreint des dimensions matricielles géométriques. L'optimisation d'un processus informatique est un cheminement qui vise à réduire la quantité de ressources nécessaires au traitement et à mieux utiliser les ressources disponibles afin d'accroître l'efficience. La réduction de la taille des cubes matriciels, du nombre de couvertures et de l'étendue spatiale utile à l'analyse constitue une approche privilégiée pour tendre vers l'exécution à la volée des analyses spatiales matricielles. Toutefois, cette approche limite la capacité d'analyse du cube en réduisant considérablement le nombre de faits, d'axes d'analyse et de membres dans chacune des dimensions. Pour la suite, il sera question d'explorer d'autres pistes théoriques d'optimisation, en utilisant la dimension matricielle géométrique, afin de tendre vers une exécution à la volée des analyses spatiales matricielles dans les cubes matriciels.

Le processus d'analyse spatiale selon De Smith [De Smith et *al.*, 2007] comporte quatre étapes qui sont illustrées à la figure suivante. Dans le cadre de cette recherche, il est sous-entendu que les étapes de *définition, de planification et de sélection* sont réalisées en amont par l'utilisateur du SOLAP. Cette recherche se concentre tout particulièrement sur l'étape d'observation, de mesure et d'analyse. Finalement, l'interprétation des résultats est également un élément hors portée du contexte de ces travaux.

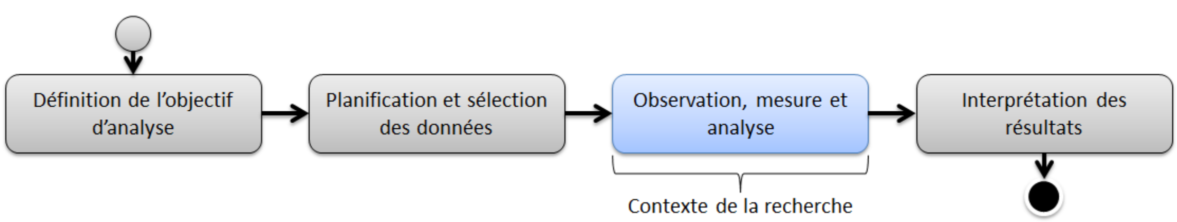


Figure 4.1 — Schématisation de l'analyse spatiale

Les structures de données OLAP sont organisées pour accroître la performance des interrogations. Cette performance provient notamment de l'agrégation des faits détaillés qui sont structurés à l'aide de dimensions contenant des hiérarchies. Au chapitre précédent, il a été démontré que les couvertures matricielles sont compatibles avec la création de dimensions spatiales matricielles sous forme de hiérarchies de couvertures matricielles. Ce chapitre vise à présenter différentes pistes théoriques d'optimisation des analyses spatiales matricielles dans les dimensions matricielles géométriques, c.-à-d. celles intégrant les couvertures par les niveaux d'une hiérarchie (figure ci-dessous).

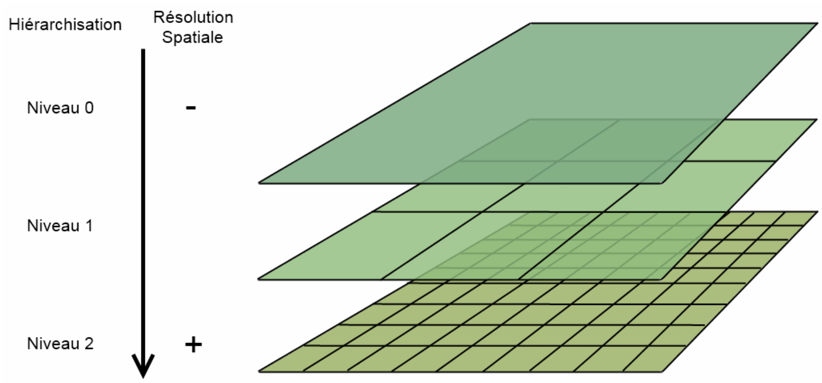


Figure 4.2 — Hiérarchisation géométrique de couvertures matricielles

Cette structure comporte un intérêt particulier, car elle supporte une hiérarchisation spatiale géométrique (variation de la résolution spatiale entre les niveaux) des couvertures matricielles formant la dimension. Cette approche réduit grandement le volume de données lorsque les cellules plus grossières, situées dans les niveaux supérieurs près de la racine, sont exploitées sur une même étendue spatiale. Dans ce chapitre, **nous explorons des façons de tirer avantage de cette hiérarchisation géométrique des couvertures matricielles pour l'optimisation de l'exécution à la volée des analyses spatiales matricielles.**

En premier lieu, il sera question de concevoir des algorithmes d'analyse spatiale matricielle spécialisés dans l'exploitation de la hiérarchie de couverture matricielle. Ensuite, une option de calcul préalable de statistiques, par opposition au calcul préalable des résultats, est abordée afin de favoriser l'analyse spatiale matricielle à la volée. Finalement, les caractéristiques d'un index de cube matriciel seront décrites ainsi que les conditions d'utilisation pour favoriser l'exécution à la volée des analyses spatiales matricielles.

4.2 Algorithmes multi-niveaux pour l'optimisation de l'analyse spatiale matricielle

Les dimensions spatiales matricielles intégrant les couvertures matricielles par les niveaux permettent la mise en œuvre de hiérarchies géométriques de couvertures matricielles. Ce type de hiérarchie (*cf.* section 3.4) permet de faire varier la résolution spatiale des couvertures matricielles qui composent les niveaux de la hiérarchie. Les différents niveaux de détails permettent d'obtenir de l'information plus généralisée près de la racine et plus détaillée vers les feuilles de l'arborescence. La figure précédente présente un exemple de ce type de hiérarchie. En appliquant la hiérarchisation aux couvertures matricielles, la structure arborescente ainsi créée ressemble aux pyramides matricielles qui sont mises en œuvre pour alléger et accélérer l'affichage cartographique et le traitement d'image [Spatial-Analyst, 2013; Ress-ArcGIS, 2013]. Nous croyons qu'il est concevable théoriquement que la forte structuration des hiérarchies de couvertures matricielles soit favorable à l'exécution à la volée des analyses spatiales matricielles.

Dans un premier temps, il sera question de la représentation mathématique de l'exploitation des couvertures matricielles et des modèles d'optimisation. Par la suite, deux méthodes d'optimisation des analyses spatiales matricielles seront explorées afin de démontrer leur contribution théorique à l'exécution à la volée des analyses spatiales matricielles. Finalement, une discussion portera sur l'application de ces méthodes d'analyse spatiales matricielles dans les cubes matriciels.

4.2.1 Représentation mathématique de l'exploitation de couvertures matricielles

L'analyse spatiale matricielle consiste à lire et interpréter les valeurs thématiques contenues dans les cellules du domaine spatial analysé. Ces valeurs thématiques sont mises en relation les unes avec les autres en fonction des opérateurs d'analyse spatiale matricielle (*cf.* 2.3.1 Opérateurs d'analyse spatiale matricielle) exploités par l'algorithme d'analyse. Les opérateurs d'analyse spatiale matricielle sont : thématiques; métriques; et topologiques. Également, il est possible de combiner et composer ces opérateurs simples pour en faire des opérateurs plus complexes. Les valeurs thématiques des cellules sont obtenues à l'aide de fonctions élémentaires de balayage des couvertures (*cf.* 2.3.2 Portée des fonctions d'analyse spatiale matricielle). Ces fonctions sont : locales; focales; ou zonales. Ces deux dernières constituent des combinaisons de fonctions locales organisées afin de faire ressortir des patrons « *patterns* » ou simplement analyser un sous-échantillonnage de données.

Les analyses spatiales matricielles reposent sur l'exécution d'algorithmes linéaires et séquentiels [Jensen, 2005; De Smith et *al.,* 2007; Tomlin, 1990]. Ce type d'algorithme possède une complexité mathématique [Wiki-BigO, 2013; Perlmonks, 2013] exprimée par la forme générale $\boldsymbol{\theta}$ **(*N*)**. Cette notation permet de décrire le comportement d'une fonction linéaire et ultimement le temps nécessaire à son exécution. La valeur de $\boldsymbol{N}$ correspond au nombre de cellules composant la couverture matricielle analysée de façon séquentielle. Sous cette forme, le temps d'exécution du traitement, cellule par cellule, est directement proportionnel au nombre de cellules du domaine spatial analysé. Dans le cas où plusieurs couvertures matricielles sont impliquées dans la résolution de l'analyse spatiale matricielle (ex. : superposition de thématiques), le temps d'exécution est de la forme générale $\theta\ (N_1+N_2+\ldots+N_x)$ où x exprime le nombre de couvertures traitées dans l'analyse spatiale matricielle. Ce cas décrit notamment les algorithmes additifs de « Map Algebra » [Tomlin, 1990].

Les algorithmes d'analyse spatiale matricielle reposant sur des fonctions focales, comme les algorithmes de lissage (ex. : « *high pass* » et « *low pass* ») ou de rehaussement (ex. : filtre Laplacien) [Jensen, 2005], sont également des algorithmes linéaires. Les cellules matricielles sont lues de façon séquentielle à l'intérieur de la fenêtre. Cette fenêtre se déplace sur toute la zone correspondante au domaine spatial de l'analyse de façon séquentielle et la lecture de chaque cellule composant la fenêtre est répétée à chaque déplacement[11]. La figure suivante illustre le déplacement de la fenêtre de convolution de l'algorithme « *texture unit* » ainsi que la séquence de lecture des cellules [Jensen, 2005]. Dans cet exemple, les cellules sont lues dans un sens horaire, la valeur de la cellule zéro est calculée et la fenêtre est ensuite déplacée vers les cellules voisines.

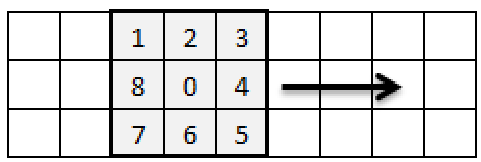


Figure 4.3 — Déplacement d'une fenêtre sur une couverture matricielle algorithme « texture unit »[12]

Les fonctions focales constituent des compositions de fonctions linéaires de la forme générale $\theta\ (N*M)$ où N est associé au nombre de cellules composant le domaine spatial et M le nombre de cellules composant la fenêtre de convolution. Enfin, une analyse spatiale matricielle focale sur plusieurs couches aurait la forme générale $\theta\ ((N_1+N_2+\ldots+N_x)*M)$.

Pour leur part, les fonctions zonales constituent des échantillons de couverture globale et sont parcourues par des fonctions locales ou focales. Ces fonctions possèdent une forme générale identique aux fonctions qui les parcourent, mais sur une étendue restreinte exprimée par αN. Les formes générales de résolution des analyses spatiales sont résumées dans le tableau suivant.

[11] Cette méthode suggère une implantation de base ou « naïve » de ce type d'algorithme et a pour but d'établir un référentiel commun à l'ensemble des fonctions focales. Une version optimisée serait dépendante de facteurs externes (ex. : système d'exploitation, langage de programmation, taille de la mémoire), mais ultimement le temps nécessaire à sa résolution serait tout de même exprimé par une fonction linéaire.

[12] [Jensen, 2005]

Tableau 4.1 — Forme générale des fonctions d'analyse spatiale matricielle

Fonction	Forme générale
Locale	
• une seule couche	**θ (*N*)**
• plusieurs couches (ou couvertures)	**θ (N_1+N_2+… +N_x)**
Focale	
• une seule couche	**θ (*N*M*)**
• plusieurs couches (ou couvertures)	**θ ((N_1+N_2+… +N_x)**M*)**
Zonale [locale]	
• une seule couche	**θ (*αN*)**
• plusieurs couches (ou couvertures)	**θ *α*(N_1+N_2+… +N_x)**
Zonale [focale]	
• une seule couche	**θ (*αN*M*)**
• plusieurs couches (ou couvertures)	**θ (*α*(N_1+N_2+… +N_x)**M*)**

Toutes ces formes exprimant la complexité des analyses spatiales matricielles sont en fait des algorithmes linéaires présentant une pente plus ou moins prononcée. Pour réduire significativement le temps d'exécution nécessaire au traitement des analyses spatiales matricielles, il est essentiel de changer la forme générale de la fonction.

L'approche proposée par Marchand [Marchand, 2004] consiste à calculer préalablement les résultats d'analyse spatiale et conserver ces résultats dans le cube. Ainsi, les résultats peuvent être indexés et accédés rapidement. La forme générale **θ *(log N)*** est très commune pour décrire le temps d'interrogation des méthodes d'indexation appartenant aux familles « *B-Tree* », « *R-Tree* », « *Quadtree* » et pour les arbres binaires. L'exception à cette règle réside dans la famille des index basés sur les tables de hachage « *hash table* » ou « *bitmap* ». Ces index permettent d'accéder aux résultats de requêtes en un temps exprimé par l'intervalle **[θ (1), θ (*N*)]**. Des algorithmes hybrides comme le « *B^x-Tree* » ont été mis au point pour combiner les forces du « *B-Tree* » et des tables de hachages afin d'obtenir des temps de requêtes encore plus faibles. Toutefois, cette approche de calcul préalable va à l'encontre de l'objectif de cette recherche qui vise à explorer les facteurs favorisant l'exécution à la volée des algorithmes d'analyse spatiale matricielle dans les cubes matriciels. **La section qui suit présente l'utilisation d'algorithmes multi-niveaux exploitant la hiérarchie de la dimension matricielle géométrique à la façon des index arborescents sans toutefois calculer préalablement les résultats d'analyse spatiale matricielle**.

4.2.2 Exploiter la hiérarchie de la dimension matricielle géométrique

Les analyses spatiales matricielles sont conçues pour exploiter des couvertures matricielles. Elles ne sont pas conçues pour exploiter les hiérarchies de couvertures matricielles présentant différents niveaux de granularité. Les algorithmes exploitant des données matricielles à différents niveaux de granularité sont regroupés sous l'appellation d'algorithmes multi-échelles ou « *multiscale algorithms* ». Ces algorithmes sont utilisés notamment en traitement d'image, en classification multivariée et en détection de points chauds ou « *clusters* » en datamining spatial [Jensen, 2005; De Smith et *al.,* 2007]. Ces algorithmes exploitent une arborescence thématique et spatiale calculée préalablement. Ils interrogent les niveaux agrégés puis les niveaux plus détaillés itérativement en fonction de la pertinence et de l'importance relative du phénomène observé. Pour traiter l'information, les algorithmes multi-échelles nécessitent une structuration de l'information selon différents niveaux de granularité qui s'apparentent fortement à la hiérarchisation des couvertures matricielles (*cf.* section 3.4). À titre d'exemple, les algorithmes multi-échelles, en traitement d'image, exploitent des pyramides de type « *MIP map* » pour accélérer le traitement ou réduire les effets de pixellisation [Wiki-Mipmap, 2013]. La figure suivante présente un exemple de « *MIP map* » (image originale à gauche et « *MIP map* » à droite).

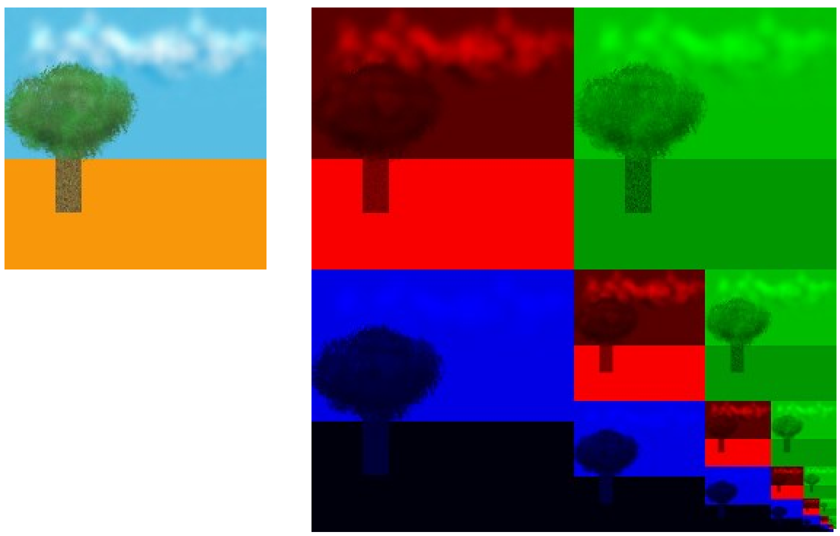
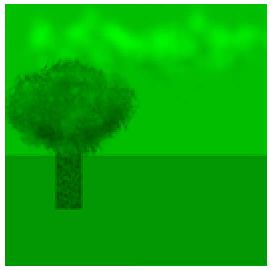
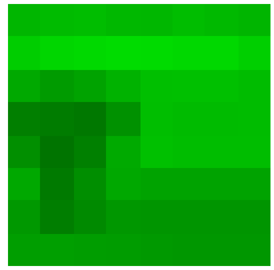


Figure 4.4 – Image RGB décomposée par couleur et échantillonnée à plusieurs résolutions pour former un « MIP map »

Les algorithmes multi-échelles exploitent ces différentes résolutions de l'image pour réduire l'étendue de la recherche à la zone d'intérêt puis en procédant de façon itérative et incrémentielle sur cet échantillon de cellules afin de produire une sélection. À titre d'exemple, la sélection de l'objet arbre sur l'image avec un algorithme traditionnel, nécessite de balayer l'image entière et de comparer la valeur numérique de chaque cellule avec un intervalle de valeur définie à partir d'un seuil de tolérance. En se basant sur l'extrait de l'image précédente et d'une seule bande (verte), il est nécessaire de parcourir 17 131 cellules (131 x 131) pour délimiter l'objet (figure a). En utilisant un algorithme multi-échelles, il est possible d'identifier une zone d'intérêt restreinte en parcourant 64 cellules (8 x 8) pour identifier grossièrement l'objet (figure b) et dans un deuxième temps parcourir cette zone d'intérêt représentant environ 40 % de l'image afin d'obtenir la sélection finale.





A) Extrait de l'image haute résolution (131 x 131 cellules)

B) Extrait de l'image faible résolution (8 x 8 cellules)

Figure 4.5— Exemple de traitement multi-échelles

La méthode de traitement, exploitant les algorithmes multi-échelles, permet de transformer l'exécution de la sélection d'une forme générale **θ (*N*)** en une exprimée par **θ *(log N)***. La forme logarithme décrit le temps d'exécution pour la recherche dans des arbres binaires ou dans toutes formes d'arborescences balancée [Perlmonks, 2013]. Ces algorithmes ne nécessitent pas de calculer préalablement les résultats, mais plutôt de structurer hiérarchiquement l'information thématique et spatiale pour la rendre plus facilement accessible. Appliquée au SOLAP, les algorithmes d'analyse spatiale matricielle multi-niveaux exploiteraient les faits matriciels agrégés afin d'identifier les faits matriciels détaillés pertinents (échantillonnage) et produire le résultat d'analyse à la volée à partir de ce sous-ensemble de cellules. Nous considérons que l'exploitation multi-échelles constitue une avenue théorique intéressante pour accélérer le traitement de certaines analyses spatiales matricielles à l'aide d'une dimension matricielle géométrique.

> **Algorithme d'analyse spatiale matricielle multi-niveaux** : Procédure d'analyse spatiale exploitant plusieurs niveaux de granularité des dimensions matricielles géométriques dans le traitement des analyses spatiales matricielles. Les niveaux de granularité sont utilisés de façon itérative et incrémentielle, du plus grossier vers le plus détaillé, afin de réduire progressivement l'étendue spatiale significative pour le traitement.

Ce type d'algorithme permet, par exemple, d'analyser l'adjacence de deux espèces de conifères afin d'évaluer le risque de propagation de la tordeuse des bourgeons de l'épinette (voir figure ci-dessous). Considérant que la tordeuse préfère l'épinette blanche à l'épinette noire [MRN-Tordeuse, 2013], le risque de contamination des peuplements d'épinettes noires est plus élevé à proximité des peuplements d'épinettes blanches.

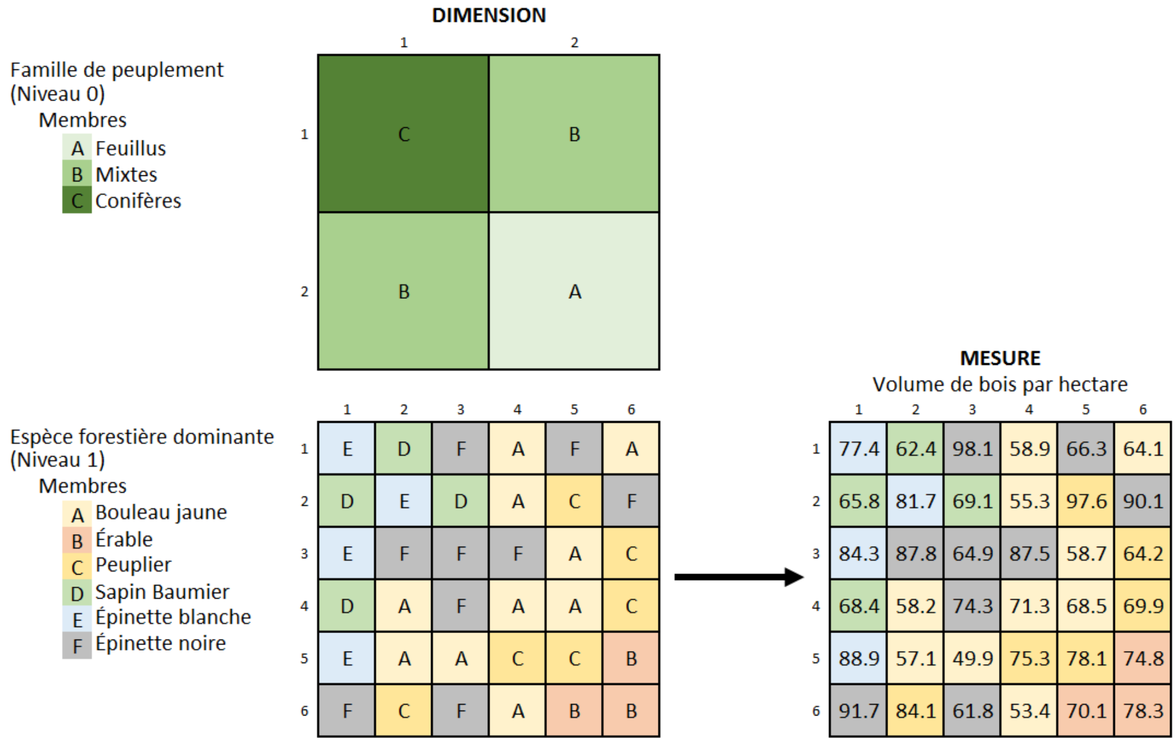


Figure 4.6— Dimension matricielle géométrique appliquée à la foresterie et mesure associée

Dans un premier temps, l'algorithme d'analyse spatiale matricielle multi-niveaux, à l'aide d'un opérateur thématique, recherche les cellules où il y a une présence de conifères dans le niveau le plus agrégé (Famille de peuplement – Mixte ou Conifère). Ensuite, par un processus de forage dans la dimension, les cellules pertinentes du niveau détaillé sont sélectionnées afin d'obtenir une étendue spatiale réduite qui est pertinente pour l'analyse spatiale matricielle. Finalement, un opérateur thématique et topologique permet de sélectionner les peuplements d'épinettes noires adjacents aux peuplements d'épinettes blanches pour calculer le volume de bois par hectare qui est le plus à risque. La figure ci-dessous résume le processus d'analyse spatiale matricielle multi-niveau.

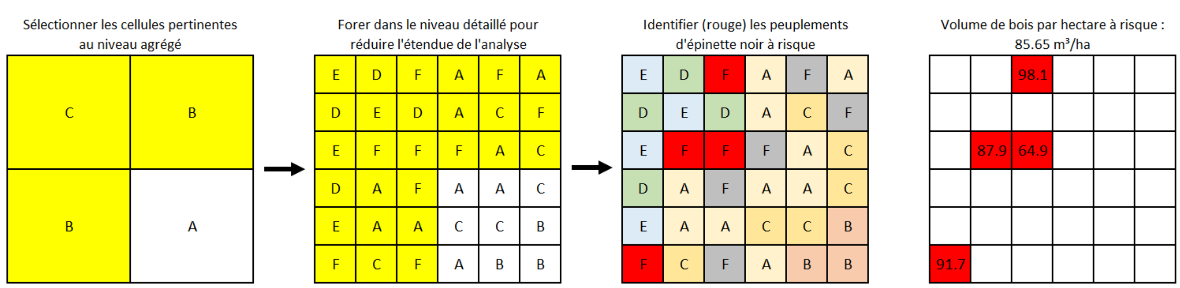


Figure 4.7 — Exemple de processus d'analyse spatiale matricielle multi-niveaux

En comparaison, pour obtenir un résultat semblable avec un algorithme d'analyse spatiale matricielle traditionnel, il aurait été nécessaire de parcourir toutes les cellules du niveau détaillé. Ce résultat est obtenu sans perte d'information en exploitant la structure de la dimension matricielle géométrique. Cet exemple théorique démontre un gain substantiel au niveau des ressources nécessaires au traitement de l'analyse spatiale matricielle par la réduction du nombre de cellules à interroger.

Les algorithmes d'analyse spatiale matricielle multi-niveaux constitue une solution théorique. La revue de littérature n'a pas permis d'identifier des algorithmes d'analyse spatiale matricielle multi-niveaux pour les dimensions matricielles géométriques dans les cubes matriciels. L'élaboration de ce type d'algorithme et la démonstration empirique de leur efficacité devront faire l'objet de travaux futurs.

4.2.3 Optimiser la dimension géométrique matricielle pour favoriser les fonctions focales

Les fonctions focales visent à calculer une nouvelle valeur thématique pour une cellule à partir des valeurs thématiques des cellules voisines. Le nombre de cellules voisines, l'ordre dans lequel les valeurs sont interprétées et le poids accordé aux valeurs dépend de l'algorithme d'analyse spatiale matricielle employé (*cf.* 2.3.3 Balayage des couvertures matricielles). Pour certains algorithmes d'analyse spatiale matricielle, la distance séparant la cellule d'intérêt et ses cellules voisines influence le poids relatif qui sera accordé à la valeur de cette cellule. Dans le contexte de cette recherche, seuls les algorithmes d'analyse spatiale matricielle exploitant des fonctions focales utilisant une fenêtre de forme carrée comportant 3 x 3, 5 x 5 et 7 x 7 cellules ont été considérés.

Ce type d'algorithme permet notamment d'identifier et de mesurer automatiquement (résolution spatiale uniforme sur l'ensemble de la couverture) les frontières des peuplements forestiers à l'égard d'autre type d'utilisation du sol (ex. : agriculture, urbain). Ce résultat est obtenu en rehaussant la couverture matricielle à l'aide d'un filtre de Sobel [Jensen, 2005] appliqué à une fenêtre de 3 x 3 puis en dénombrant les cellules résultantes dans leur contexte spatial (résolution spatiale).

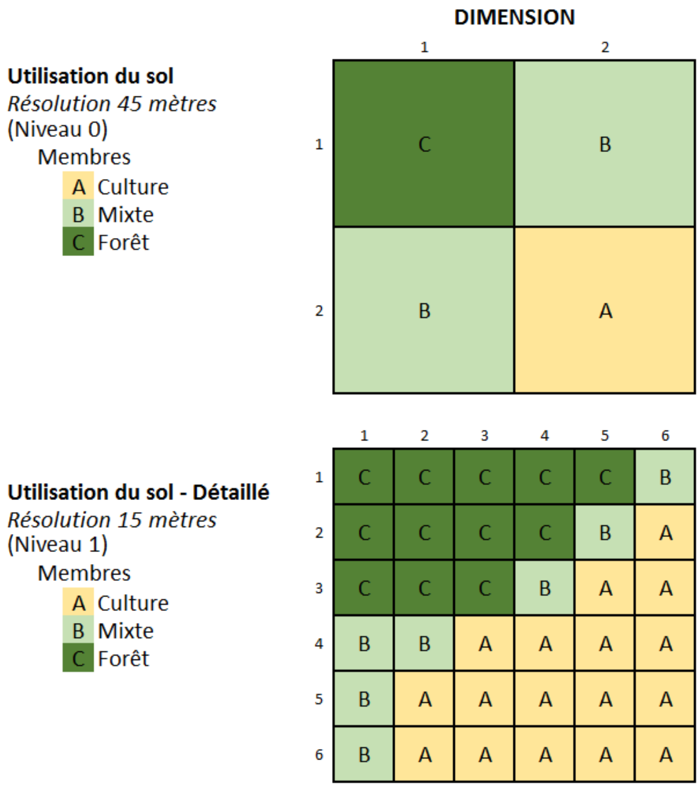


Figure 4.8 — Dimension matricielle géométrique appliquée au domaine de la foresterie (Utilisation du sol)

Ce type de filtre permet d'identifier les changements thématiques et spatiaux dans la couverture matricielle. Ainsi, les cellules où les valeurs passent de la culture (A) à la mixité (B) puis à la forêt (C) ou directement de la culture (A) à la forêt (C), tout en considérant le voisinage immédiat des cellules, seront rehaussées.

Dans un premier temps, l'algorithme d'analyse spatiale matricielle multi-niveaux recherche les cellules où il y a présence de milieu forestier ou mixte dans le niveau le plus agrégé (Utilisation du sol). Ensuite, un forage spatial sur les membres sélectionnés dans la dimension permet de sélectionner les cellules pertinentes du niveau détaillé. Cette façon de procéder permet d'obtenir une étendue spatiale réduite en comparaison à l'ensemble du domaine spatial. Finalement, un opérateur thématique et topologique exploitant une fenêtre focale (filtre de Sobel) permet de rehausser l'échantillon de la couverture matricielle afin d'obtenir la frontière entre le milieu forestier et la culture. Les cellules conservées sont dénombrées et reportées dans leur contexte spatial (résolution) pour obtenir la longueur de cette frontière. La figure ci-dessous résume le processus d'analyse spatiale matricielle multi-niveaux appliqué à une fonction focale.

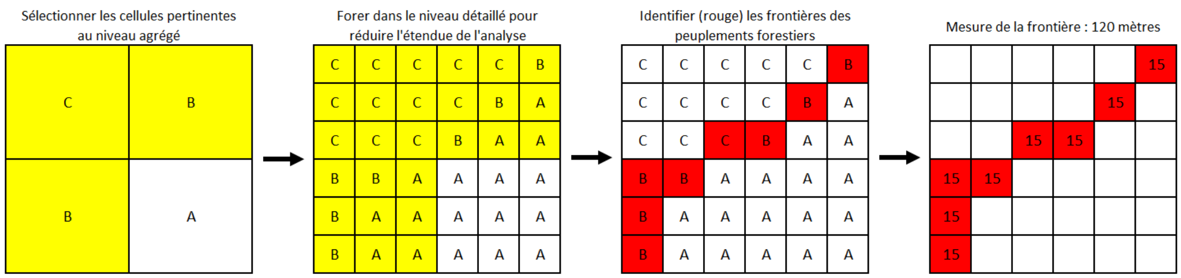


Figure 4.9 — Exemple de processus d'analyse spatiale matricielle multi-niveaux (fonction focale)

L'algorithme d'analyse spatiale matricielle multi-niveaux décrit précédemment exploite l'arborescence de la hiérarchie afin de tendre vers un temps d'exécution se rapprochant de la forme **θ *(log N)***. Le gain de performance est obtenu réduisant la taille de l'étendue spatiale de l'analyse (***N***) à sa plus simple expression, c.-à-d. aux cellules pertinentes (***N'***). Toutefois, afin d'obtenir un maximum de gain de performance pour le temps d'exécution, il est nécessaire d'optimiser la hiérarchie. En effet, le recours à des l'algorithme d'analyse matricielle multi-niveaux utilisant une fonction focale dans une structure qui n'est pas optimisée peut entrainer une consommation indue de ressources informatiques.

Le chapitre précédent (*cf.* 3.4.1 Hiérarchisation des couvertures matricielles) décrit la variation de la résolution spatiale (**δ**) entre deux niveaux adjacents (ρ : ρ-1; ρ+1) d'une hiérarchie de dimensions spatiales matricielle. Cette valeur **δ** exprime aussi la relation de subordination qui existe entre une cellule d'un niveau agrégé (parent) et les cellules du niveau détaillé (enfant). Une dimension spatiale matricielle géométrique se caractérise par **δ** ϵ [1, N] où N est constant pour les niveaux adjacents. Une dimension spatiale matricielle non-géométrique (ou thématique) se caractérise par **δ** ϵ [1]. La figure suivante illustre deux exemples de valeur **δ** exprimant la relation parent-enfant entre les cellules dans une hiérarchie de couvertures matricielles.

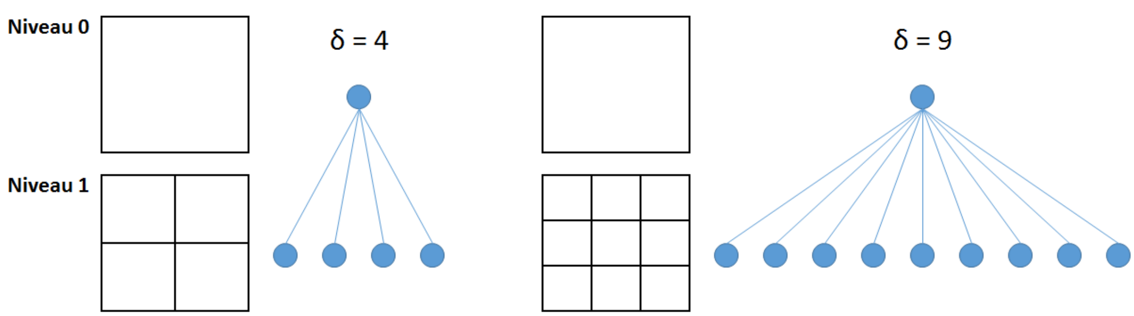


Figure 4.10 — Variation de la résolution spatiale entre deux niveaux d'une hiérarchie de couvertures matricielles

D'entrée de jeu, ce ne sont pas toutes les dimensions matricielles géométriques qui sont favorables à l'exécution à la volée des fonctions focales. Pour obtenir une valeur **δ** favorable à l'exécution à la volée des analyses spatiales matricielles focales, il est nécessaire d'**optimiser la dimension matricielle géométrique**. Une situation favorable consiste en une cellule du niveau agrégé permettant d'obtenir une ou plusieurs fenêtres focales de dimension appropriée lorsqu'elle est forée. Par opposition, une situation défavorable survient lorsque plusieurs cellules d'un niveau agrégé sont nécessaires pour obtenir une seule fenêtre focale au niveau détaillé. La figure ci-dessous illustre trois dimensions matricielles géométriques. L'une d'entre elles (figure a) n'est pas favorable à l'exécution à la volée, alors que les deux autres (figures b et c) sont favorables à l'exécution à la volée des analyses spatiales matricielles multi-niveaux. À chaque itération, la structure qui n'est pas favorable entraine la lecture de cellules agrégées en plus grand nombre. À l'opposé, la structure optimisée ne produit pas cette consommation excessive de ressources informatiques pour le traitement.

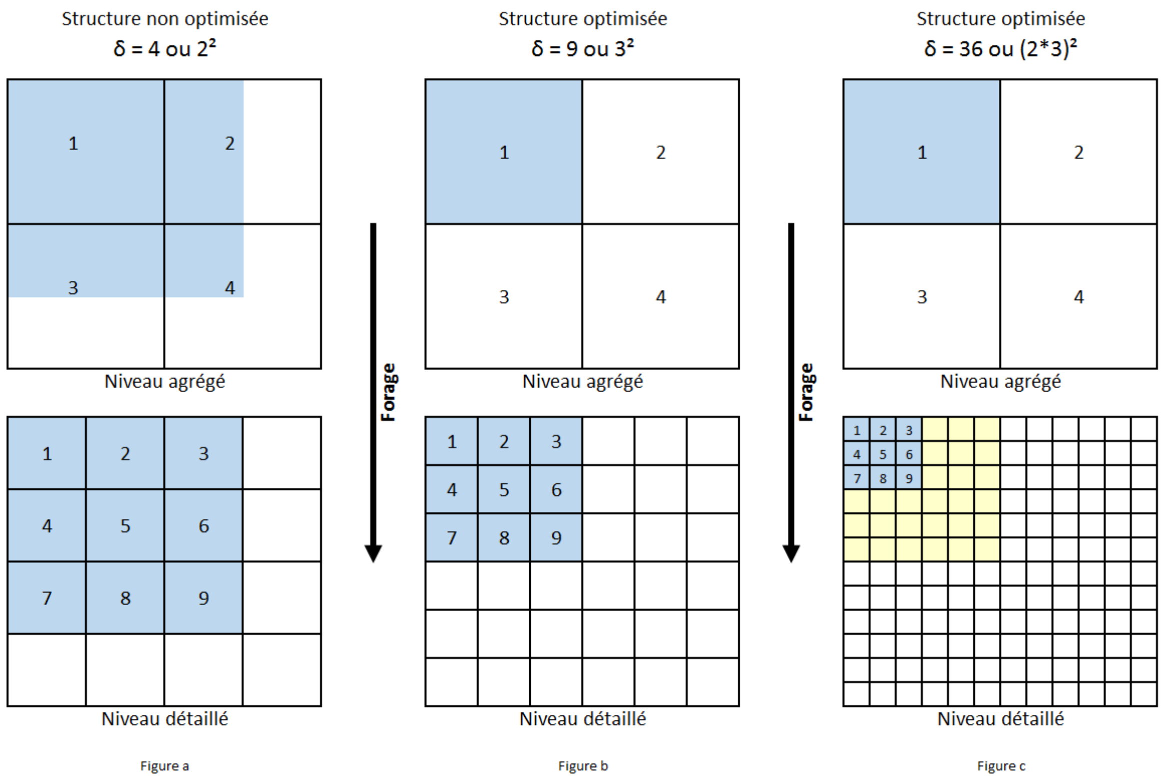


Figure 4.11 — Optimisation de la hiérarchie géométrique de couverture matricielle pour supporter l'analyse spatiale matricielle à la volée à l'aide d'une fonction focale (3x3)

Ainsi, pour produire des analyses avec des fenêtres de convolution de 3 x 3 la valeur **δ** doit produire une racine carrée entière. Aussi, le résultat de cette racine carrée doit être un multiple de 3 et d'un coefficient entier. La formule suivant décrit les valeurs possibles de **δ** entre deux niveaux adjacents afin d'être optimisée pour le traitement des fonctions focales de 3 x 3.

$$\boldsymbol{\delta} = (\alpha 3)^{\beta}$$

Où $\alpha \in \mathbb{N}$ et $\beta = \{2n, n \in \mathbb{N}\} = 2\mathbb{N}$

Ce même raisonnement est applicable aux autres tailles de fenêtres de fonction focales. Le tableau ci-dessous présente les formules d'optimisation pour obtenir la valeur **δ** pour différentes dimensions de fenêtres de convolution.

Tableau 4.2 — Optimisation de la relation parent-enfant des cellules de deux niveaux contigus

Taille de la fenêtre	**Variation** δ
3 x 3	$(\alpha 3)^{\beta}$ où $\alpha \in \mathbb{N}$; $\beta = \{2n, n \in \mathbb{N}\} = 2\,\mathbb{N}$
5 x 5	$(\alpha 5)^{\beta}$ où $\alpha = \mathbb{N}$; $\beta = \{2n, n \in \mathbb{N}\} = 2\,\mathbb{N}$
7 x 7	$(\alpha 7)^{\beta}$ où $\alpha = \mathbb{N}$; $\beta = \{2n, n \in \mathbb{N}\} = 2\,\mathbb{N}$

Finalement, il faut rappeler que les algorithmes multi-niveaux d'analyse spatiale matricielle constituent une solution théorique. L'optimisation de la dimension matricielle géométrique, pour favoriser l'analyse spatiale matricielle à la volée, s'appuie sur la proposition d'analyse spatiale matricielle multi-niveaux. Nous considérons que l'exploitation multi-échelles constitue une avenue intéressante pour accélérer le traitement de certaines analyses spatiales matricielles à l'aide d'une dimension matricielle géométrique. La revue de littérature n'a pas permis d'identifier des algorithmes d'analyse spatiale matricielle multi-niveaux pour les dimensions matricielles géométriques dans les cubes matriciels. Toutefois, l'élaboration de ce type d'algorithme et la démonstration empirique de leur efficacité devront faire l'objet de travaux futurs.

4.2.4 Constats et recommandations à propos des algorithmes multi-niveaux et pour l'optimisation de l'analyse spatiale matricielle

Le traitement des analyses spatiales matricielles possède un temps de résolution linéaire exprimé par la forme générale **θ (*N*)**. Pour réduire le temps de traitement, nous avons opté pour la réduction de l'étendue spatiale pertinente à l'analyse dans la couverture matricielle. Le but était de proposer une façon de diminuer la taille de la variable ***N*** sans toutefois affecter la validité des résultats des analyses. La piste de solution envisagée consiste à exploiter la structure du cube, mais plus particulièrement la dimension spatiale matricielle géométrique. Cette structure offre de bonnes performances d'interrogation exprimées par la forme générale **θ *(log N)*** ou **[θ (1), θ (*N*)]** selon le type d'indexation employé. Pour ce faire, il est nécessaire d'**élaborer des algorithmes d'analyse spatiale matricielle spécialisée** capable d'exploiter les niveaux géométriques des hiérarchies de couvertures matricielles. Ces algorithmes d'analyse spatiale matricielle multi-niveaux utilisent l'arborescence de la dimension matricielle géométrique pour sélectionner les données pertinentes à l'analyse et ainsi réduire la taille de l'étendue spatiale de l'analyse. Ainsi les fonctions d'analyse locales, focales et zonales consomment le minimum de ressource nécessaire pour leur traitement. Les analyses spatiales matricielles multi-niveaux constituent une proposition, qui est favorable à l'exécution à la volée des analyses spatiales matricielles dans un contexte SOLAP.

Les fonctions d'analyse spatiale matricielle focales ont ceci de particulier qu'elles nécessitent plusieurs itérations sur les mêmes cellules afin de produire le résultat d'analyse spatiale matricielle. C'est pourquoi leur temps d'exécution est exprimé par la forme générale **θ (*N*M*)**. La valeur ***M*** correspond au nombre de cellules composant la fenêtre focale et c'est cette fenêtre qui doit parcourir l'ensemble des cellules pertinentes de la couverture matricielle. La valeur ***M*** étant constante en fonction de l'algorithme d'analyse, la piste de solution envisagée pour favoriser l'exécution à la volée de ce type d'analyse demeure la réduction de la variable ***N*** (étendue pertinente à l'analyse). Pour obtenir une dimension matricielle géométrique favorable à l'analyse spatiale matricielle à la volée, il est nécessaire d'**<u>optimiser la dimension matricielle géométrique</u>** pour supporter adéquatement les fonctions focales. L'optimisation consiste à s'assurer que la relation entre une cellule d'un niveau agrégé et les cellules d'un niveau plus détaillé exprimées par **δ** permet d'obtenir toutes les cellules pertinentes pour alimenter une fenêtre focale lorsqu'une cellule agrégée est forée. S'il s'avère nécessaire de sélectionner plusieurs cellules d'un niveau agrégé pour peupler une fenêtre focale avec les cellules correspondantes d'un niveau détaillé, il y aura des délais indus dans le traitement de l'analyse spatiale matricielle. Ici, la notion de cube matriciel spécialisé pour faciliter l'analyse spatiale matricielle à la volée prend tout son sens. Les dimensions matricielles géométriques doivent être conçues pour supporter les fonctions focales afin d'en assurer l'efficience par la suite. Il y a donc un **effort de planification additionnel à faire lors de la conception des dimensions matricielles géométriques afin de favoriser l'exécution à la volée des analyses spatiales matricielles**.

4.3 Calcul préalable de statistiques pour supporter les analyses spatiales matricielles multi-niveaux

Les travaux précédents pour la production d'analyses spatiales dans les cubes SOLAP ont démontré qu'il était nécessaire de calculer préalablement les analyses et d'interroger les résultats [Marchand, 2004, Brisebois, 2003] pour obtenir des temps de réponse adéquats (temps < 10 secs). Or, l'analyse spatiale matricielle à la volée constitue le point focal des travaux de cette recherche et elle est incompatible avec le calcul préalable des résultats. Il n'est toutefois pas exclu de calculer préalablement certaines statistiques, découlant des faits matriciels, qui puissent accélérer le traitement des analyses spatiales matricielles multi-niveaux. Nous croyons que le calcul préalable de ces statistiques, reposant sur la structure de la dimension matricielle géométrique, permet théoriquement de rendre plus performant certains algorithmes d'analyse spatiale matricielle multi-niveaux.

Dans un premier temps, la notion de statistiques spécialisées dans l'optimisation des performances des analyses spatiales matricielles multi-niveaux sera abordée. Par la suite, un exemple d'application. Finalement, des constats et recommandations, à l'égard des statistiques spécialisées pour optimiser le traitement des analyses spatiales matricielles multi-niveaux, viendront clore cette section.

4.3.1 Caractérisation des statistiques conçues pour l'optimisation des algorithmes d'analyse spatiale matricielle multi-niveaux

Le calcul préalable est un concept connu et maîtrisé en OLAP. Il permet notamment calculer les agrégations, à différents niveaux de granularité, des faits détaillés à l'aide des hiérarchies des dimensions. Les faits agrégés synthétisent les faits détaillés ce qui contribue à la performance des interrogations dans les structures OLAP. De la même façon, les statistiques spécialisées pour l'optimisation des analyses spatiales matricielles multi-niveaux permettent théoriquement d'accélérer le traitement de certains algorithmes. Ces statistiques sont le résultat de fonctions agrégatives (distribution, algébrique, holistique) [Gray et al., 1996; Shekhar et al., 2001] de nature statistique ou géostatistiques. Ces agrégations sont calculées selon les différents niveaux de granularité des dimensions matricielle géométrique. Elles synthétisent l'information thématique retrouvée dans les mesures du cube. Il s'agit de fonctions zonales, d'un point de vue de l'analyse spatiale matricielle, qui extraient les informations pertinentes selon tous les croisements des axes d'analyse. Le domaine de valeurs, les valeurs minimales et maximales, la somme, la moyenne et l'écart-type sont des exemples de statistiques qui peuvent théoriquement contribuer à l'exécution à la volée des analyses spatiales matricielles.

Considérant qu'il n'existe pas d'algorithmes d'analyse spatiale matricielle multi-échelles conçues pour les structures de données OLAP spatiales à ce jour, il est difficile de prédire quelles seront les statistiques pertinentes ou spécifiques à certains besoins d'analyse. Dans le cadre de cette recherche, nous nous sommes concentrés sur les statistiques décrivant les valeurs thématiques minimales et maximales (bornes). Ces valeurs sont obtenues en parcourant les valeurs thématiques des cellules détaillées (enfant) de chaque cellule agrégée (parent), et ce pour tous les croisements des axes d'analyse. Un tableau à la fin de cette section présente les statistiques qui comportent en théorie un intérêt pour favoriser l'analyse spatiale matricielle à la volée. La figure ci-dessous illustre la façon de calculer les statistiques spécialisées représentant les valeurs thématiques maximales et minimales. La relation de subordination, établie entre la cellule de la dimension matricielle géométrique, permet d'établir la portée de la recherche des valeurs maximales et minimales.

Ces statistiques permettent d'accélérer la recherche de valeurs thématiques clés. À titre d'exemple, ce type de statistique facilite la recherche des peuplements d'épinette noire où le rendement est supérieur à 90 m³/ha et qui sont à risque de contamination élevée par la tordeuse d'épinette dû à leur proximité avec un peuplement d'épinette blanche. Ainsi, les efforts de protection ou de récoltes pourront être concentrés dans ces peuplements à haut rendement et à risque.

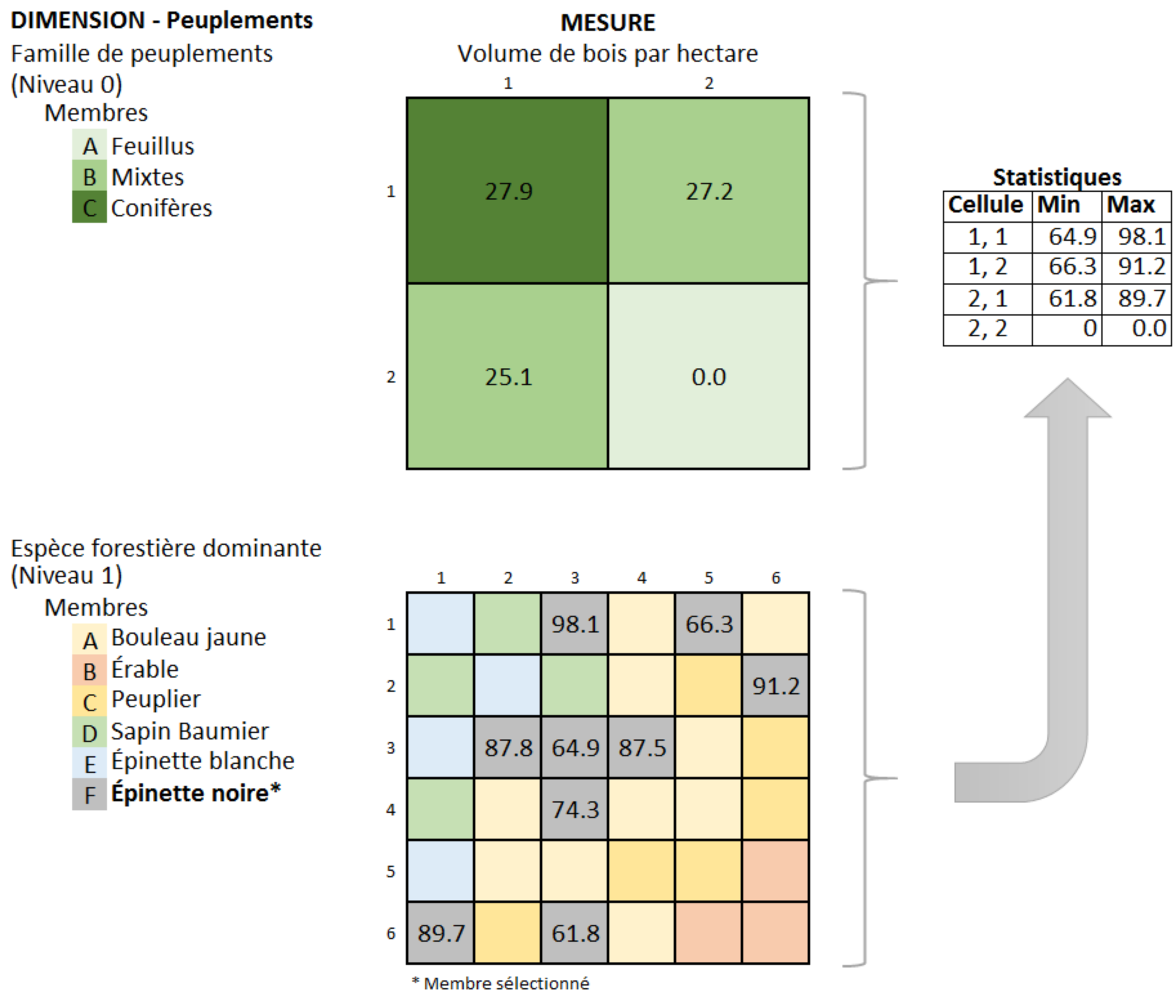


Figure 4.12 – Calcul préalable des statistiques dans une dimension matricielle géométrique et thématique

Dans un premier temps, l'algorithme d'analyse spatiale matricielle multi-niveaux recherche dans les statistiques les cellules où la valeur thématique maximale de « volume par hectare » est supérieure à 90 m³/ha. Ces cellules pertinentes sont sélectionnées au niveau agrégé sur la base des statistiques calculées préalablement et un forage spatial sur ces membres permet d'obtenir une étendue spatiale d'intérêt réduite. Finalement, un opérateur thématique et topologique sélectionne les peuplements d'épinettes noires adjacents aux peuplements d'épinettes blanches répondant au critère de volume. La figure ci-dessous résume le processus d'analyse spatiale matricielle multi-niveaux optimisé à l'aide des statistiques calculées au préalable.

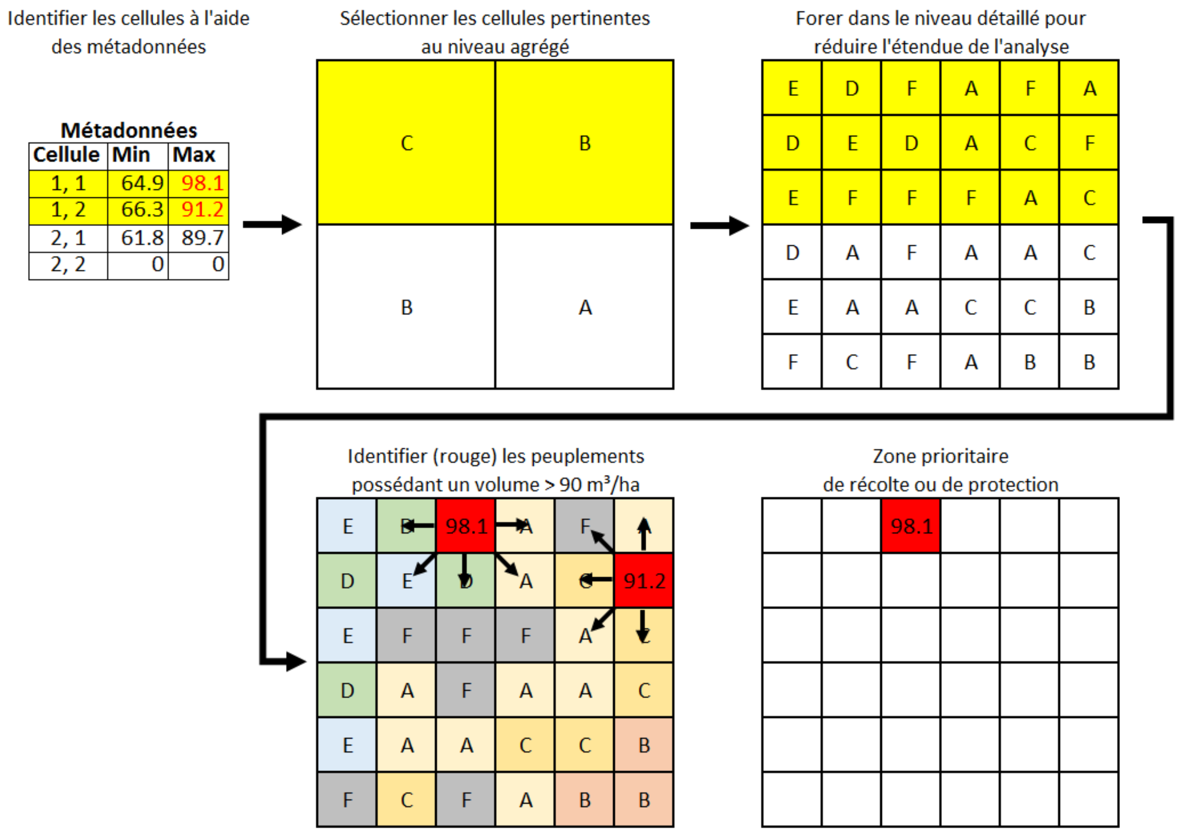


Figure 4.13 — Processus d'analyse multi-niveaux exploitant les statistiques calculées préalablement

En comparaison, ce type de résultat aurait pu être obtenu sans l'apport des statistiques, mais aurait nécessité de parcourir un plus grand nombre de cellules au niveau agrégé et au niveau détaillé. Enfin, en théorie, si ces statistiques sont triées et indexées à l'aide d'une table de hachage conventionnelle, il serait possible d'atteindre une forme générale équivalente à **θ (1)** pour la résolution de cette étape préliminaire de sélection de l'étendue spatiale d'intérêt [Perlmonks, 2013]. Cet exemple démontre un gain potentiel théorique de performance en interrogeant les statistiques spécialisées calculées préalablement pour identifier des candidats potentiels au traitement de l'analyse spatiale matricielle.

4.3.2 Constats et recommandations à l'égard des statistiques spécialisées pour les analyses spatiales matricielles multi-niveaux

Les statistiques spécialisées pour supporter l'analyse spatiale matricielle multi-niveaux représentent une avenue intéressante afin de réduire le temps de traitement nécessaire et ainsi tendre vers un traitement à la volée. Ces **<u>statistiques spécialisées doivent être planifiées lors de la modélisation et calculées lors de la mise en œuvre du cube matriciel</u>**. Le gain de performance est obtenu au détriment de l'espace de stockage nécessaire pour le cube matriciel. En effet, l'espace nécessaire pour conserver ces statistiques peut s'avérer une limitation, mais devrait être largement compensé par le gain de performance pour l'analyse spatiale matricielle. Également, il faut mentionner que ces statistiques sont calculées pour les faits agrégés du cube à partir des faits détaillés. Il y a donc autant de statistiques que de faits agrégés, ce qui représente un volume d'information plus limité que s'il était nécessaire, à titre d'exemple, de calculer des statistiques pour tous les faits du cube matriciel. Aussi, il est envisageable qu'un système de gestion de bases de données multidimensionnelles hybride (HOLAP) puisse hypothétiquement calculer à la demande et conserver ces statistiques selon l'utilisation qui en est faite. Ainsi, les analyses subséquentes, dans un contexte semblable, seraient plus performantes. Toutefois, le recours aux structures de données HOLAP demeure une perspective théorique qui n'est pas abordée plus en détail dans le cadre de ces travaux. Le tableau suivant présente les statistiques qui ont été analysées dans le cadre de ces travaux.

Tableau 4.3 — Statistiques pour l'optimisation du traitement des analyses spatiales matricielles

Statistiques	**Exemple d'application**
Bornes (Min, Max) ou intervalle de valeurs	Permet de connaître les valeurs thématiques maximales et minimales d'un sous-ensemble de cellules indépendamment de leur position individuelle. La question « Est-ce qu'il y a des cellules possédant un volume de bois par hectare supérieur à 80 m³? » permet de discriminer un sous-ensemble et saurait être répondue sans explorer individuellement les cellules du niveau détaillé à chaque interrogation.
Somme	Renseigne sur l'addition ou la quantité totale des valeurs thématiques dans un sous-ensemble de cellules indépendamment de leur position individuelle. La question « Est-ce que le volume de bois total dans ce sous-ensemble est supérieur à 300 m³? » permettrait de connaître la valeur potentielle d'une zone sans interroger chaque cellule du niveau le plus détaillé à chaque interrogation.
Moyenne arithmétique	Renseigne sur la tendance des valeurs dans un sous-ensemble de cellules indépendamment de leur position individuelle. La question « Est-ce que le volume de bois moyen locale dans ce sous-ensemble est supérieur au volume de bois moyen global? » pourrait être répondue et renseigner sur des zones possédant un potentiel d'intérêt sans interroger chaque cellule du niveau plus détaillé.

Écart-type et Variance	Renseigne sur la distribution des valeurs dans un sous-ensemble de cellules et constitue une des composantes du calcul de l'autocorrélation spatiale. La question « Est-ce que le volume de bois par hectare est homogène (écart-type faible) dans ce sous-ensemble de cellules? » pourrait être répondue sans interroger le volume des cellules du niveau plus détaillé à chaque interrogation.
Énumération des valeurs possibles	Permets de valider la présence d'une valeur thématique dans un sous-ensemble de cellules sans en connaître la position ou la fréquence. La question « Est-ce qu'il y a des cellules pour lesquelles le volume de bois par hectare est compris entre X m³ et Y m³? » pourrait être répondue aisément sans interroger chaque cellule du niveau plus détaillé une à une lors de l'analyse spatiale matricielle.

Le calcul préalable de statistiques supportant l'analyse spatiale matricielle multi-niveaux constitue une option particulièrement intéressante pour deux raisons. La première est que ces statistiques sont potentiellement utiles à un vaste éventail de processus et d'algorithmes d'analyse spatiale matricielle. Cette piste de solution constitue donc une option polyvalente pour les analyses à produire. La seconde raison est qu'il n'est pas nécessaire de prévoir toutes les analyses spatiales matricielles lors de la conception du cube et de calculer préalablement leurs résultats. Ainsi, le cube demeure flexible et ne nécessite pas d'être mise à jour à toutes les fois qu'une nouvelle analyse spatiale matricielle est nécessaire pour les utilisateurs du cube. Enfin, comme il s'agit d'une méthode reposant sur l'arborescence de la hiérarchie, le temps d'exécution de la sélection est exprimé par une forme **θ *(log N)*** plutôt que la forme traditionnelle **θ (*N*)**. Le temps de traitement de l'analyse spatiale matricielle demeure pour sa part inchangé, mais comme il repose sur un domaine spatial d'intérêt restreint (***N'***) il est théoriquement moins long à exécuter. Finalement, il faut rappeler que les algorithmes d'analyse spatiale matricielle multi-niveaux constituent une proposition de solution théorique. La revue de littérature n'a pas permis d'identifier des algorithmes d'analyse spatiale matricielle multi-niveaux pour les dimensions matricielles géométriques dans les cubes matriciels. L'efficacité empirique de ces pistes de solution devra être démontrée dans des travaux futurs.

4.4 Index pour les analyses spatiales matricielles multi-niveaux dans les cubes matriciels

Dans le contexte d'une structure de données OLAP, le but d'un index est d'organiser les faits du cube de façon à accélérer la lecture des valeurs thématiques en lien avec le contexte multidimensionnel d'interrogation (membres sélectionnés). Les index non spatiaux utilisent les valeurs des clés primaires des membres des dimensions dans la table de faits pour créer l'indexation et accélérer le traitement des requêtes. Pour leur part, les index spatiaux utilisent la relation spatiale (ex. : topologie) présente dans les niveaux de la dimension spatiale afin de s'assurer que les valeurs qui sont spatialement situées près les unes des autres sont aussi logiquement rapprochées dans la structure de données de la table de fait. Cette technique est également appelée « *spatial clustering* ». Il est évidemment possible de combiner plusieurs aspects thématiques, spatiaux et temporels pour créer des index adaptés.

Il est généralement reconnu que les requêtes spatiales engendrent une consommation importante de ressources informatiques lors de leur traitement. C'est le cas également dans le domaine multidimensionnel spatial. « Malgré les méthodes classiques de réduction de consommation de ressources (vues prétraitées, fragmentation, partitionnement et indexation), l'interrogation spatiale des hiérarchies est relativement longue à traiter » (Traduction libre) [Siquiera et *al*., 2012]. Pour remédier à ces temps de traitements importants, l'auteur de cet article propose la création d'un index de type « Bitmap hiérarchisé » qui conserve les relations spatiales entre les membres des niveaux des hiérarchies spatiales. Cet index se nomme HSB « *Hierarchical Spatial Bitmap* » et est employé dans un contexte SOLAP pour des données géospatiales vectorielles. La discussion de la présente section élabore sur le fait que l'index HSB n'est pas adapté aux hiérarchies de couvertures matricielles et à l'analyse spatiale matricielle à la volée.

Cette section vise à identifier des caractéristiques théoriques d'une indexation spatiale favorable à l'exécution à la volée des analyses spatiales matricielles multi-niveaux. En premier lieu, il sera question de la notion d'index spatial pour les dimensions matricielles géométriques. Par la suite, la façon de structurer cet index pour favoriser l'analyse spatiale matricielle multi-niveaux pour les fonctions focales sera abordé. Finalement, une discussion sur le fait que l'indexation HSB n'est pas adaptée à l'analyse spatiale matricielle à la volée et des recommandations pour la création d'un index spécialisé viendront clore cette section.

4.4.1 Index spatial pour les dimensions matricielles géométriques

Les concepts de cube matriciel conçu pour supporter l'analyse spatiale matricielle à la volée et de dimension matricielle géométrique impliquent certaines caractéristiques structurantes pour la définition de l'index spatial spécialisé. Tout d'abord, cet index doit être compatible avec les couvertures matricielles, supporter l'énumération séquentielle des cellules (cf. 2.1.2 Couverture matricielle) qui les composent et être conciliable avec les méthodes de balayage des couvertures par les algorithmes d'analyse spatiale matricielle (cf. 2.3.3 Balayage des couvertures matricielles). Ceci dit, il doit également être capable d'exploiter les hiérarchies de couvertures matricielles, c.-à-d. les relations topologiques à l'intérieur d'une couverture, mais aussi entre les couvertures qui composent la hiérarchie. Enfin, cet index spatial doit être compatible avec l'aspect multidimensionnel du cube.

Le balayage correspond à la séquence par lesquels les cellules matricielles sont parcourues afin de produire le résultat d'analyse. Ces paramètres sont importants, car plusieurs algorithmes d'analyse (ex. : Krigeage, filtre Laplacien et « *Texture Unit* ») reposent sur une séquence précise pour produire le résultat désiré. Le chapitre 2 (*cf. 2.3.3 Balayage des couvertures matricielles*) présente plus en détail les méthodes de balayage des couvertures matricielles (figure ci-dessous).

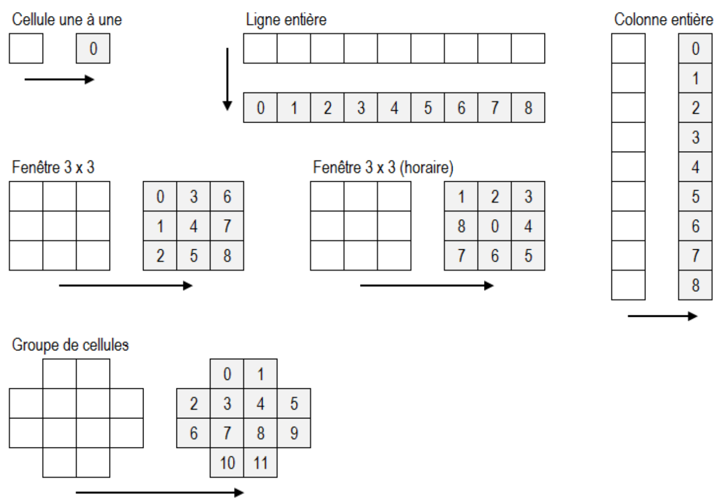


Figure 4.14 — Méthodes de balayage des données matricielles

Les courbes de remplissage ou « *space-filling curves* » furent élaborées notamment pour parcourir les données matricielles sans engendrer d'écart trop important d'un point de vue géospatial entre les cellules comme c'est le cas avec les balayages linéaires.

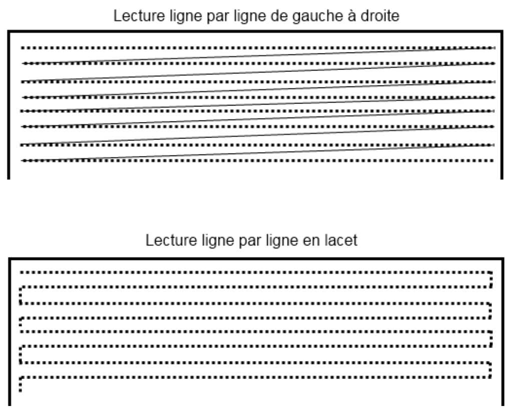


Figure 4.15 — Exemples de balayage linéaire

Il existe un grand nombre de courbes de remplissage compatibles avec les couvertures matricielles (ex. : Hilbert, Peano, z-order, Moore et Sierpinski) et d'autres non compatibles (ex. : Koch, Dragon et Gosper), car elles sont conçues pour d'autres types de tessellations qui ne sont pas carrées. La figure suivante illustre un exemple des courbes de remplissage de Hilbert du premier au troisième ordre.

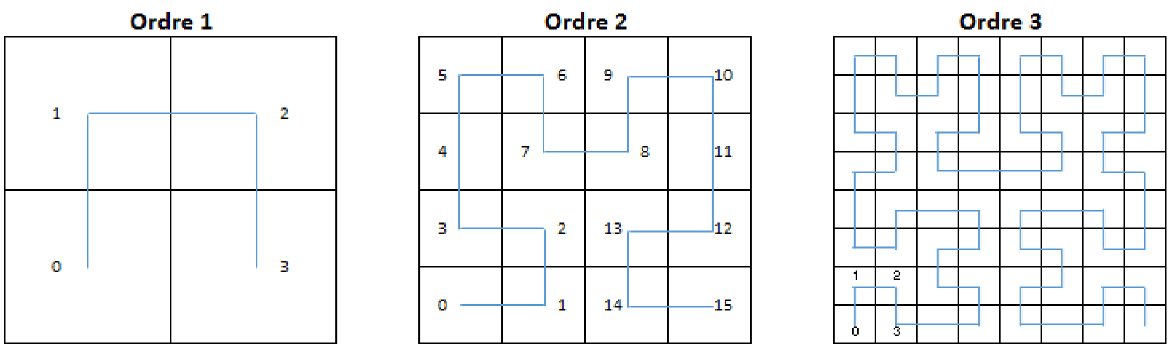


Figure 4.16 — Courbe de Hilbert (Ordre 1 à 3) appliqué à des couvertures matricielles

L'ordre des courbes de remplissage permet de suivre l'évolution des niveaux des hiérarchies de couvertures matricielles géométriques. Ainsi, il y a une relation linéaire entre l'ordre de la courbe et la résolution spatiale. La figure suivante illustre un exemple de hiérarchie géométrique des couvertures matricielles et le parcours d'une courbe de Hilbert sur ces niveaux hiérarchiques.

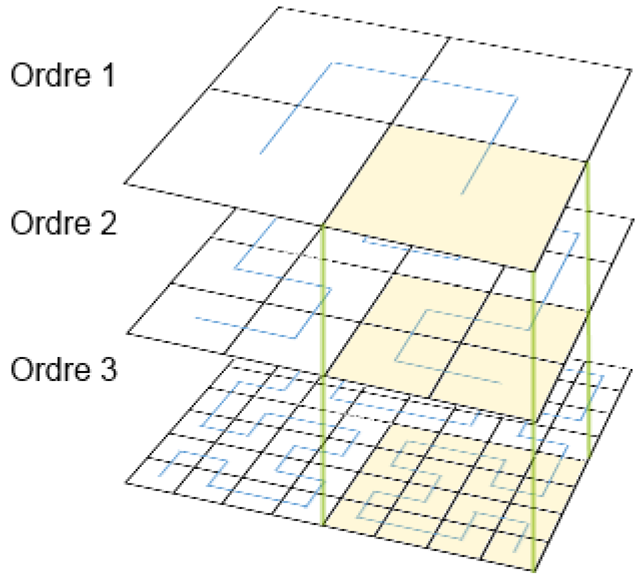


Figure 4.17 — Hiérarchisation de la courbe de Hilbert (Ordre 1 à 3) appliqué à des couvertures matricielles

Les courbes de remplissage sont utilisées pour construire des index spatiaux hiérarchiques (ex. : Hilbert pour le R-Tree). La performance d'interrogation de ces index provient du fait qu'il s'agit de fonctions supportant une énumération séquentielle au même titre que les couvertures matricielles, mais aussi qu'elle permet d'explorer les niveaux de détails des hiérarchies pour atteindre progressivement le niveau le plus détaillé. Enfin, les courbes de Peano ou de Hilbert peuvent être utilisées pour les applications multidimensionnelles et sont compatibles avec les hypercubes [Laurini Milleret-Raffort, 1993]. Pour chaque dimension additionnelle, il suffit d'ajouter une coordonnée pour situer la courbe dans ce nouvel espace et connaître l'emplacement dans l'énumération séquentielle des membres de cette dimension.

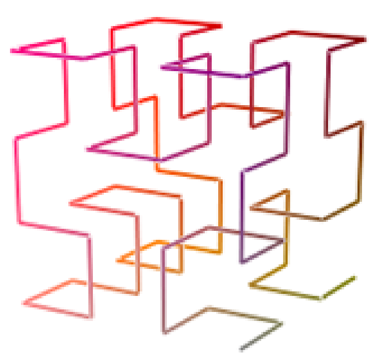


Figure 4.18 — Application 3D d'une courbe de Hilbert[13]

[13] Source : http://www.cgtools.net/data_generation_edges.php

Les courbes de remplissage constituent des candidates de premier plan pour la création d'index dans le contexte des dimensions matricielles géométriques. Notamment, elles sont compatibles avec les couvertures matricielles, supportent l'énumération séquentielle des cellules et peuvent être hiérarchisées. Toutefois, elles possèdent toutes une séquence précise d'énumération qui n'est pas nécessairement adaptée à toutes les fonctions d'analyse spatiale matricielle. En effet, les fonctions d'analyse nécessitant un balayage focal ou zonal linéaire (ex. : krigeage) possèdent des exigences spécifiques. La prochaine section s'intéresse à l'optimisation de la dimension géométrique matricielle et de la courbe de remplissage pour supporter les analyses spatiales matricielles multi-niveaux utilisant des fonctions focales.

4.4.2 Optimiser l'index pour les analyses spatiales matricielles multi-niveaux focales

Pour les analyses spatiales matricielles multi-niveaux les index créés à partir de ces courbes de remplissage possèdent un intérêt majeur pour les fonctions locales. Ce qui n'est toutefois pas toujours le cas pour les fonctions focales d'analyse spatiale matricielle. En effet, ces courbes reposent la plupart du temps sur des bases binaires (base 2) tout comme les systèmes informatiques pour lesquels elles ont été conçues. Or, les fonctions focales visées par cette recherche nécessitent des bases 3, 5 ou 7 afin de créer des fenêtres de 3 x 3, 5 x 5 ou 7 x 7. L'utilisation d'un index de type « *R-Tree* » qui est basé sur une courbe de Hilbert entrainera une consommation de ressources plus importante que nécessaire, car un plus grand nombre de cellules doit être consulté par l'algorithme multi-niveaux lors de l'opération de forage. Cette situation est comparable à celle décrite pour l'optimisation des hiérarchies géométriques à la section précédente (cf. 4.2.3 Optimiser la dimension géométrique matricielle pour favoriser les fonctions focales). La figure suivante illustre un exemple de courbe de Hilbert appliquée à une hiérarchie de couverture matricielle pour laquelle la valeur **δ** (nombre de cellules enfant composant relation parent-enfant de deux niveaux adjacents) est exprimée par une base 2. Selon cet exemple, il est nécessaire d'interroger minimalement quatre (4) cellules du niveau agrégé (parent) afin d'obtenir une fenêtre de 3 x 3 au niveau plus détaillé (enfant) par une opération de forage sur les membres.

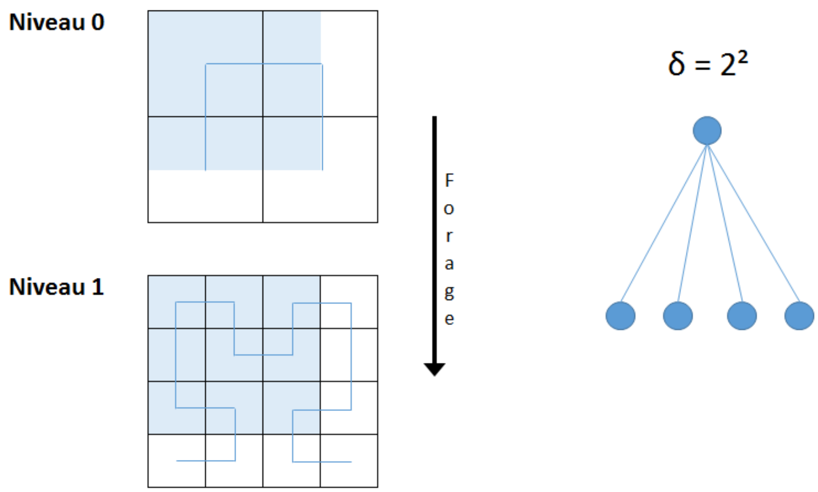


Figure 4.19 — Exemple non optimisé pour l'analyse spatiale matricielle multi-niveaux ayant recours à une courbe de Hilbert

Pour remédier à cette situation, il est nécessaire de recourir à des courbes de remplissage dont la base est équivalente à celle de la fonction focale (3, 5 ou 7). La base ternaire (base 3) proposée par Peano permet de diviser l'espace en neuf (9) cellules identiques et de les parcourir de façon séquentielle à la manière d'une fonction. La figure suivante illustre une courbe de remplissage de Peano du premier au troisième ordre.

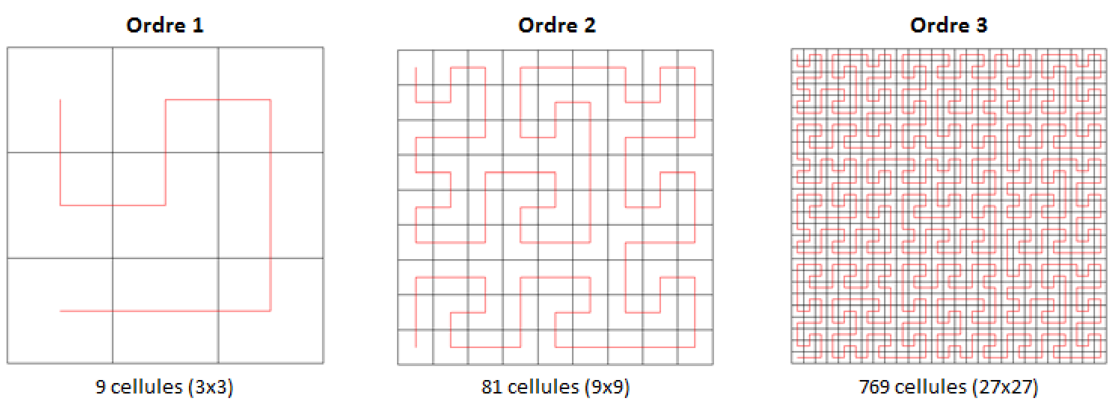


Figure 4.20 — Exemple de courbe de Peano (Ordre 1 à 3) appliqué à des couvertures matricielles

La courbe de Peano à base ternaire est théoriquement plus favorable à l'indexation des cubes matriciels et des dimensions matricielles géométriques destinées à l'analyse spatiale matricielle multi-niveaux, car elle est compatible avec les fonctions focales exploitant des fenêtres de 3 x 3. Ainsi, lors d'une opération de forage (figure ci-dessous) d'un niveau agrégé vers un niveau détaillé, il n'y aura pas de cellules lues inutilement au niveau agrégé. Le recours à une hiérarchie de couverture matricielle possédant une valeur **δ** compatible avec les fonctions focales de l'analyse spatiale matricielle multi-niveaux et un index comportant la même base devrait théoriquement procurer de meilleures performances.

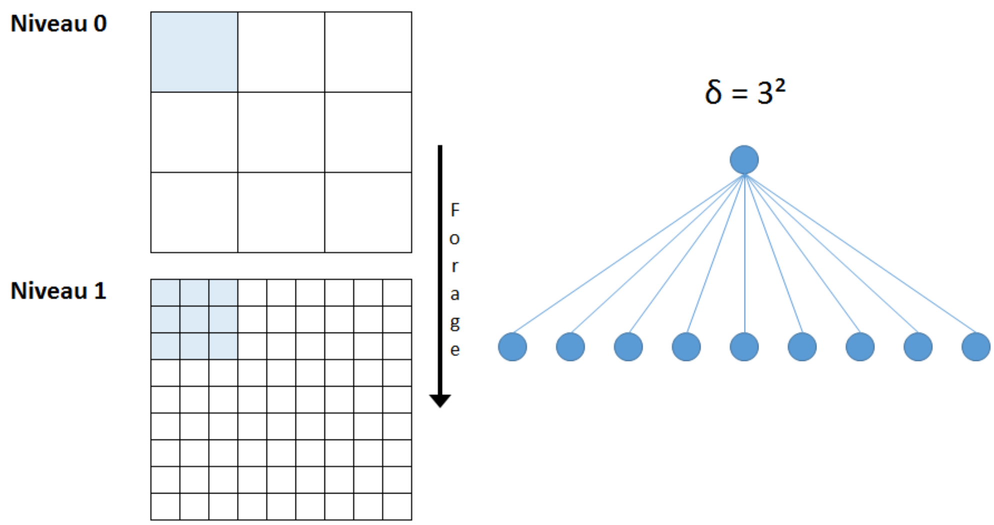


Figure 4.21— Exemple optimisé pour l'analyse spatiale matricielle multi-niveaux ayant recours à une courbe de Peano à base ternaire

Finalement, il faut rappeler que les algorithmes d'analyse spatiale matricielle multi-niveaux constituent une proposition de solution théorique. La revue de littérature n'a pas d'identifier de méthodes d'indexation implémentant une courbe de Peano à base ternaire. L'élaboration de ce type d'algorithme et la démonstration empirique de leur efficacité devront faire l'objet de travaux futurs.

4.4.3 Constats et recommandations à l'égard de l'index pour les analyses spatiales matricielles multi-niveaux

La mise en œuvre d'index pour les cubes matriciels exploitant des dimensions matricielles géométriques est un concept nouveau. Il nécessite d'élaborer des algorithmes d'analyse spatiale matricielle spécialisée capable d'exploiter les niveaux géométriques des hiérarchies de couvertures matricielles. Aussi, ils requièrent que les systèmes de gestion de base de données soient capables d'implémenter ce type de structure et de les interroger efficacement. Un index hiérarchique, multi-niveaux et conçu pour les cubes spatiaux comme le « *Hierarchical Spatial Bitmap* – HSB » [Siquiera et *al*., 2012] constitue un pas dans la bonne direction, mais n'est pas adapté pour les hiérarchies de couvertures matricielles.

En effet, du point de vue de l'analyse spatiale matricielle multi-niveaux, l'index HSB ne considère par les particularités des hiérarchies de couvertures matricielles. Cet index vise à structurer hiérarchiquement les rectangles minimaux englobants ou « *MBR – Minimal Bounding Rectangle* » des objets vectoriels à différentes échelles afin d'obtenir un temps de requête exprimé par la forme générale **θ *(log N)***. Or, les hiérarchies de couvertures matricielles possèdent une topologie explicite. En effet, les cellules d'une même couverture possèdent un positionnement relatif exprimé à l'aide des identifiants de ligne et de colonne facilitant ainsi la recherche des cellules voisines. Également, une telle hiérarchie possède une relation unitaire (*1..N*) entre les cellules parents et enfants des couvertures adjacentes. Ce type de relation est possible avec des données vectorielles, mais n'est pas obligatoire dans les dimensions vectorielles géométriques. Enfin, le rectangle minimal englobant de chaque géométrie ne se superpose pas à celui des cellules voisines comme c'est souvent le cas pour les objets vectoriels. Aussi, cette méthode d'indexation ne considère pas les particularités des analyses spatiales matricielles. Les analyses spatiales matricielles parcourent un sous-ensemble de cellules de façon séquentielle afin de produire les résultats d'analyse. L'assemblage logique sous forme de groupes de cellules est nécessaire pour produire certains types d'analyse spatiale matricielle (*cf. 2.3.3 Balayage des couvertures matricielles*). C'est le cas notamment des analyses par filtre de convolution et les analyses basées sur l'inverse de la distance qui sont en réalité des fonctions focales [Tomlin, 1990]. Il est donc nécessaire d'**élaborer des méthodes d'indexation spécifiques aux cubes matriciels favorisant l'analyse spatiale matricielle à la volée**.

Nous considérons que les indexations reposant sur les courbes de remplissage ou « *space filling curves* » constituent la seule avenue envisageable pour favoriser les fonctions focales d'analyse spatiale matricielle à la volée. Un index favorisant l'exécution à la volée des analyses spatiales matricielles à l'aide de dimension matricielle géométrique doit comporter les caractéristiques suivantes :

- Permettre de parcourir, de façon séquentielle, les cellules d'une couverture matricielle;
- Être multi-niveaux (ou multi-échelles) afin de parcourir les hiérarchies de couvertures matricielles;
- Posséder un système de référence multidimensionnel afin de permettre l'interrogation de cubes;

Les courbes de remplissage sont compatibles avec l'ensemble de ces caractéristiques. Ceci dit, un index de cube matriciel basé sur une courbe de remplissage n'est pas adapté à toutes les fonctions d'analyse spatiale matricielle. En effet, il ne favorise pas la mise en œuvre de fonctions d'analyse linéaire (zonale) telles que requises pour le krigeage. Aussi, la base de l'index doit être adaptée à la taille des fenêtres de convolution employées dans l'analyse spatiale matricielle. Il n'est pas envisageable de concevoir un index comportant une base unique compatible avec toutes les tailles de fenêtres de convolution.

Il est donc nécessaire de **planifier, lors de la conception du cube, le type d'analyse spatiale matricielle qui sera supporté par le cube matriciel**.

Finalement, il faut rappeler que les algorithmes d'analyse spatiale matricielle multi-niveaux constituent une proposition de solution théorique. La revue de littérature n'a pas permis d'identifier des algorithmes multi-niveaux exploitant les dimensions matricielles géométriques dans un contexte SOLAP. L'efficacité empirique de cette piste de solution devra être démontrée dans des travaux futurs.

4.5 Conclusion

Ce chapitre visait à explorer des façons d'optimiser le traitement des analyses spatiales matricielles dans un contexte SOLAP. Nous avons entrepris de tirer avantage de la hiérarchisation géométrique des couvertures matricielles afin d'optimiser l'exécution à la volée des analyses spatiales matricielles. Dans le cadre de ces travaux de recherche, différents scénarios d'optimisation ont été envisagés et plusieurs furent rejetés. Notamment, l'indexation selon la méthode HHCode – Helical Hyperspatial Code [Varma, 1990] fut considérée pour sa capacité à référencer des grands volumes de données dans un espace multidimensionnel. Cependant, cette méthode d'indexation nécessite une relation stricte **δ** = 8 (un parent pour huit enfants) ce qui était restrictif et entrainait une consommation excessive de ressources pour le traitement des analyses spatiales matricielles reposant sur une fonction focale (3 x 3, 5 x 5, 7 x 7). L'analyse et le rejet de cette méthode d'indexation a notamment permis de déterminer qu'il fallait considérer les relations entre les cellules de niveaux adjacents pour les fonctions focales.

La nécessité de concevoir des algorithmes d'analyse spatiale matricielle spécialisés, capable d'exploiter la structure des hiérarchies de couvertures matricielles, s'est imposée d'elle-même au cours de cette recherche. Ces algorithmes multi-niveaux permettent notamment de réduire le domaine spatial d'analyse pertinent en cours d'exécution de façon à interroger, au niveau le plus détaillé, seulement un sous-ensemble pertinent de cellules. Également, la nature de la relation entre les cellules de niveaux adjacents fut considérée comme un facteur déterminant de la performance des analyses spatiales matricielles, ceci particulièrement pour les fonctions focales. En effet, l'étude de cette relation permet de prédire une consommation excessive de ressources informatiques lors du forage spatial qui est une étape essentielle des algorithmes d'analyse spatiale matricielle multi-niveaux. Ensuite, afin d'optimiser les algorithmes d'analyse spatiale matricielle multi-niveaux, il fut proposé de calculer préalablement des statistiques utiles pour le traitement de nombreuses analyses spatiales matricielles. Ces statistiques, planifiées lors de la modélisation du cube, doivent en théorie permettre d'accélérer la recherche de valeurs thématiques pertinentes. Finalement, les caractéristiques d'un index spatial favorisant l'analyse spatiale matricielle à la volée furent décrites. Cet index doit être compatible avec les fonctions d'analyse spatiale matricielle, notamment les fonctions focales, afin d'optimiser les ressources nécessaires au traitement de ces algorithmes. En conclusion, il apparaît théoriquement possible de produire des analyses spatiales matricielles à la volée dans les cubes matriciels en mettant à profit les pistes de solution élaborées dans le cadre de cette recherche. Cependant, l'analyse spatiale matricielle dans les cubes matriciels n'est pas triviale et exige un effort de planification, lors de la conception des dimensions matricielles géométriques et du cube matriciel, afin de favoriser l'exécution à la volée.

Chapitre 5. Discussions et travaux futurs

5.1 Objectifs et résultats

Les structures de données OLAP et SOLAP sont conçues pour interroger des axes d'analyse et obtenir en quelques opérations simples les résultats du croisement de ces variables indépendantes. Le faible temps nécessaire pour obtenir ces résultats est une caractéristique fondamentale et constitue en bonne partie l'intérêt de l'OLAP pour les applications d'aide à la décision. Dans un contexte d'analyse spatiale à la volée, un temps réponse satisfaisant ($t_{requête}$ < 10 secondes) est difficile à obtenir sans calculer préalablement l'ensemble des croisements spatio-temporels et stocker ces résultats d'analyse pour une interrogation future. Ce constat s'applique aux cubes spatiaux vectoriels, mais aussi pour les cubes matriciels. À ce jour, aucune étude ne décrit le potentiel ni la façon dont les cubes matriciels peuvent être utilisés en complément des cubes vectoriels afin de minimiser le calcul préalable des analyses spatiales. Cette situation prévaut pour l'ensemble de la littérature rencontrée sur les SOLAP et non seulement pour les travaux pionniers faits au Centre de Recherche en Géomatique (CRG). Incidemment, l'incertitude concernant la capacité des cubes matriciels à répondre à l'exigence de performance d'exécution est très grande et difficile à quantifier. C'est dans ce contexte que ces travaux de recherche MSc ont été entrepris. Ce mémoire constitue le résultat final et présente une synthèse des réflexions qui ont eu cours dans le cadre de ces travaux de recherche.

L'objectif de cette recherche visait à **explorer les avantages et les considérations à exploiter les cubes spatiaux matriciels afin de produire des analyses spatiales à la volée dans un contexte décisionnel**. Cet objectif a été atteint par l'achèvement des deux sous-objectifs de recherche. Le premier sous-objectif consistait à **identifier les causes de la consommation de ressources par les analyses spatiales matricielles**. Dans le contexte SOLAP matriciel, la discontinuité du domaine spatiale et la disparité des caractéristiques des couvertures furent identifiées comme les principales sources externes de consommation additionnelle de ressources induisant un temps de traitement supérieur au seuil visé. Il nous apparaissait essentiel d'intervenir en amont de l'observation de la discontinuité du domaine spatiale d'analyse et de pousser plus loin le cadre théorique en intégrant explicitement la notion de couverture matricielle dans les composantes du cube. Cette démarche visait à mieux gérer l'homogénéité des couvertures et la continuité sur le domaine spatial de l'analyse. Nos efforts se sont concentrés sur l'intégration des couvertures matricielles dans les dimensions géométriques matricielles (exclusion des dimensions hybrides et mixtes) et directement dans les faits détaillés du cube sans l'intervention de dimensions géométriques matricielles. Cette réflexion a donné naissance à sept (7) scénarios théoriques d'intégration des couvertures dans les cubes (ex. : couverture associée au fait détaillé, au niveau d'une dimension ou au membre d'une dimension). De ces scénarios, seulement cinq (5) permettent d'assurer la continuité sur le domaine spatial d'analyse. Les résultats de ces observations ont mené à étendre la définition du cube matriciel afin d'y inclure formellement la notion de couverture matricielle. Il fut également indispensable de redéfinir la dimension géométrique matricielle afin d'y intégrer la notion de hiérarchie thématique. Cette notion est essentielle à l'intégration et au partage d'une couverture entre les membres d'une dimension ou d'une hiérarchie sans obligatoirement faire varier la résolution spatiale. Ensuite, il fut nécessaire d'étendre la clé d'identification du modèle physique des hiérarchies SOLAP afin d'y ajouter deux nouvelles formes de hiérarchies favorables à l'analyse spatiale matricielle à la volée. Ces nouvelles hiérarchies, qui devaient être compatibles avec les hiérarchies de couvertures matricielles, sont le résultat de l'étude de la relation de subordination entre les cellules des niveaux des couvertures des dimensions matricielles géométriques. Enfin, dans le but d'identifier les sources internes de consommation de ressources par les analyses spatiales matricielles, la complexité mathématique des algorithmes d'analyse spatiale matricielle fut analysée. Les résultats de cette analyse ont permis d'associer la forme générale de complexité **θ (*N*)** (algorithme linéaire) aux algorithmes d'analyse spatiale matricielle. Dans ce contexte, le temps nécessaire à la résolution est une fonction de la taille ***N*** de la couverture analysée (nombre de cellules formant la couverture). Plutôt que de tenter de réduire la taille de la couverture analysée, nous avons entrepris d'identifier des moyens d'optimiser la forme générale de l'algorithme en exploitant la hiérarchie de couverture matricielle. Les moyens proposés pour arriver à la forme **θ *(log N)*** sont exposés dans le second sous-objectif.

Le second sous-objectif consistait à **identifier des moyens d'optimiser les ressources pour le traitement des analyses spatiales matricielles**. L'analyse des scénarios d'intégration des couvertures matricielles dans les composantes du cube a permis d'identifier les scénarios les plus favorables à l'analyse spatiale matricielle à la volée. Selon les critères établis (continuité du domaine spatial et homogénéité des caractéristiques des couvertures), les scénarios les plus favorables à l'exécution à la volée des analyses spatiales produisent des cubes comportant une seule couverture, mais plusieurs couches distinctes (variables thématiques). Toutefois, il est nécessaire d'uniformiser préalablement le domaine de valeurs des mesures du cube en exploitant un seul type de variable numérique (ex : entier ou réel). Ce procédé d'optimisation permet d'éviter le recours à la fusion des couvertures[14] lors de l'exécution de l'analyse spatiale matricielle. Dans un deuxième temps, nous avons entrepris de transformer la forme générale **θ (*N*)** décrivant le temps nécessaire à l'exécution de l'algorithme d'analyse spatiale matricielle à la forme **θ *(log N)*** en exploitant la structure des hiérarchies de couvertures

[14] La **fusion des couvertures** ou des couches d'une même couverture est une étape préalable à l'analyse spatiale matricielle, lorsque plusieurs couches ou couvertures sont impliquées dans le traitement, afin d'uniformiser la résolution spatiale et le domaine de valeurs.

matricielles. Afin d'y parvenir, nous avons établi qu'il est nécessaire de concevoir des algorithmes d'analyse spatiale multi-niveaux capables d'exploiter les hiérarchies de couvertures matricielles à la façon d'un arbre binaire.

Cette proposition, inspirée des algorithmes multi-échelles issus du domaine du traitement d'image, est novatrice pour le domaine d'application SOLAP. En effet, la revue de littérature n'a pas permis de trouver d'exemple de ce type d'algorithme d'analyse spatiale matricielle. Dans le but de favoriser davantage l'exécution à la volée en exploitant les algorithmes multi-niveaux, nous avons également proposé de calculer préalablement des statistiques pour accélérer la recherche de valeurs thématiques à l'aide de la hiérarchie de couvertures matricielles. Le recours à une énumération des valeurs ou aux bornes maximales et minimales d'un sous-ensemble de cellule favorise à notre avis le traitement de certaines analyses spatiales matricielles. Enfin, nous nous sommes penchés sur les méthodes d'indexation qui seraient compatibles à la fois avec les fonctions locales, focales et zonales des algorithmes d'analyse spatiale matricielle. Il fut déterminé qu'il serait nécessaire de mettre au point un index spatial exploitant la hiérarchie de couvertures matricielles basé sur les courbes de remplissage de Peano pour favoriser à la fois les fonctions locales et focales d'analyse spatiale matricielle. Ce type d'index est compatible avec les aspects multidimensionnels, multi-échelles et l'énumération séquentielle, ce qui en fait un candidat de choix pour ce type d'application. Cependant, la base de la courbe de remplissage, c.-à-d. le nombre de cellules parcourues avant de produire une variation dans le patron de la courbe, doit être un multiple de la taille de la fenêtre focale de l'analyse spatiale matricielle afin de ne pas engendrer une consommation indue de ressources lors du traitement. Il est donc nécessaire de connaître la taille des fenêtres focales à mettre en œuvre avant de déterminer la base de la courbe de remplissage à employer. Enfin, les moyens identifiés pour optimiser la consommation de ressources pour le traitement des analyses spatiales matricielles ne sont pas exhaustifs, c.-à-d. qu'elle est basée uniquement sur l'exploitation de la hiérarchie de couvertures matricielles. Toutefois, ces pistes de solutions sont favorables à l'exécution à la volée des analyses spatiales matricielles dans le SOLAP sans nécessiter le calcul préalable des résultats.

L'atteinte de ces objectifs s'inscrit dans l'objectif global d'accroitre la base de connaissances spécifique à l'analyse des problématiques matricielles décisionnelles. La compréhension accrue des caractéristiques de l'exécution à la volée devrait permettre de proposer des pistes de solution visant à optimiser les ressources nécessaires au traitement des analyses spatiales matricielles dans le contexte posé. Les algorithmes d'analyse spatiale matricielle sont nombreux et diversifiés. Ces travaux de recherche ont initié l'exploration des avantages et des considérations à exploiter les cubes matriciels afin de produire des analyses spatiales matricielles à la volée. Nous avons identifié des causes internes et externes de la consommation des ressources qui accroissent le temps nécessaire au traitement des analyses spatiales matricielles. Nous avons également proposé des moyens d'accélérer le traitement dans un contexte SOLAP afin de tirer avantage de la forte structuration des membres des dimensions sous forme de hiérarchie combinée à celle des couvertures matricielles. Finalement, nous avons fait évoluer le cadre théorique entourant l'intégration des données matricielles dans les cubes spatiaux, afin de mieux l'adapter aux besoins spécifiques de l'analyse spatiale matricielle. Cette recherche se voulait exploratoire, la section qui suit met en lumière des pistes de recherche pour des travaux futurs qui sont issus de réflexions inexplorées dans le cadre de ces travaux et à l'approfondissement des concepts élaborés.

5.2 Travaux futurs

La présente recherche s'est intéressée à une problématique générale qui va au-delà de ce qu'il est possible de solutionner à l'intérieur d'un mémoire. Il était donc nécessaire de poser certaines limites concernant la portée du projet. En premier lieu, aucune considération n'a été apportée aux aspects de visualisation et de navigation dans les résultats d'analyse spatiale matricielle, et ce malgré le contexte d'utilisation intimement lié à l'outil visuel SOLAP. Dans ces travaux, l'aspect analytique est strictement mis de l'avant tout en considérant le contexte géodécisionnel comme un atout indéniable. Pour la suite, il serait intéressant d'explorer la combinaison des analyses spatiales visuelles et des analyses spatiales matricielles basées sur des algorithmes afin d'obtenir des résultats à la volée. Notamment, le domaine de la composition d'image de synthèse par superposition de couches décrit par Jensen [Jensen, 2005] pourrait selon nous être un complément pour obtenir des résultats à la volée. En second lieu, le stockage des résultats d'analyse et leurs réutilisations, par exemple pour la composition d'analyses plus complexes, n'ont pas été abordés dans ce mémoire. Nous croyons qu'il serait potentiellement utile de conserver les résultats des analyses spatiales matricielles déjà produites pour les interroger à nouveau en les intégrant dans une dimension d'analyse spatiale. De plus, le stockage de résultats de certaines analyses, à la manière d'une structure de données HOLAP, pourrait peut-être permettre d'atteindre un équilibre entre ce qui doit être calculé préalablement et ce qui doit être calculé à la volée (performance d'interrogation vs espace de stockage). Enfin, ce projet de recherche présente les enjeux théoriques de la modélisation spatiale matricielle multidimensionnelle et de l'exécution à la volée des analyses spatiales matricielles. La validation des pistes de solution proposées, par l'élaboration de cube matriciel et d'algorithmes multi-niveaux, permettrait de quantifier les gains de performance. Finalement, les exemples proposés dans ces travaux de recherche concernent l'utilisation du territoire et l'exploitation forestière. Nous croyons que le domaine de l'imagerie médicale, notamment par résonnance magnétique, pourrait grandement bénéficier d'analyses spatiales matricielles à la volée pour l'aide à la décision. Nous croyons que ce domaine d'application du SOLAP matriciel constitue une avenue de recherche et d'application de ces travaux qu'il serait pertinent d'explorer dans le cadre de travaux futurs.

Bibliographie

[Balayoghan], Balayoghan, V.B., Space filling curves dans University of Texas, Consulté le 2010-03-30, [En ligne], http://userweb.cs.utexas.edu/users/vbb/misc/sfc/0index.html

[Bédard et al., 2001] Bédard, Y., T. Merrett & J. Han, 2001, Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery, Geographic Data Mining and Knowledge Discovery, Taylor & Francis, Vol. Research Monographs in GIS, No. Chap. 3, p. 53-73

[Bédard, 2004] Bédard, Y., 2004, Amélioration des capacités décisionnelles des SIG par l'ajout d'un module SOLAP (Spatial On-Line Analytical Processing), Univesité Aix-Marseille, École Polytechnique Universitaire de Marseille, Filière Génie Industriel et Information, 8 avril (Professeur Invité)

[Bédard, 2004-B] Bédard Y., Larrivée S., Proulx M.-J., Nadeau M., 2004, Modeling Geospatial Databases with Plug-Ins for Visual Languages : A Pragmatic Approach and the Impacts of 16 Years of Research and Experimentations on Perceptory, S. Wang et al. (Eds.) : COMOGIS Workshops ER2004, LNCS 3289, pp. 17-30, 2004., Shanghai

[Bédard, 2007] Bédard, Y., S. Rivest, & M.-J. Proulx, 2007, Spatial On-Line Analytical Processing (SOLAP) : Concepts, Architectures and Solutions from a Geomatics Engineering Perspective, Dans: Robert Wrembel & Christian Koncilia (ed(s)), Data Warehouses and OLAP : Concepts, Architectures and Solutions from a Geomatics Engineering Perspective

[Bédard et al., 2008] Bédard, Y., J. Han, 2008, Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery, In: Geographic Data Mining and Knowledge Discovery, 2e edition, Chap. 3, Taylor & Francis.

[Bédard et al., 2008-B] Bédard Y., Larrivée S., Modeling with Pictogrammic languages, In : Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Bernier, 2006] Bernier, E., Y. Bédard, 2006, Développement de technologies géospatiales Livrable 2 : Rapport et démonstrateur technologique pour la création de données, des métadonnées et leur utilisation, Rapport de recherche pour Recherche Défense- Valcartier, Mars, 69p.

[Bernier, 2008] Bernier, E., 2008, Réalisation de l'application SOLAP : Documentation des choix conceptuels et exemples d'analyse, Rapport de recherché pour le projet GEOIDE « Un outil Web interactif innovateur pour mieux comprendre les vulnérabilités de santé liées au climat », Mars, 27p.

[Bernier et al., 2008] Bernier, E., Y. Bédard, P. Gosselin, T. Badard, 2008, La technologie SOLAP comme support à la prise de décision en groupe : Exemple d'application en santé publique, Colloque (Outils pour décider ensemble), 5-6 juin, Québec, Canada

[Bernier, 2009] Bernier, E., P. Gosselin, T. Badard, Y. Bédard, 2009, Easier Surveillance Of Climate-Related Health Vulnerabilities Through A Web-Based Spatial Olap Application, International Journal of Health Geographics, Vol8, No. 18.

[Brodeur, 2008] Brodeur J., Badard T., Modeling with ISO 191xx Standards, In : Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Caron, 1998], Caron, Pierre-Yves, Étude du potentiel de OLAP pour supporter l'analyse spatio-temporelle, Mémoire présenté à la Faculté des études supérieures de l'Université Laval, Québec, 1998, 129 p.

[Chaker W et al., 2009] Chaker, W., M. Proulx, B. Moulin, Y. Bédard, 2009, Modélisation, Simulation et Analyse d'Environnements Urbains Peuplés : Approche multi-agent pour l'étude des déplacements multimodaux, Revue Internationale de Géomatique, Vol. 19, No. 4, pp. 413-441

[Couture, 1997] Couture M., Fournier, R.-P., La recherche en sciences et génie — Guide pratique et méthodologique, Les presses de l'Université Laval, Québec, Qc., 1997, 261 p.

[CCT, 2005], Résolution spatiale dans Ressources Naturelles du Canada, Gouvernement du Canada, Consulté le 2010-04-03, [En ligne], http://www.ccrs.nrcan.gc.ca/glossary/index_f.php?id=562

[Dawson, 2005] Dawson, Christian W., Projects in Computing and Information Systems — A student's guide, Pearson Education Limited, Edinburgh Gate, England, 2005, 246 p.

[De Smith, 2007] De Smith, Michael J., Goodchild, Michael F., Longley, Paul A., Geospatial Analysis, A comprehensive guide to principles, techniques and software tools, Matador on behalf of The Winchelsea Press, Leicester, Second Edition, 2007, 437 p.

[Elbassioni, 2008] Elbassioni K., Elmasry A., Ibrahim K., Indexing Schemes for Multi-dimensional Moving Objects, In : Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus.

[Goffe, 2011] Romain Goffe, Top-down irregular pyramids for large hostological images segmentation, Thèse présentée à la Faculté des Sciences Fondamentales et appliquées de l'Université de Poitiers, France, 2011, 197p.

[Gupta, 2008] Gupta S., Ravishankar C. V., Spatio-temporal Queries on Road Networks, Coding Bases Methods, In: Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Hadjieleftheriou, 2008] Hadjieleftheriou M., Kollios G., Vassilis J. T., Gunopulos D., Indexing Spatio-temporal archives, In : Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Kollios, 2008] Kollios G., Tsotras V., Mobile Object Indexing, In: Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Jensen, 2005], Jensen, J. R., Introductory digital image processing - a remote sensing perspective, Prentice Hall Series in Geographic Information Science, Third Ed., N.J., 2005, 525 p.

[Jensen, 2008-A] Jensen C., Lin D., Ooi, B. C., Indexing of Moving Objects, Bx Tree, In: Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Jensen, 2008-B] Jensen C., Lin D., Ooi, B. C., Indexing of Moving Objects, Bx Tree, In: Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Larrivée, 2005] Larrivée, S., Y Bédard & J Pouliot, 2005, How to Enrich the Semantics of Geospatial Databases by Properly Expressing 3D Objects in a Conceptual Schema, R. Meersman et al. (Eds.): OTM Workshops 2005, LNCS 3762, pp. 999 – 1008.

[Laurini, 1993] Laurini, Robert, Milleret-Raffort, Françoise, Les bases de données en géomatique, Traité des nouvelles technologies série Géomatique, Hermes, Paris, 1993, 340 p.

[Lisboa, 2008] Lisboa Filho J., Iochpe C., Modeling with UML Profile, In : Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Marchand, 2004] Marchand, P., The Spatio-Temporal Topological Oparator Dimension, a Hyperstructure for Multidimensional spatio-temporal Exploration and Analysis, Mémoire présenté à la Faculté des études supérieures de l'Université Laval, Québec, 108 p.

[McHugh, 2008], McHugh, R.-M., Intégration de la structure matricielle dans les cubes spatiaux, Mémoire présenté à la Faculté des études supérieures de l'Université Laval, Québec, 2008, 127p.

[Miller, 2008] Miller J. H., Time Geography, In: Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Miralles, 2008] Miralles, A. Lubourel T., 2008, Modeling with Enriched Model Driven Architecture, In: Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[MRN-Tordeuse, 2013] MRN – Tordeuse des bougeons de l'épinette, [En ligne]. Adresse URL : http://www.mrn.gouv.qc.ca/forets/fimaq/insectes/fimaq-insectes-insectes-tordeuse.jsp (Page consultée le 11 octobre 2013)

[OQLF, 2004], Grand dictionnaire terminologique, Office Québecois de la langue Française, Consulté le 2010-04-01, [En ligne], http://www.granddictionnaire.com

[Parents, 2008] Parents C., Spaccapietra S., Zimanyi E., Modeling and multiple perceptions, In : Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Perlmonks, 2013] Perlmonks, An informal introduction to O(N) notation, [En ligne]. Adresse URL : http://www.perlmonks.org/?node_id=227909 (Page consultée le 4 avril 2013)

[Perceptory, 2006] Perceptory 2006, Équipe de BD spatiales, CRG, Université Laval, Consulté le : 2009-12-18, [En-ligne], Dernière modification : 2005-12-12, http://sirs.scg.ulaval.ca/perceptory/

[Proulx, 2008] Proulx, M., Y. Bédard, 2008, Fundamental Characteristics of Spatial OLAP Technologies as Selection Criteria, Location Intelligence 2008, April 29, Santa Clara, CA, USA.

[Ress-ArcGIS, 2013] ArcGIS Ressources, ArcGIS Help 10.1 – Raster pyramids, [En ligne]. Adresse URL : http://resources.arcgis.com/en/help/main/10.1/009t/009t00000019000000.htm (Page consultée le 4 avril 2013)

[Rigaux, 2002] Rigaux P., Scholl M., Voisard A., Spatial Databases with application to GIS, Academic Press, Morgan Kaufmann Publishers, San Francisco, CA. USA, 2002, 409 p.

[Rivest et al., 2005] Rivest, S., Y. Bédard, M.-J. Proulx, M. Nadeau, F. Hubert & J. Pastor, 2005, SOLAP : Merging Business Intelligence with Geospatial Technology for Interactive Spatio-Temporal Exploration and Analysis of Data, Journal of International Society for Photogrammetry and Remote Sensing (ISPRS) "Advances in spatio-temporal analysis and representation", 60 (1), pp. 17-33.

[Rivest, 2007] Rivest, S., Y. Bédard, 2007, Spatial databases, Dans : K.K. Kemp (ed(s)), Encyclopedia of Geographical Information Science, pp. 62-66.

[Saltenis, 2008] Saltenis, S., Indexing the positions of continuously moving objects, In: Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Siquiera et al., 2012] Siquiera T. et al., The SB-index and the HSB-index: efficient indices for spatial data warehouses, GeoInformatica, 2012, Vol. 16(1), 165-205

[Spatial-Analyst, 2013] Spatial-Analyst, Pyramid layers for large raster maps, [En ligne]. Adresse URL : http://spatial-analyst.net/ILWIS/htm/ilwis/ilwis_objects_raster_maps_pyramids.htm (Page consultée le 4 avril 2013)

[Spofford, 2006] Spofford, George and al., MDX Solutions, Wiley Publishing inc., Indianapolis, Indiana, Second Edition, 2006, 716 p.

[Tryfona, 2008] Tryfona N., Jensen C.S., Spatio-temporal Database Modeling with Extended Entity-Relationship Model, In : Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Tao, 2008] Tao Y., Queries in Spatio-Temporal Databases, Time Parameterized, In : Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Tomlin, 1990], Tomlin, C.D., Geographic information systems and cartographic modelling. Prentice Hall, N.J., 1990, 249 p.

[Vaisman, 2009-A] Vaisman, A., E. Zimanyi., 2009, A Multidimentional Model Representing Continuous Fields in Spatial Data Warehouses, ACMGIS09, Seattle, WA, USA.

[Vaisman, 2009-B] Vaisman, A., Zimanyi E., What is Spatio-Temporel Data Warehousing, DaWaK '09 : Proceedings of the 11th International Conference on Data Warehousing and Knowledge Discovery, Springer-Verlag, Berlin, Heidelberg, 2009, p. 9-23

[Waller, 2004] Waller, Lance A., Gotway, Carol A., Applied Spatial Statistics for Public Health Data, Wiley series in probability and statistics, John Wiley and Sons inc., N.J., 2004, 494 p.

[Wiki-BigO, 2013] Wikipedia, Big O notation – Wikipedia, the free encyclopedia, [En ligne]. Adresse URL : http://en.wikipedia.org/wiki/Big_O_notation (Page consultée le 4 avril 2013)

[Wiki-Mipmap, 2013] Wikipedia, Mipmap – Wikipedia, the free encyclomedia, [En ligne]. Adresse URL : http://en.wikipedia.org/wiki/Mipmap (Page consultée le 4 avril 2013)

[WMTS, 2012] OpenGIS Web Map Tile Service, [En ligne]. Adresse URL : http://www.opengeospatial.org/standards/wmts (Page consultée le 4 avril 2013)

[Worboy, 1995] Worboys, Michael F., GIS — A computing perspective, Tayloy & Francis Ltd., London, 1995, p.371

[Xiong, 2008] Xiong X., Mokbel M. F., Aref W.G., Continuous Queries in Spatio-Temporel Databases , In : Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Yiu, 2008] Yiu M. L., Tao Y., Indexing, BDual Tree, In: Encyclopedia of GIS, Shekhar & Xiong Eds, Springer, XL, 2008, 1377 p., 723 illus

[Zobel, 2004] Zobel, J., Writing for computer science, Springer Verlag, London Limited, Second edition, 2004, 270 p.